International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 214 - 219, 2015, ISSN:- 2319–8656

An Unsupervised Change Detection in Satellite IMAGES Using MRFFCM Clustering

S.Venkata Lakshmi
Dept. of CSE
Panimalar Institute of Technology
Chennai,India

K.Sathyamoorthy
Dept. of CSE
Panimalar Institute of Technology
Chennai,India

T.K Senthil Kumar
Dept. of ECE
Rajalakshmi Institute of Technology
Chennai,India

Abstract*:* This paper presents a new approach for change detection in synthetic aperture radar images by incorporating Markov random field (MRF) within the framework of FCM. The objective is to partition the difference image which is generated from multitemporal satellite images into changed and unchanged regions. The difference image is generated from log ratio and mean ratio images by image fusion technique. The quality of difference image depends on image fusion technique. In the present work; we have proposed an image fusion method based on stationary wavelet transform. To process the difference image is to discriminate changed regions from unchanged regions using fuzzy clustering algorithms. The analysis of the DI is done using Markov random field (MRF) approach that exploits the interpixel class dependency in the spatial domain to improve the accuracy of the final change-detection areas. The experimental results on real synthetic aperture radar images demonstrate that change detection results obtained by the MRFFCM exhibits less error than previous approaches. The goodness of the proposed fusion algorithm by well-known image fusion measures and the percentage correct classifications are calculated and verified.

Keywords-*image change detection, MMRM, LRM, MRFFCM, SAR, fuzzy clustering*

1. INTRODUCTION

Satellite images and geographic mapping is the key to many applications and it depends on the accurate information about the land area in a region.Upon detecting changes in the geographic area at different times is important in many domains.The images are generated by synthetic aperture radar (SAR)[1][2].The unsupervised[11] change detection approach is classified in to 3 steps.1)preprocessing 2) difference image generation 3)analysis of difference image. Usually satellite images are subject to different errors. The geometric errors are systematic, which means that they can be identified and corrected. The images affected by external errors ie.due to movement of air or spacecraft also to be noticed and corrected.The purpose of internal errors correction is removal of geometrics in images.

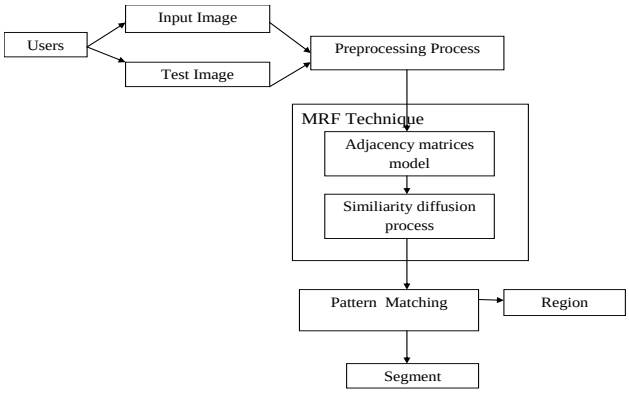


Figure.1. Process of change detection

The noise removal [5] in images is the next process and of most of the SAR images suffer from the presence of speckle noise [4]. The Markov random field is a granular noise that inherently exists in and degrades the quality of SAR images. The next step is to identify the change by comparing two images by pixel by pixel and a new image is generated which is the difference image.

The figure 1 shows the process of the image with MRF technique and the discrimination of changed and unchanged regions are separated using fuzzy clustering algorithms. A mean ratio and log ratio operator is used to identify and by improving MRFFCM clustering algorithms to identify the change areas.The MRFFCM (Markov Random Field Fuzzy C-Means clustering)[4] algorithm is proposed to focus on modification of membership to reduce the effect of speckle noise. The information provided by neighbor pixel serve as the spatial context and it is considered to center pixel in FCM progress.The point wise prior probability is of positive correlation to the new membership, which indicates that we are able to modify the new membership by modifying the point wise prior probability indirectly. This approach is time consuming. The analysis shows that the new approach not only does well in reducing speckle noise but also reduce time complexity.

2. RELATED WORK

An approach for sub pixel change detection in satellite imagery[6], detects the land cover changes using remote sensed images. It provides fine information dedicated to a specific range of application and its characteristics result in between fine spatial resolution and fine spectral resolution at high time frequency .Related another work

represents the changes in satellite imagery by incorporating image fusion and soft clustering techniques. Similar work is analyzed in multitemporal remote sensing images using neural based algorithm

3.1 DETECTION OF CLUSTER CHANGE:

In general, Unsupervised Change detection algorithms usually take two digitized images as input and return the locations where the differences between the two images can be identified. To accomplish this task,preprocessing is necessary focused at rendering two images with respect to spatial and spectral domains.Concerning the spatial domain, the coregistered two images are verified whether they are at same coordinates in the images.The crucial step,if performed in accuracy may lead to registration of noise .Illumination can be assumed to change smoothly with respect to pixel coordinates. Therefore in many situations,we can divide an original scene into different areas of interest (AOI),so that differences in illuminance conditions will be constant for each pixel in a given specified area of interest. It is worth-noting that selecting the AOI dimensions should involve a trade-off between obtaining uniform illumination conditions and relying on sufficient statistics for carrying out the change detection process.

The two registered and corrected images (a linear or non linear) combinations of the spectral bands of such images are compared pixel by pixel in order to generate a further image called as "difference image". The difference images is computed in such a way that pixels are associated with land cover changes present graylevel values significantly different from those of pixels associated with unchanged areas. Toexploit the spatial –contextual information[5], input patterns are generated corresponding to each pixel in the difference image, DI considering its spatial neighborhood system as follows.

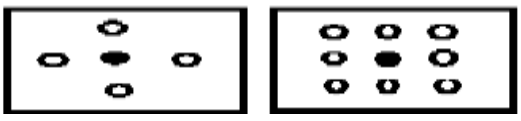


Figure. 2.*a) 1st neighborhood b) 2nd neighborhood*

Now after,generating a two dimensional pattern corresponding to the pixel at position (m,n) of DI by considering the gray value of pixels as one feature and average of the gray value of 8 neighbors as another feature .However when SAR images are considered, changes are obtained by analyzing the image resulting from the applications of the ratio operator to the considered couple of SAR images with mean and log ratios.Hence image fusion technique is introduced to generate a difference image by using complementary information from a mean ratio image and log ratio image. Image fusion is a process of fusing two or more images into a single fused image, there by relevant information in the images are combined. . So this single fused image will be more informative than any of the input images.

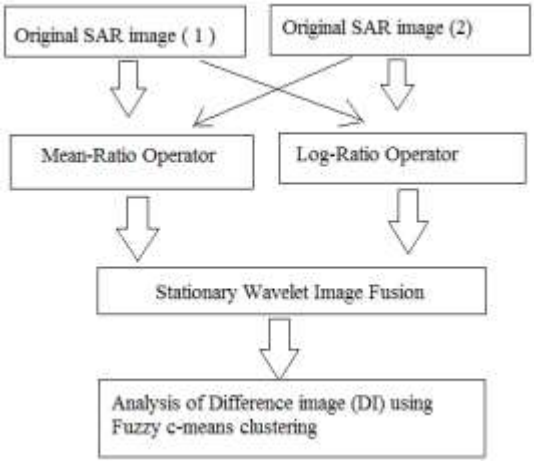


Figure 3: Process of Image Fusion based on the Stationary Wavelet Transform

In the newly produced difference image,the pixel value of no change area approach zero and the pixel value of changed area are positive or negative .The majority of fusion technique are based on wavelet transformation but the image fusion is resulting with shift variant and additive noise in fused image and it doesn't preserve edge a number of clusters of the image. The information loss is more thereby, the clarity of the fused image is also reduced .Ground truth image is finally obtained in which a pixel on the SAR image is compared to what is there in reality (at the present time) in order to verify the contents of the image. It helps to determine the accuracy and thereby minimize errors in the classification.

3.1 Cluster change analysis using Minimum Mean Ratio Method (MMRM):

The MMRM is an algorithm through which the change in the two images can be extracted in spatial domain. Mean is the measure of pixel similarity in an image. Subtracting the mean value from each pixel will produce the resultant image which is discriminative. Hence comparing the mean subtracted images will give the difference image.

Mean-Ratio operator=1-min[e1/f1,f1/e1]

Algorithm steps are:

1. Finding the mean of the two images (m1 and m2).
2. Finding the mean ratio m1 / m2 and m2 / m1.
3. Normalize the mean ratio using the equation [1- min (m1/m2, m2/m1)].
4. Keep these normalized mean as the threshold and make the entire pixel below the threshold as '0' and above the threshold as '1'.

Figure. 4.Mean ratio operator

3.2 Cluster change analysis using log ratio method (LRM):

The logarithmic operator is characterized by enhancing the low-intensity pixels while weakening the pixels in the areas of high intensity. Therefore, the information of changed regions that is obtained by the log-ratio image may not able to reflect the real changed trends in the maximum extent because of the weakening in the areas of high-intensity pixels.

Log-Ratio operator=log1- log2

Algorithm steps are:

1. Finding the mean of the two images (m1 and m2).
2. Log ratio is computed using log(m1)-log(m2).
3. Keep this log ratio as the threshold and make the entire pixel below the threshold as '0' and above the threshold as 1.

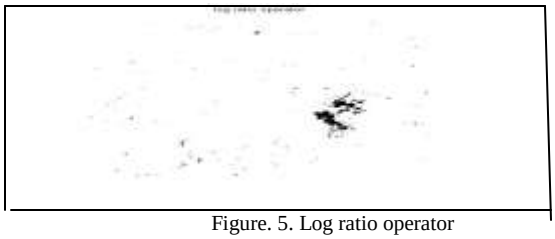


Figure. 5. Log ratio operator

3.3 Fusing MMRM and LRM using wavelet:

The implemented wavelet fusion algorithm steps are: Decompose the MMRM image into low-low, low-high, high-low, high-high frequency region using daubechies wavelet. Decompose the LRM image into low-low, low-high, high-low, high-high frequency region using daubechies wavelet.

1. Find the average of every pixels of the low-low region
2. Compare a pixel of the low-high region of the MMRM image with LRM image. Keep the pixel with the lesser value in the resultant image.
3. Compare a pixel of the high-low region of the MMRM image with LRM image. Keep the pixel with the lesser value in the resultant image.
4. Compare a pixel of the high-high region of the MMRM image with LRM image. Keep the pixel with the lesser value in the resultant image.
5. Perform inverse discrete wavelet transform[12] on step 3,4,5,6 images.

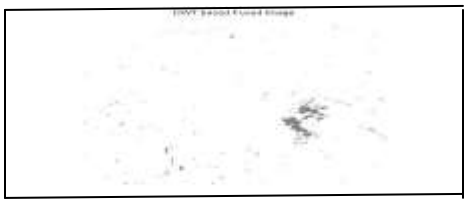


Figure. 6.DWT based fused image

4. USING FUZZY MODIFIED BY MARKOV RANDOM FIELD

The Fuzzy c means is well known and popular to distinguish the changed class from the unchanged class in the domain of image segmentation. Here is the algorithm.

FCM ALGORITHM :

Input : Unlabeled data set consisting of n patterns

Output: Protypes V and fuzzy partition matrix U

Step 1: set fusilier m and initialize u randomly

Step 2: Compute each cluster center using (5)

Step 3 : Compute all possible distances d_{ik}*: i=1,2,…c ;k=1,2,..n*

Step 4 : Update fuzzy partition matrix using (6)

Step 5 : Compute $\Delta=||(u_t-u_{t-1})||$ *; t denotes tth iteration*

Step 6 : Check if $\Delta < \varepsilon$ *,where* ε *is a pre defined small positive constant*

Step 7 : If above condition is not true goto step 2

Step 8 : Stop

Description:

- Choose the number of clusters.
- Assign randomly to each point coefficient for being in the cluster.
- Repeat until the algorithm are converged (that is ,the coefficient change between two iterations is no more than ,the given sensitivity threshold):
- Compute the centroid of each cluster
- for each point compute its coefficient of being in the clusters

The MRFFCM improves FCM by modifying the membership of each pixel according to the MRF based spatial context. Thecalculated the mean values of the two clusters and the cluster whose center is close to the origin (of the featurespace) is labeled as unchanged and another one as changed. The pixels corresponding to changed one are marked as black (graylevel zero)and unchanged ones are marked as white(graylevel 255)

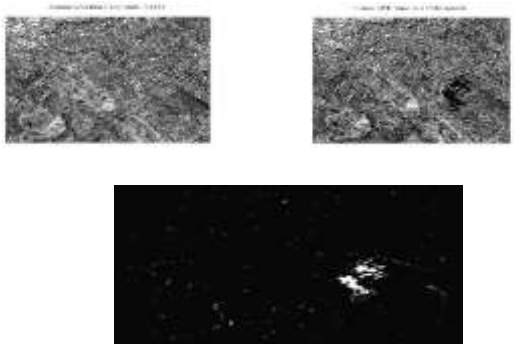


Figure.7.SAR image 1,2 and Segmented image

The figure shows the segmented image after the matching process.

5. DATASETS DESCRIPTION AND EXPERIMENTAL SETTINGS

The main by making qualitative visual analysis and quantitative analysis of the experiment results. The preprocess typical corrections have been done before we apply the proposed method .A log ratio and Mean ratio operators are employed to generate difference images. A classical KI method is employed for comparative analysis. A Markov Random Field Fuzzy c-means clustering algorithm[10] (MRFFCM) being a progressive clustering algorithm is employed. By presenting the numerical results on the three datasets we will show the performance of the proposed method. The first data set represents an area near the city of Bern, Switzerland while the second represents city of Ottawa and third represents the region of Yellow river Estuary in China.

5.1. BERN DATASET:

The first dataset represents selection (301x301) pixels of two SAR images acquired by the European remote sensing 2 satellite SAR sensor over the city of Bern, Switzerland, in April and May 1999 respectively. Between the two dates, river is flooded in some parts of the cities of Thun and Bern and the airport of Bern entirely. Hence the valley between Bern and Thun was selected as a test site for detecting flooded areas. Here the MRFFCM, shows some changed regions are not detected (appearing as a low value of FN).FP of MRFFCM do not exhibit the best, whereas FN, PCC, KC which serves as an overall evaluation are best. It shows that it is less time consuming and time complexity. It does not engender a high time complexity but capable of improving detecting accuracy effectively

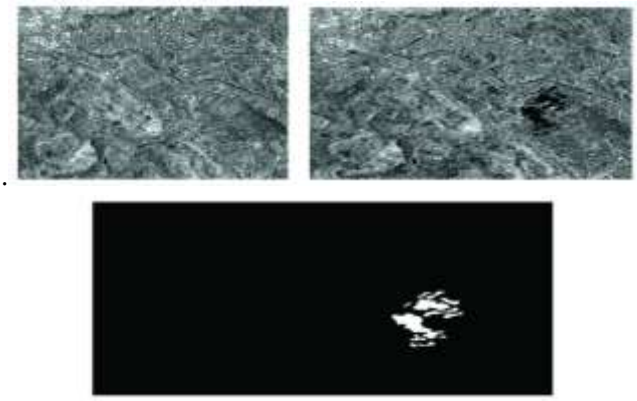


Figure 8.Bern dataset (a) Image acquired in April 1999 (b)Image acquired in May 1999(c) Ground Truth image

5.2.OTTAWA DATASET

The second dataset is a section of (290x350) pixels of two SAR images of the city Ottawa acquired by RADARSAT SARsensor. They were provided by defense research and development Canada(DRDC),Ottawa. It shows the image acquired in May 1997 during the summer flooding, shows the image acquired after the summer flooding.

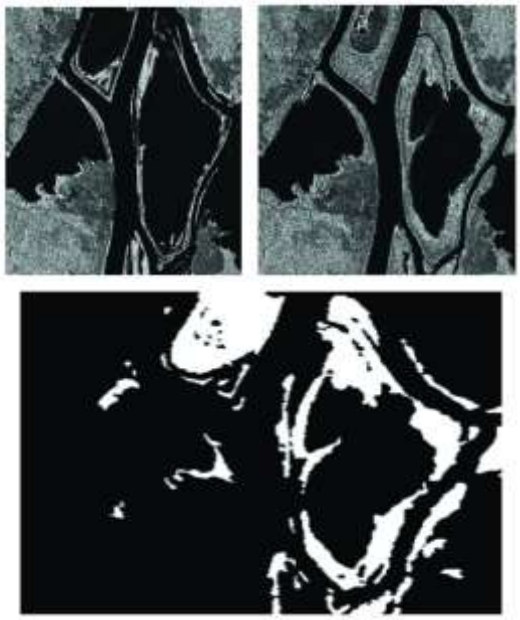


Figure 9. Ottawa dataset (a) Image acquired in May 1997 (b) Image acquired in August 1997(c) Ground Truth image

5.3. YELLOW RIVER DATASET

The third dataset comes from two SAR images acquired by RADARSAT-2 at the region of Yellow river Estuary in China in June 2008 and June 2009 respectively.The

original SAR images acquired by RADARSAT-2 are shown and with the size of 7666x7692.They are too huge to detail the information for our purposes, so only a small typical area of size 306x291 pixels is selected to compare the change detection results of different approaches.The final map generated by FCM lists those high values of FP. The white spots are eliminated to different degrees by modifying FCM.

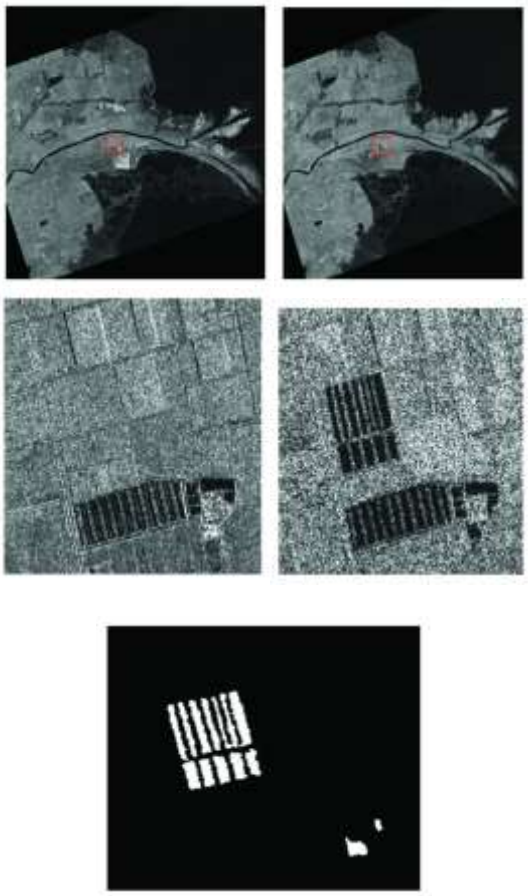


Figure. 10.Yellow river dataset (a) Original Image acquired in 2008 (b) Original Image acquired in 2009 (c) Selected area in (a) (d) Selected area in (b) .(c) Ground Truth image of the selected area.

The following measures are computed for change detection results.True positive (TP) is the number of pixels that are detected as the changed area both in the reference image and in the change detection result. True negative (TN) is the number of pixels that are detected as the unchanged area both in the reference image and in the change detection result.False negative (FN) is the number of pixels that are detected as changed area in the reference image and as unchanged area in the change detection result. False positive (FP) is number of pixels that are detected as unchanged area in the reference image and as changed area in the change detection result.Overall error (OE) is the total number of decision errorswhich equals the sum of FN and FP. The percentage correct classification (PCC) is calculated as follows:

PCC =TP + TN
TP + TN + FP + FN.

Table 1. PCC calculation

Datasets	*FP*	*FN*	*OE*	*PCC*	*KC*	*T/s*
BERN	2615	6	2621	0.99	1	56.8
OTTAWA	2248	113	2361	0.99099	1	71.0
YELLOW RIVER	2617	4	2621	0.99	1	66.1

In addition, Kappa statistic taking into account of commissionand omission errors is employed for evaluating theperformance comparison of the methods. Kappa statisticscalculated as follows:

Kappa =PCC –PRE/1 –PRE

where

PRE = ((TP + FP)(TP + FN) + (FN + TN)(TN + FP))/(TP + TN + FP + TN)2.

If the change detection result and the reference image are in full agreement, the Kappa value is 1. If there is no agreement between the change detection result and the reference image, the Kappa value is 0.

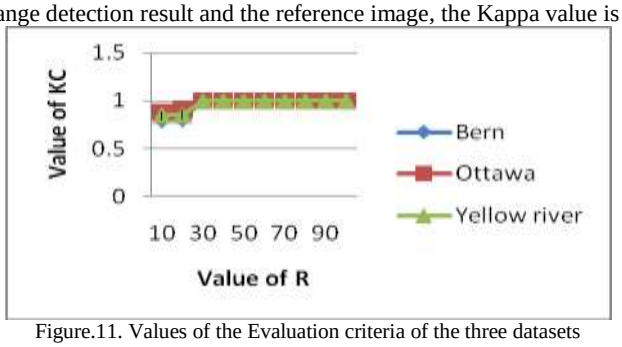


Figure.11. Values of the Evaluation criteria of the three datasets

The testing curve of the parameter R on the three datasets is employed.

6. CONCLUSION

The analysis of multitemporal SAR images has been done. This approach is based on the universal utilized FCM algorithm and the MRF model. After the difference image has been generated through the log-ratio operator, the MRF method is noted in the procedure of FCM algorithm. The energy function is altered by utilizing not only the membership but also the number of same class of neighborhood pixels. Thus we are able to decide whether the central pixels are in the homogenous region or in the heterogeneous region.

The new approach does not consider the use of any prior knowledge about the scene but it considers only the use of gray-level intensity. So it is an unsupervised approach. The main advantages of change detection approach is reducing the speckle noise, making the computations are simpler and it has low complexity. To compute the SAR images by the technique of image fusion, the image is normally converted from discrete waveform to Stationary wavelet transform (SWT).This technique is achieved and efficiency is also increased.

7.REFERENCES

[1] M. Gong, Z. Zhou, and J. Ma, “Change detection in synthetic aperture radar images based on image fusion and fuzzy

clustering," *IEEE Trans.Image Process.*, vol. 21, no. 4, pp. 2141–2151, Apr. 2012.

[2]S. Krinidis and V. Chatzis, "A robust fuzzy local information C-meansclustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5,pp. 1328–1337, May 2010.

[3]X. Zhang, H. Li, and C. Qi, "Spatially constrained fuzzy-clustering-basedsensor placement for spatiotemporal fuzzy-control system," *IEEE Trans.Fuzzy Syst.*, vol. 18, no. 5, pp. 946–957, Oct. 2010.

[4]D. M. Tsai and S. C. Lai, "Independent component analysis-based background subtraction for indoor surveillance," *IEEE Trans. Image Process.*,vol. 18, no. 1, pp. 158–167, Jan. 2009.

[5]A. Robin, L. Moisan, and S. Hegarat-Mascle, "Ana-contrario approach robust to registration noise for change detection in VHR multispectral images," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1877–1889, Jul.2010.

[6]S. S. Ho and H.Wechsler, "Amartingale framework for detecting changes for subpixel change detection in satellite imagery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1977–1993, Nov. 2010.

[7]S. Marchesi, F. Bovolo, and L. Bruzzone, "A context-sensitive technique in data streams by testing exchangeability," *IEEE Trans. Pattern Anal.Mach.Intell.*, vol. 32, no. 12, pp. 2113–2127, Dec. 2010.

[8]H. C. Huang, Y. Y. Chuang, and C. S. Chen, "Multiple kernel fuzzy clustering,"*IEEE Trans. Fuzzy Syst.*, vol. 20, no. 1, pp. 120–134, Feb. 2012.

[9]F. Chatelain, J.-Y. Tourneret, and J. Inglada, "Change detection in multisensory SAR images using bivariate gamma distributions," *IEEE Trans.Image Process.*, vol. 17, no. 3, pp. 249–258, Mar. 2008.

[10]W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy C-means clustering algorithms incorporating local information for image segmentation," *Pattern Recog.*, vol. 40, no. 3, pp. 825–838, Mar. 2007.

[11]Ashish Ghosh ,Niladri Shekhar Mishra , Susmita Ghosh , "Fuzzy clustering algorithms for unsupervised change detection in remote sensing images", *journal paper* in *information sciences* .

[12]Ferdinando Di Martino, Vincenzo Loia , Salvatore Sessa ,"Direct and Inverse Fuzzy Transforms for Coding/Decoding Color Images in YUV Space" , *Journal of Uncertain Systems* Vol.3, No.1, pp.11-30, 2009.

[13] Nguyen XuanVinh ,Julien Epps and James Bailey, "Information Theoretic Measures for Clusterings Comparison:Variants, Properties Normalization and Correction for Chance",*Journal of Machine Learning Research 11 (2010)*

Ms.S.Venkata Lakshmi is currently working as a Assistant Professor Grade 1, Computer Science and Engineering, at Panimalar Institute of Technology, Poonamalle, Chennai 600062.She has completed her graduate in M.S.University and Post graduate in DR.MGR university where she is currently working toward the Ph.D degree. She has published research papers in International, National conference proceedings. Her research interest includes image processing and medical imaging.

Mr.K.Sathyamoorthy is currently working as a Assistant Professor, Computer Science and Engineering, at Panimalar Institute of Technology, Poonamalle, Chennai 600062. He has completed his graduate in Anna University and Post graduate in Sathyabama University. His researchinterest includes image processing and soft computing.

Mr. T.K Senthilkumar is currently working as a Assistant Professor, Electronic and Communication Engineering at Rajalakshmi Institute of Technology, Poonamalle, Chennai 600124.He has completed his graduate and Post graduate in Anna University. He is currently working toward the Ph.D degree. He has published research papers in International, National journals. His research interests includeimage / signal processing and VLSI.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 220 - 225, 2015, ISSN:- 2319–8656

A Case Study upon Non-functional Requirements of Online Banking System

Harmeet Kaur
Punjab Technical University
Jalandhar, Punjab, India

Shahanawaj Ahamad
DCScSWE, College of
Computer Sc. & Engineering
University of Ha'il, K.S.A.

Gurvinder N. Verma
Shri Sukhmani Institute of
Engineering & Technology
Derabassi, Punjab, India

Abstract: The proper working of online banking is essential for development and advancement over the world and influencing organizations, society and individuals. Online banking, a vital activity within the financial markets is experiencing real change fuelled by advances in information systems techniques. Non - functional requirements (NFRs) are important software requirements that have been evaluated and specified right from the beginning of software development cycle while specifying the functional software requirements (FRs). Nfrs could be thought of as additional requirements that must be fulfilled by the functional requirements. The dissatisfaction of Nfrs is one of the indispensable purposes behind the failure of software projects. Non- functional requirements for instance accuracy, usability, performance and security are often crucial to online banking system. Accordingly the nonfunctional requirements or Nfrs should be given attention as early as possible in a software lifecycle and must be shown in the software design before conferring to the detailed design. The inspiration driving this paper is to examine how the assessment of Nfrs of online banking system serves to systematically make choice among different alternatives.

Keywords – Reengineering, Legacy, Non-functional Requirement, Requirement engineering, Usability.

1. INTRODUCTION

The success of a software system can be accessed on the basis of its fulfillment of two interdependent types of requirements. First and foremost, the client focused requirements that focus on the functionalities given to the users of the system. Second, the context focused requirements which can be system, procedure and human necessities. The later type of requirements is suggested as non-functional requirements (Nfrs). Nfrs are systems goals that may impact the operational environment and design decisions an employer may look for the progression of the software. Besides other things Nfrs incorporate operational environment: hardware and software interfaces, accuracy, performance etc. Functional requirements portray the noticeable external input and output interactions with the system under scrutiny, despite the fact that non-functional requirements are those that drive unique conditions and qualities on the system to develop. Thus, system acceptance testing is centered on both functional and non-functional system's requirements.

The NFR has various definitions in the literature and the industry. [19] Characterizes it as "Umbrella term to cover all those requirements which are not explicitly defined as functional". Nfrs can be grouped either using intra model dependency view or intermodal dependency view. In the intra model dependency view, Nfrs are refined into a hierarchy comprising of a root NFR classification and various child refinements, for example, decomposition and operationalization [19].The disintegration is a unique procedure where a NFR is portrayed using child sub-Nfrs [19]. For example, the security NFR can be disintegrated into smaller sub-Nfrs so that it can be addressed in a better way. The operationalization is an extraordinary methodology where a NFR is refined into operations, functions, information representations and architecture design decisions that are important to address the NFR effectively.

2. NON FUNCTIONAL REQUIREMENTS

In competing situations time-to-market, robustness, and quality are some of the components which are critical for measuring the success of system. As the business requirements are changing rapidly there is a need to change and develop. In the system development process requirement specification is the important and significant step to be considered. Subsequently, there is a strong need to formulate the methodologies and standards by which one can capture the requirements. In addition to providing functional requirements or services the software system should posses the nonfunctional requirements as they are important for considering the quality of the software and amazingly impact the design and execution decisions a developer may make. These also have an impact on the acceptability of the software by the intended clients or users.

3. THE NATURE OF NFRS

There are different clarifications why non- functionality is complicated to handle. To appreciate why, it has to research what a NFR is and the distinctions that exist contrasted with functional requirements. A NFR may be seen as how the system act as per specific attributes, for instance performance, modifiability and security. A functional requirement also is seen as what the system must be capable of performing. Functional requirement are particular entities that are not difficult to understand

and administer. In fact they are easy to verify because of their temperament, which infers that it is not difficult to give a boolean answer whether the requirement is satisfied i.e. "Is it conceivable to upload a document?" Concerning NFR it is not that irrelevant to ask essential inquiries and expect a boolean answer that shows if the requirement is fulfilled or not, this is a due to nature of these requirements. NFR could be seen as subjective, since different people decipher and evaluate the Nfrs contrastingly. This is exasperates even increasingly because of how small and uncertain NFR for the most part are depicted in software requirement specification.

From time to time in distinctive organizations the systems are built according to the need, there for the explanation and criticalness of the various Nfrs changes from system to system. Which is the motivation behind why Nfrs can be seen as relative? Not precisely how Nfrs are translated or how important they are, could be seen as relative but how the requirements are attained. May be existing techniques could be created or upgraded to acknowledge particular Nfrs, there for the methodology "one solution fits all" is not suitable. Moreover Nfrs can be communicating, which implies that when understanding one requirement this could help solving or overturn an alternate one. In the light of worldwide effect of the aforementioned requirements local solutions for the NFR is often not sufficient. For example, performance centered security. Assume that the system will be made secure at the cost of performance.

4. WHY NON-FUNCTIONAL REQUIREMENTS

Functional requirements exhibit the behavior of the system that fulfills client's objectives or tasks whereas non- functional requirements consolidate constraints and qualities. Qualities are properties of the system that are examined by its and in this manner will affect their level of satisfaction with the system. Constraints are not subject to negotiation and, not at all like qualities, are (theoretically at any rate) off-limits all through design trade-offs. Contextual constraints are characteristics of the "super-system " or the advancement affiliation that oblige the change somehow. Examples join the target working system or fittings stage by virtue of the customer environment, or the inclination set of open visionaries by virtue of the improvement association. Nfrs are in general called as quality prerequisites [4] [2] the business asks for the more non-viable perspectives to be satisfied in information schemas other than its usefulness. Later works proposed by [12] [21] [7] have demonstrated that complex reasonable models must administer nonfunctional requirements. Mistakes in light of oversight of Nfrs or to not suitably administering them are around the most costly and generally hard to reexamine [21] [13] [9]. Non-functional requirements are always related to a viable need [7].

5. ROLE OF NON-FUNCTIONAL REQUIREMENTS IN A SYSTEM

There is much verbal confrontation about the part non-functional requirements play in the requirement elicitation and system advancement process. Systems and software designers succumb to two camps concerning this inquiry, with the line of division being the "qualitative" nature of non functional requirements A Qualitative requirement is equal to an enormous need, which is something that must be avoided at any cost. Recently, [4],[14],[16] HAVE SHOWN THAT there has been a push to make non- functional requirements more quantitative in nature. For example, one requirement may be: "The system might allow anonymous login for unlimited functions." A second need expresses: "The system may have defensive firewalls to affirm fluctuating levels of access." Implicitly, the stakeholder is getting two non-functional requirements: Ease of use and security. By and large, when seen uninhibitedly, non- functional requirements can be addressed routinely. On the other hand, there are times, (for example the one above) when they show an accident which begins ripple affect felt by distinctive requirements– trying to fulfill one need. The developer is sacrificing another requirement.

6. NON-FUNCTIONAL REQUIREMENTS IGNORED

Because of less consideration paid to a various non-functional requirements: interoperability and testability NASA's Mars Climate Orbiter and Polar Lander were lost in Mars environment late in 1999.As far as the interoperability is concerned, the orbiter was delivered by different organization: Lockheed Martin Aeronautics (main contractor) and the Jet Propulsion Laboratory (JPL). In transforming push for the unit, Lockheed Martin researchers utilized English estimations, while JPL experts utilized the metric system. This oversight was recently perceived after the test was conveyed, a definite explanation behind the disappointment. Moreover, the modules that were made for the Polar Lander were not as brief as first considered. All through system testing, one module was denied (there was a known issue with that module) and tests were run.

At a later time, the banned module was tried autonomously and final conveyance was made. It was not until the system failed in space that scientists recognized that the exclusion of the module all through system testing (they simply ran one full test), influenced the usefulness of the landing gear. These two missteps cost United States citizens $319m.Another example which alludes to the absence of attention to non-functional requirements can impact framework unfavorably. Among distinctive clarifications and as various non functional requirements were ignored all through the system advancement stage the London ambulance system was deactivated after its deployment. Some of the non functional requirements which were disregarded during the deployment of the ambulance

are reliability (vehicles location), cost (stress on the best value), usability (poor control of data on the screen), performance (the system did what should do however the performance was unsatisfactory), [15] [3]

7. THEORATICAL NFR FRAMEWORK

The NFR Framework offers a systematic methodology for characterizing NFRs. It offers extraordinary capacity to see for all significant NFRs and their interdependencies. In addition, it likewise helps software designers to fathom the vital activities for guaranteeing quality. It even captures and documents design decisions and rationales in addition to providing traceability for derived specifications and requirements

The NFR Framework offers a precise approach for describing Nfrs. It offers remarkable ability to see for all noteworthy Nfrs and their interdependencies. Likewise, it similarly helps software designers to understand the essential activities for ensuring quality. It even catches and reports design decisions and methods of reasoning in addition to providing traceability for derived specifications and requirements.

7.1 Security

Security is the feature of the system which ensures that system must be protected from the unintentional or malignant harm; unauthorized access to the data is not permissible. For the safety purpose the data must be backed up after certain period of time say 24 hours and the backed up data must be stored in a secure location. In online banking system the application must be able to send or receive the information to or from the server and client in an encrypted way. Security must stick to some standard and plans.

The security is significant subject of online banking as client is more worried about the security of the account, personal data and transactions. The information kept in the system ought to be precise and complete.

7.2 Performance

The term performance alludes to the capacity of the system or software to process as many as transactions per second as submitted to it without failure. Despite the fact that the system is functional and reliable if it fails to make efficient use of resources such as CPU cycles, disk space etc its performance is not good that is the performance of the system is not up to the mark. Performance measure how well the system can perform and whether the software will have the capacity to reach its response time targets. Performance additionally measures that how effectively the software will have the capacity to scale with countless activities for every second, moment, or hour. The online banking system is a multi- user system, which implies distinctive clients can access the system simultaneously and the system will work accurately and proficiently. So the client is more worried about the performance of the online banking.

The term performance suggests the ability of the system or software to process the same number of transactions every second as submitted to it without failure. Performance is measured in term of how well the system makes optimized or maximum use of the resources without failure. It measure how well the system can perform and whether the software will have the ability to achieve its response time targets. Performance moreover measures that how successfully the software will have the ability to scale with numerous activities submitted to it in every second, minute, or hour.

7.3 Usability

As online banking is carried by various types of clients i.e. whether they have knowledge of computers or not so the application designed for online baking must be easy to use and enable the client to manage their accounts or transactions with simplicity. The application must have graphical user interface and it must have the ability to provide informative error messages. The qualities of the ease of use which can be measured are learning time it points out the time required to learn the application, number of errors while working with the normal speed furthermore the likeness of the client to measure the system i.e. the client fulfillment in utilizing the framework. The interfaces of the system ought to be clear, easy and simple to use and understand. To expand the ease of use, on-line help and customer care executive ought to be incorporated into the system to encourage online banking.

7.4 Availability

The online banking should be available round the clock. It means for how long the system is available for its users or clients and for how long the system will be operational. As far as the online banking system is concerned availability of the system is of uttermost importance as the business is round the clients and clients should be able to avail the benefits of online baking without any constraints and round the availability makes it happen and thus it is of importance for the bank and the clients as well. The online banking system must have the availability of 99% if not hundred than the availability of the system must be nearly 100%.

7.5 Confidentiality

Client should be able to access the online banking account after successful authentication. The data entered by the client is not accessible to other clients using online banking. As far as the confidentiality of the is concerned it means to maintain the secrecy as online banking is round the clock to access the account i.e. any where any time so it is important that the software of the online baking must provide the facility to maintain the secrecy of the clients and the clients should have their own passwords and user names and these must be automatically become inaccessible to the person who so ever wants to have the unauthorized access to the account.

7.6 Reliability

Reliability reflects the capacity of the software to maintain its performance over the time. It implies how well the system performs in peak hours. A robust system is one which has the capacity to handle the bugs without failure i.e. how effortlessly it handles the bugs

because of data or handling, surprising conditions while working conditions furthermore the software imperfections, and if the system is robust it is reliable also. The application must self contained, consistent and complete with in itself. The failure rate in the online banking system should be least or negligible as the system is supposed to be reliable. Reliability of the system depends on the failure free transactions and how fast the system is able to recover from the failure.

7.7 Operability

It is concerned with the how well the software will work in distinctive environment. The online banking application must have the capacity to operate on any gadget i.e. hand held gadgets or desktops or laptops without failure the change in the environment ought not to hamper the operability of the software. The online banking system ought to be operational on any hand held gadgets and it ought to have the capacity to show the different types of currencies used in distinctive nations. While in operation the system ought to give the consistency of the operation. It is supposed to give ease of operation and controls by the client. The online banking system ought to be operational on any hand held gadgets and it ought to have the capacity to show the different types of currencies used in distinctive nations. While in operation the system ought to give the consistency of the operation. It is supposed to give ease of operation and controls by the client.

7.8 Traceability

Traceability refers to the capability for tracing the status of the transaction and account on account number basis. Traceability is an important aspect in the banking industry, where it makes tracking of transaction possible. It should contribute to the safety of the transaction. Online banking should enable the user to trace the state of his/her transaction at any time.

7.9 Recoverability

Recoverability implies the ability to restore your software to the point when failure occurred. The ability to recoup quickly from a system failure depends not simply on having backup of the data, also on having a predefined plan for recuperating that information. in case of online banking system it is obliged that the software ought to be tested to check it will have the capacity to recuperate from the failures or not to do so the software is subject to failures via doing deeds which prompts its failures e.g. restarting the machine when the online banking application is running, or it is getting information and so on. Software can get back from failures at the very instance when the failures occur. The online banking must have the capacity to recover itself from the failures which may happen due to internet problem or because of some other reason. The online banking ought to have the capacity to recover the influenced information after failure and must have the ability to restore the sought level of execution or performance.

7.10 Visibility

It alludes to condition of having the capacity to see online banking empowers the client to see the login screen and the configuration of the online banking application as per the client desire. As the customer is dealing with his or her transactions or accounts online the application must be composed in such a way, t that every single feature of the application is obvious to the client as the client logins. The different features of the application must be self contained so that even a novice can utilize it without hesitation. All the features of the application must be visible and easy to understand and use.

8. RESEARCH METHODOLOGY

8.1 Selecting Samples

The main objective of this study was concentrated on the role of non functional requirements to assess the quality for reengineering. The study was intended to find the importance of non functional requirements like security, performance, usability availability, operability etc in online banking system. Therefore, the sample for this study was selected from the users of online banking.

8.2 Data Collection Procedures

Data was assembled from the clients using online banking. A set of questionnaire was given to online banking users. The first part of the questionnaire comprises of the general data of the respondent. Non functional requirements were utilized in second part. The last part comprises of online banking user's recommendations if any. While filling up the questionnaire, each part of the survey was clarified to the respondents.

8.3 Research Design

A sample of 122 online banking customers was drawn, however 65 respondents could manage to provide the requisite information which was further utilized to find out the perception of users how and which out of 12 non functional requirements are of importance for them while dealing with online banking.

9. RESULTS AND DISCUSSION

While measuring the information with the help of statistical tools, to be specific Total Weightage Score system (TWS), it was observed that 67% of target respondents were male and 33% were females, of which most of the respondents were lying in the age group of 20-30 yrs i.e. 41.54%, followed by age of 40-50 yrs i.e. 30.77% and 12.30% were lying between the age of 30-40 yrs while the rate of respondents in the age gathering of 50-60 yrs and 60-70 yrs was discovered to be less i.e. 9.24 and 6.15% respectively. The information gathered was analyzed and total weightage score was computed and from the observation of the analysis it was discovered that in the online banking system client is more worried about the security followed by performance, usability and availability of the online banking services. From the graph indicated in Fig.

reliability, visibility and confidentiality are closely related with total weightage score ranging from 252,240 and 233.The total score for operability and accuracy was discovered to be 198 and 165 respectively though the total weightage score for portability, recoverability and traceability varies between 135 and 126.

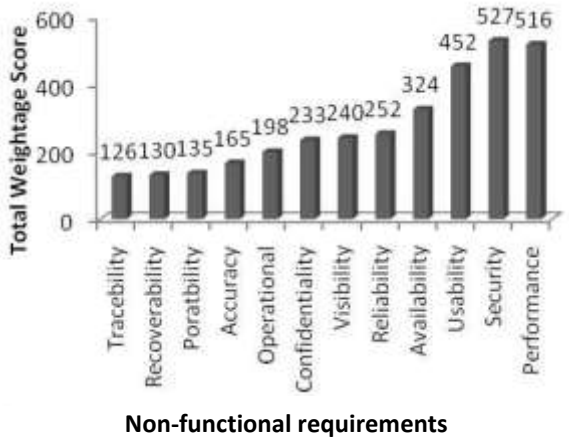


Figure 1: Weightage of NFRs in online banking system

During the study it was observed that security is of utmost importance amongst the other non functional requirements. Security is of concern for the user of the online banking system and also for the success of the online banking system as client or user is more worried about the security issues followed by performance which is affected by the speed and response of the connection over which the user is working. To have hassle free operations user is concerned about the usability and availability of the online banking system. For proper functioning the system must be reliable so another non functional requirement of user concern is reliability where as portability, recoverability, traceability, accuracy, usability and confidentiality makes the online banking system easy to use and user friendly

10. CONCLUSION AND FUTURE SCOPE

In this paper, non functional requirement for online banking system has been exhibited. The non-functional requirement corresponds to the quality of the system but in online banking system Nfrs are considered as requirements of paramount importance for the system. The security is one of the non functional requirements which is thought to be of absolute vitality and is considered as one of the primary prerequisites that determine the success or failure of the system. Performance and usability are also of importance as far as the proper working and running of the system is concerned. It has been proposed that while considering the quality requirements of the system the non functional requirements ought not to be ignored for proper working of the system. This work would help the software engineering community and to the researcher in enhancing the notion of Non functional requirements for online banking system and furthermore inspire them to come up with pre-specified benchmarks of Nfrs. The present work will be further extended as per the interest of the research problem and will be put to statistical investigation for increasing accuracy of the work and to improve the precision later on.

11. REFERENCES

[1] Basili, V. and Rombach, H., "Tailoring the software process to project goals and environments", Proceedings of the 9th International Conference on Software Engineering (ICSE-9), IEEE Computer Society Press, pp. 345-357 March, 1987.

[2] Boehm, Barry e In, Hoh., " Identifying Quality-Requirement Conflicts", IEEE Software, pp. 25-35, March 1996.

[3] Breitman,K. K., Leite J.C.S.P, Finkelstein Anthony, "The World's Stage: A Survey on Requirements Engineering using a Real-Life Case Study", Journal of the Brazilian Computer Society, No. 1, Vol. 6, pp.13-37, Jul. 1999.

[4] Chung, K. L., "Representing and Using Non-functional Requirements for Information System Development: A Process Oriented Approach", Ph.D. Thesis, also Tech. Rpt. DKBS-TR-93-1, Department of Computer Science, University of Toronto, June 1993.

[5] Chung L, Nixon B, Yu E. Using non-functional requirements to systematically select among alternatives in architectural design In: Proceedings of the first international workshop on architecture for software systems, Seattle, WA, 1995.

[6] Chung L, Nixon B, Yu E. Dealing with change: an approach using non-functional requirements. Requirements Eng pp.238–260,1996.

[7] Chung, Lawrence et al., Non-Functional Requirements in Software Engineering Boston: Kluwer Academic Press 2000.

[8] Cysneiros, L.N., Leite, J.C.S.P, "Driving NFR to Use Cases and Scenarios", XV Brazilian Symposium on Software Engineering, 2001.

[9] Cysneiros L.N., Leite J.C.S.P., "Integrating non-functional requirements into data model", In: Proceedings of the 4th international symposium on requirements engineering – Ireland. IEEE Computer Society Press, Los Alamitos, pp. 162–171, 1999.

[10] Daniel, E., "Provision of electronic banking in the UK and the Republic of Ireland", International Journal of Bank Marketing, 17(2), pp.72-82, 1999.

[11] D. Arnold, et al., "Scenario-Based validation: Beyond the user requirements notation", in Australian Software Engineering Conference (ASWEC), Los Alamitos, CA, USA, pp. 75- .84, 2010.

[12] Dardenne A., van Lamsweerde A, Fickas, S., "Goal Directed Requirements Acquisition", Science of Computer Programming, Vol. 20, pp: 3-50, 1993.

[13] Ebert, Christof, "Dealing with Nonfunctional in Large Software Systems", Annals of Software Engineering, 3, pp. 367-395,1997.

[14] Franch, X., "Systematic Formulation of Non-Functional Characteristics of Software", Proc. 3rd Int. Conf. on Requirements Engineering, Colorado Springs, Colorado, USA, 1998.

[15] Finkelstein, A. and Dowell J., "A comedy of Errors: The London Ambulance Service Case Study", Proceedings of the Eighth International Workshop on Software

Specification and Design, IEEE Computer Society Press pp. 2-5, 1996.

[16] Jacobs, Stephan, "Introducing Measurable Quality Requirements: A Case Study", Fourth IEEE International Symposium on Requirements Engineering, 1999.

[17] J. H. Hill, et al., "Unit Testing Non-functional Concerns of Component-based Distributed Systems", presented at the Proceedings of the 2009 International Conference on Software Testing Verification and Validation, 2009.

[18] Jureta, I., S. Faulkner, P.-Y. Schobbens, "A More Expressive Softgoal Conceptualization for Quality Requirements Analysis", Proc. of IEEE Int. Conf. on Conceptual Modelling (RE06), pp.281-295.

[19] Kassab , M , " Formal and Quantitative Approach to Non Functional Requirements Modeling and Assessment in Software Engineering", Doctoral dissertation Concordia University , Montreal, Canada,2009.

[20] K. Saleh and A. Al-Zarouni, "Capturing non-functional requirements using the user requirement notatio", Proceedings of the International Research Conference on Innovations in Information Technology (IIT 2004), Dubai, pp. 222-230, 2004.

[21] Mylopoulos J, Chung L, Yu E, Nixon B., "Representing and using non-functional requirements: a process-oriented approach", IEEE Trans Software Eng, 18(6), pp.483–497, 1992.

[22] Mylopoulos J, Chung L, Yu E., "From object-oriented to goal oriented requirements analysis", Commun. ACM, 42(1), pp.31–37, 1999.

[23] Paech, B. & Kerkow, D., "Non-functional requirements engineering – quality is essential", 10th International Workshop on Requirements Engineering – Foundation for Software Quality, pp. 237-250, 2004.

[24] Q. Tran, "A CASE Tool for the Non-Functional Requirements Framework", M.S. Thesis, Dept. of Computer Science, Univ. of Texas Dallas, 1998.

[25] R. Sprague, "A framework for the development of decision support systems", MIS Quarterly, Vol.4, No. 4, 1980.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 226 - 228, 2015, ISSN:- 2319–8656

Position Tracking and Fuzzy Logic

Mohd. Asim
MGM CoET Noida
Uttar Pradesh, India

Riya Malik
MGM CoET, Noida
Uttar Pradesh, India

Shagun Bhatt
MGM CoET, Noida
Uttar Pradesh, India

Abstract: In this paper, the problem of tracking the position of an object from a remote place is solved by applying the rules of fuzzy logic. The domain of image analysis and artificial intelligence has been integrated so as to track the position of the object and get results that are more precise. The object has been identified and then tracked until it has exceeded the range of a camera. Various techniques have been used at various levels of the entire process. Fuzzy rules have been made so as to accomplish the goal. A video has been first converted into frames and then using technique of background subtraction, the object has been identified and then tracked which is an optimal solution as well as precise.

Keywords : Fuzzy Logic, Position Tracking, Fuzzy Inference System.

1. INTRODUCTION

Position Tracking using Fuzzy Logic is the leading field in the world of research. Today lots of research is being done in the field of Tracking the position of a remote object. The need to track position of a person or an object can be due to various reasons and so are the procedures. Numerous number of procedures has been developed in order to fulfil this need. Technologies and software are also present but it is our will in order to use which way of solving the problem of position tracking.

Position Tracking*:* In this, first a video has been captured located in an area and then that video has been converted into frames using functions of matlab. The whole process has been divided into three steps:

First is the step of object identification in which the technique of background subtraction has been used. In this, first the intensity values of all the pixels are calculated and by background modelling the pixels intensities are arranged. A threshold value is set above which all the values are considered to be of the object of interest and rest are considered to be the pixels of the background. In this way, the object is identified and background subtraction is performed in order to get the desired object. Second step is to correctly classify the different objects which have been identified in the background according to the rules. The system is made to learn some rules through the process of supervised learning. The third and the final step is to precisely track the position of the object. This is the most crucial step and involves very close analysis of the step to step movement of the object. The whole procedure of doing this process is being explained in this paper.

2. LITRATURE SURVEY

Position Tracking using fuzzy logic has many implementations among which the first that we came across was a Robot Soccer System, consisting of Robots and a Soccer ball. In this the Robots are divided into two teams and they both are playing a match. All the robots in this game are designed in such a way that each one of them will initially detect the position of the ball and then try to fetch a goal. The ball will be detected by a camera placed on the head of the robot. This camera once captures the image of the ball then detects the angle between opponent player ball and if there is no interruption then the ball is kicked a goal is fetched [1].

The second paper that we studied was on how can vehicles track an obstacle in front if it and then automatically takes decision based on the situation. This paper tells us about an Automatic Breaking System used in the cars which can detect the obstacles or object present in front of it. This was done using the condition that is the object present in front makes an angle less than 60 degrees then the object is ignored and if it I more than 60 degrees then the automatic breaking system gets ON and breaks are applied. This paper tells us how can objects be detected in front and then according to the situation occurring the decision is taken[2].

The methodology of Colour Based Segmentation is another such technique used in the field of position tracking this is applied by recognising the colour of the target and then throughout the process matching it and keeping a track of it, so as to track the displacement of an objectfrom initial position to final one. This method is simple yet efficient as only the colour conversion method is used and tracking is done in the whole process. Since the process needs to track the path so it also uses the Fuzzy Logic Controller to determine the position of the object and decline fake and unnecessary movements and track the right path [3].

Tracking can also be done by simple colour conversion technique that is conversion of RGB value to its corresponding HSB value. This is done by simple image processing techniques already present in the market. Blobs are also used as a way of calling the objects and images in a converted scenario of objects. This method is also simple the main work is done in background subtraction and then converting the colours of the images. The second work that is done is the fuzzy rules that need to be defined in such a way that it does not create a problem later on [4].

While researching over position tracking we also came across the system of DINDS. In this process the object to be tracked

was observed by many cameras at a single time as the object moves from one place to another. The idea of taking so many cameras at one go was so that no single movement was missed by the system because that can cause error in tracking position. In such systems fuzzy logic is used in handing over the tracking ability from one camera embedded system to another. It gives an interesting technique of synchronisation of the system which there is a lot of movement going on [5].

3. PROPOSED TECHNIQUE

Various researches have been done in the field of position tracking. The technique proposed here firstly takes a video. The video can be of any object which is in motion but our system will only work in case of tracking of human motion. The whole process consists of four steps:

1) Identification of the object
2) Classification of the object
3) Tracking of object

3.1 Identification Technique

An image consists of a background as well as a foreground. For the precision in tracking, the object of interest has to be secluded from its background. There are various techniques using which the object can be isolated from its background but here in this paper the technique of background subtraction has been used. In this technique, a threshold value is set by image analysis. The intensities of pixels are calculated and a threshold value is set. If the current pixel value exceeds the value of the threshold then that pixel is considered as a foreground pixel else a background pixel which then set to gray level 0.

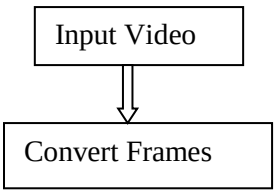
3.2 Classification Technique

In this step of object tracking, machine learning has been used so as to differentiate between various objects of foreground. The process of machine learning has been used in the process as features are stored and based on that object of interest have been chosen. Machine learning can be of following types:

1) Supervised Learning
2) Unsupervised Learning
3) Reinforcement Learning

In supervised learning, a teacher like process is present which tells the result whereas in unsupervised learning the system learns by its experience. In reinforcement learning, there is nothing like teacher but a critic tells whether the output is correct or wrong. It does not give any other information about the result. For our system, supervised learning has been chosen.

The whole process of tracking can be represented by the flowchart in Fig 1.



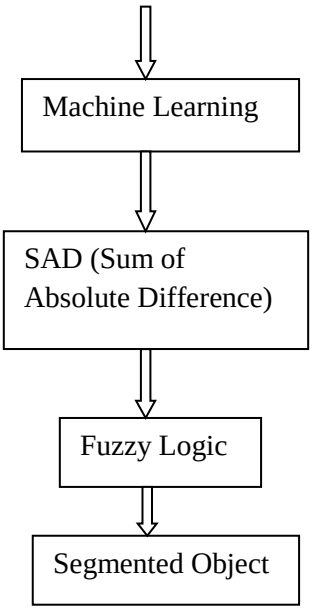


Fig 1.Flowchart showing Position Tracking.

3.3 Tracking Technique

At this level of process, the tracking of motion of a moving human object has been done. There are various techniques through which the motion of the object can be tracked. In the proposed system, the technique of SAD that is sum of absolute differences has been used. This calculates the inter-pixel difference between the two pixel locations of the two frames of same video. Then fuzzy rules are applied to the output which gives a fuzzy output which is converted into crisp value to detect the object motion.

3.4 Algorithm For identification and Tracking

Algorithm 1: Identification

```
Fr _ bw = rgb2gray (fr);

Fr _ diff = abs (double (Fr) bw) –double (bg_bw);

If   Fr_diff > thresh pixel in foreground

For j = 1: width

        For k=1: height

                If ((Fr_diff (k, j)>thresh))

                        fg (k, j) =  Fr_bw(k,j);

                        else

                                fg (k,j) = 0;

                        end

                end
```

end

end

Algorithm 2: Tracking

SAD = ∑(X-Y) for all pixels

Where X is the pixel of one frame and Y is the pixel of other frame.

4. RESULTS

Using the approach discussed above, the position of a lady in a video has been tracked. The following Fig.2 depicts a scene in which a lady has been shown walking and the respective movement has been shown by applying the first step of the process to the image which is the technique of background subtraction. All the pixels which are in the background has been set to zero and only the object of interest that is a walking lady has been shown for tracking. The above explained scenario can be seen in the following image:

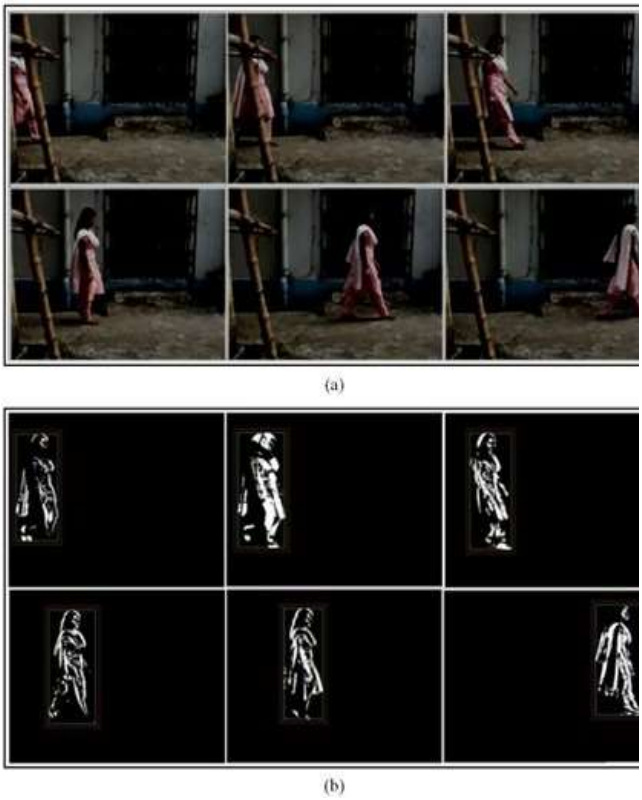


Fig 2. Frame Conversion and Background Subtraction

In the Fig.3, the technique of tracking described above has been applied to the two frames of an image and the resultant location with respect to time has been shown. The following image represents a graph showing the different movements of the lady with the values of x as well as y coordinates for four different scenarios. The above described algorithm has been used so as to obtain the results. This can be seen in the following figure:

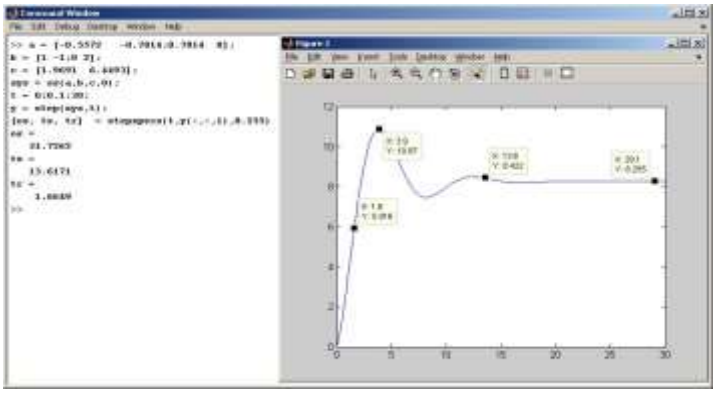


Fig 3.Tracked Path of the Lady.

5. CONCLUSION

This paper describes the process of Position Tracking using Fuzzy Logic by defining the various steps involved in doing this. This paper also gives the Algorithm used in this process. The algorithm developed clearly tells how the process is working and what all steps are used. The algorithm is divided into two parts identification and tracking. In this experiment we have used a video of a lady walking and its position is tracked.

6. REFERENCES

[1] Fuzzy Logic Path Planner and Motion Controller byEvolutionary Programming for Mobile Robots, Byung Cheol Min, Moon-Su Lee, and Donghan Kim.

[2] Neural Networks & Fuzzy Logic Elakkiya Prabha T,Pre-FinalB.Tech-IT,M.Kumarasamy College Of Engineering, Karur Kiruthika M, Pre-Final,B.Tech-IT,M.Kumarasamy College Of Engineering, Karur

[3] Colour-Based Object Tracking and Following for Mobile Service Robots, Mohamed Abdellatif, Dept. of Mechatronics and Robotics., Egypt-Japan University of Science and Technology, Alexandria, EgyptTavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[4] Fuzzy Rule-based Classification of Human Tracking and Segmentation using Colour SpaceConversion, Sivabalakrishnan.M1 andDr.D.Manjula21Research Scholar, Department of CSE, College of Engineering, Anna University,Chennai, Tamil Nadu, India, sbkrishnanm@gmail.com, 2Assistant Professor, Department of CSE, College of Engineering, Anna University, Chennai, Tamil Nadu, India.

[5] ADAPTIVE CAMERA SELECTION BASED ON FUZZY AUTOMATON FOR OBJECT TRACKING IN A MULTICAMERA SYSTEM Kazuyuki MORIOKA1, Szilvester KOVÁCS2, Péter KORONDI3,Joo-Ho LEE4, Hideki HASHIMOTO5, 1Meiji University in Tokyo, JAPAN,2University of Miskolc, 3Budapest University of Technology and Economics, HUNGARY, 4Ritsumeikan University in Kyoto, 5University of Tokyo, JAPAN

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 229 - 234, 2015, ISSN:- 2319–8656

Quest Trail: An Effective Approach for Construction of Personalized Search Engine

M. Therasa
Anna University
Panimalar Institute of Technology
Chennai, India

S.M. Poonkuzhali
Anna University
Panimalar Institute of Technology
Chennai, India

S. Hemamalini
Anna University
Panimalar Institute of Technology
Chennai, India

ABSTRACT: Personalized search refers to search experiences that are tailored specifically to an individual's interests by incorporating information about the individual beyond specific query provided. Especially people working in a software development organization (analysts, developers, testers, maintenance team members), find it increasingly difficult to get relevant results to their searches. We propose methods to personalize searches by resolving the ambiguity of query terms, and increase the relevance of search results in order to match the user's interests. Difficulty in web searches has given rise to the need for development of personalized search engines. Personalized search engines create user profiles to capture the users' personal preferences and as such identify the actual goal of the input query. Since users are usually reluctant to explicitly provide their preferences due to the extra manual effort involved, the search engine faces the entire burden of predicting the user's preferences and intentions behind a query in order to yield more relevant search results. In this paper we define a QUEST to be the objective of user's search; here we combine quest level analysis of user's search logs and semantic analysis of the user's query in order to personalize user's search results. Most personalization methods focus on the creation of one single profile for a user and apply the same profile to all of the user's queries. Hence we propose a personalized search for a software development organization by creating QUEST or domain based profile rather than individual user based profile.

Keywords: Metasearch, Content Based Filtering, Query Bundle, QB-C, Quest Trail

1. INTRODUCTION

Personalized search refers to search experiments that are tailored specifically to an individual's interests. It aims to resolve the ambiguity of query terms. To know more about the ambiguity that arises in search engines let us take the instance of "Java". When the user searches about Java there are three possibilities of results (i.e.) the results can be about Java Sea in Indonesia or about the Java coffee bean or the programming language. This is an example for ambiguity.

Difficulty in web searches has given rise to the need for development of personalised search engine. It is important to introduce personalization in a software organization where the employees are reluctant to provide information. There are two types of user behaviour (i.e.) search behaviour and browser behaviour. Search behavior [22] is everything the user enters in the search engine to search for the information needed. Browser behaviour involves surfing; user types a URL address in the browser, king a bookmark or forward page in the browser etc.

Searches can be analysed in three levels, (a) query level, (b) quest level and (c) session level. In query level it fails to capture the interleaving relationships between different quests [18]. If we analyse the search logs based on session (i.e. session level) [6] [11] [13][21] the quests will be interleaved. It is difficult to identify what the user is doing because the sessions are chronologically ordered.

If we analyse in quest level the topics will be more consistent and relevant to each other. This will help us to understand the intentions behind a user's search. Query is the search entry made by the user into the search engine (e.g.) the user types "jython versus swings" into the search box and searches. A Query Trail can be defined as sequence of user behaviour (a query followed by sequence of browsing behaviour) [14][16][28]. Quest (task) is an atomic information need (e.g.) the user needs to know what "jython" is? And compare its features with swings in "java". A quest trail represents all user activities within that particular task, such as query reformulations, URL clicks [17]. Sessionis defined as "a series of queries by a single user made within a small range of time" and the activities done by the user in that time period in a browser is known as session trail [23].

Consider the example shown in Table 1, which is a real user search session from Google (http://www.google.com). This session contains 4 different search quest: Twitter, Flipkart Kindle Books, Yahoo, and lyrics of a song. The "Yahoo" task is interleaved with the "Flipkart Kindle Books" task. The reasons causing the interleave phenomenon [18] are: (1) web search logs are ordered chronologically; (2) users often open several tabs or browsers and conduct multiple tasks at the same time.

Table1: A sample session from web search logs.

Time	Event	Value	QUEST
09:03:26	Query	Twitter	1
09:03:39	Click	www.twitter.com	1
09:06:34	Query	Flipkart	2
09:07:48	Query	Twitter	1
09:08:02	Click	twitter.com/login.php	1
09:10:23	Query	flipkart kindle	2
09:10:31	Click	kindle.flipkart.com	2
09:13:13	Query	yahoo log in	3
09:13:19	Click	mail.yahoo.com/mail	3
09:15:39	Query	flipkart kindle books	2
09:15:47	Click	flipkart.com/Kindle-eBooks...	2
09:15:59	Click	astore.flipkart.com/ Flipkart..	2
09:17:51	Query	You belong to me	4
09:18:54	Query	You belong to me lyrics	4
09:19:28	Query	Belong to me lyrics	4

In this paper we bring in two studies semantic analysis and genetic algorithm for personalizing the search process in the search engine. To get a clear picture of our study we also discuss about Meta search engine, personalization and search.

1.1 Search Engine

A search engine is a type of computer software used to search data in the form of text or a database for specified information. Search engines normally consist of spiders (also known as bots) which roam the web searching for links and keywords. They send collected data back to the indexing software which categorizes and adds the links to databases with their related keywords. When you specify a search term the engine does not scan the whole web but extracts related links from the database. Search is the heart of the web. It is how we navigate the web. All the information available in the web will become inaccessible if we don't have a search engine to enter our queries.

Search is the means through which we discover information, access services, increase our store of knowledge, and broaden our horizon .Until recently we had to rely on Boolean search. A statistical and analytical technique that uses the operators AND, OR, NOT and NEAR to create a probability model of the answers to our search query. It relies on keywords. E.g. If our query has HELP and SEO, websites having these keywords will be given as answers to the query and also because the contents in the site have the keywords HELP and SEO that are strategically located. Boolean search does not work that simply. It relies on a lot of statistical data. The good thing is that search is changing. It is changing from Boolean search that provides the 10 best probable answers in response to search query which we then have to shortlist visiting each site to a more accurate computational type of search that is typified by search query like "How old is President Obama ?" which provides the correct answer right on the search page.

Search engines on the websites are enriched with facility to search the content stored on other sites. There is difference in the way various search engines work, but they all perform three basic tasks.

Finding and selecting full or partial content based on the keywords provided.

Maintaining index of the content and referencing to the location they find the information.

Allowing users to look for words or combinations of words found in that index.

Semantic Analysis

Semantic analysis is nothing but a process of filtering that progressively eliminates more and more input strings until you are left with only valid data. Semantic search is different from Boolean search as apples are different from oranges. The transition to semantic search also marks the transition on the web as we go from websites to people [7]. The web continues to be made of websites. In websites we get to find information, consume news and buy stuff. In order to understand natural language and search queries, it has to understand what these words really mean.

Metasearch Engine

Metasearch engine is a search tool that uses other search engine's data to produce their own results from the Internet. Metasearch engines take input from a user and simultaneously send out queries to third party search engines for results. Sufficient data is gathered, formatted by their ranks and presented to the users.

Information stored on the World Wide Web is constantly expanding, making it increasingly impossible for a single search engine to index the entire web for resources. Metasearch engine is a solution to overcome this limitation. By combining multiple results from different search engines, metasearch engine is able to enhance the user's experience for retrieving information, as less effort is required in order to access more materials. A metasearch engine is efficient as it is capable of generating a large volume of data, however, scores of websites stored on search engines are all different: this can draw in irrelevant documents. Other problems such as spamming also significantly reduce the accuracy of the search. This issue is tackled by the process of fusion which improves the engineering of metasearch engine. There are many types of metasearch engines available to allow users to access specialised information in a particular field. These include Savvy search engine and Meta seek engine. The advantage of using a metasearch engine is that by sending multiple queries to several other search engines this extends the search coverage of the topic and allows more information to be found. They use the indexes built by other search engines, aggregating and often post-processing results in unique ways. Metasearch engine has an advantage over a single search engine because more results can be retrieved with the same amount of exertion. It also reduces the work of users from having individual type searches from different engines to look for resources [5].

1.2 Genetic Algorithm

In the field of artificial intelligence, a genetic algorithm (GA) is a search heuristic that mimics the process of natural selection. This heuristic (also sometimes called a Meta heuristic) is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection and crossover [24]. It is a very powerful and non-traditional optimization technique. It is based on the Darwinian Theory "Survival of the Fittest". Only the fittest will survive and reproduce and successive generations will become better and better compared to previous generations.

They are stochastic algorithms as they can come up with a different solution every time it is run on the same problem and not deterministic (an algorithm which gives the same answer for a given problem how many times it is run). GA's are creative algorithms in the sense that they make use of the concept of interaction of thousands of probabilities with each other and eventually come up with a solution. GA's are unique in that they operate from a rich database of many points simultaneously and can be incredibly efficient if programmed correctly.

The performance of the Genetic Algorithms for a particular problem can be made with regard to best fitness values obtained from it or the time taken to converge with the fairly optimal solutions because the problem might be time critical or it can also be measured in terms of diversity measures [2]. The performance can also be measured by the number of fitness functions evaluations done during the course of the run. For fixed population sizes the number of fitness function evaluations is given by the product of population size and the number of generations .The efficiency of GA varies from problem to problem and from generation to generation because some genes are solved during the first few generations but others take more time to do so, as the contribution of the genes of one individual towards the fitness function is not the same as some other genes in the same individual i.e. some genes is responsible for a high variance while others change the fitness value only minimally.

2. PROPOSED WORK

In our paper we mainly focus on personalizing the searching process for people working in a software development organization (analysts, developers, testers, maintenance team members), who find it increasingly difficult to get relevant results to their searches. We build group profiles based on either the domain in which software product is to be developed or on project basis [15].

Till now not much development is observed in web personalization field because individual web search behavior has not changed much. The main challenge in web personalization is to read the mind of the users [4]. This imposes a very big challenge because the words used for any search are limited to two or three words. Some of the issues in Web searching are (1) Structuring Queries i.e. the difficulty faced by users are properly structuring queries, namely applying the rules of a particular system, especially Boolean operators e.g., AND, OR, NOT and term modifiers e.g. '+', '!'. (2) Spelling i.e. the user tends to misspell their queries without even realizing it. (3) Query Refinement i.e. many times the users do not refine their query, even if there may be other terms that relate directly to their needed information. (4) Managing Results i.e. mostly, the user queries are extremely broad, resulting in an unmanageable number of results. Few users view more than the first ten or twenty documents from the result list.

2.1 Semantic Analysis

Analysis of the user's queries at a semantic level using vocabulary or ontology based system like ODP [8][29] or yahoo Directory [9] is semantic analysis. Optimal results from semantic analysis are chosen using genetic algorithm, where only the results that are most suitable to the users profile and interests are presented to the user [5]. Genetic algorithm aids with machine learning and supports the search engine to understand the user's mind while searching. Optimality of the results from semantic analysis is based on the user's profile that is built and the results of task analysis.

It helps in addressing the two most significant problems which is encountered during traditional content based filtering.

1. Cold start problem

2. Filter Bubble

2.1.1 Cold Start Problem

The lack of user rating leads to "cold start" problem. Initially when a user searches in a new domain he will not have the luxury of tracing recommended searches. Using semantic content based filtering and retrieving more semantically related concepts this problem can be solved.

2.1.2 Filter Bubble

Semantic analysis helps is overcoming the problem of over specialization. It means that the user is restricted to get recommendations which have strong resemblance to the one he already knows. This problem is also referred to as "Filter Bubble".

2.2 Quest Analysis

We define a quest to be the objective of the user's search (or) an atomic user information need (goal of a user's search), whereas a quest trail represents all user activities within that particular quest, such as query reformulations, URL clicks. Previously, Web search logs have been studied mainly at session[3] or query level where users may submit several queries within one quest and handle several quests within one session[6][26]. Quest level analysis of search log provides a better understanding of user's interests or goal, since it performs better in modelling user's profile. Thus the user behavior [22] can be studied and noted from the tasks he performs in the search engine. Thus task identification is important. We make use of the same task elicitation algorithm called Query Bundle - QUEST.

2.2.1 Bundling Queries into Quest

Some of the previous methods used for bundling queries into quest were WCC (weighted connected component), HTC (Head Tail component). In WCC an undirected graph for queries within a session was built. The vertices of the graph were queries and the edges were similarity scores between queries. After removing the suspicious edges with scores below a threshold, any connected component of the remaining graph is identified as a query bundle. WCC outperformed other popular clustering algorithms like Query Flow Graph [1], K-means, and DB-Scan in bundling queries into quest, as indicated in [26]. WCC was found to be better than any other previous bundling algorithms because, every query was compared with every other query before bundling queries into Quests. But the time complexity of WCC is O(k·N2), where N is the average number of queries of a session and k is the dimension of features. The overall time complexity in intolerable for search logs of massive volume.

To overcome this Orlando .S. [26] proposed another head-tail component query clustering approach (HTC) to reduce the time complexity. In this approach only the similarity between head tail components were considered for bundling queries into quest. This fails to address cases of interleaved quests. We are proposing a new approach that could reduce the time complexity while addressing the interleaving quests. We name this algorithm as QB-Q.

This algorithm bundles queries belonging to a quest or relevant quest. Say for instance there are 4 queries in a search log A, B, C, and D. WCC would have needed 6 pairs of relevance computation [4] [12] [19][25], whereas our proposed method will lesser number of relevance computation unless it is a worst case where every query is irrelevant to every other query in the search log. To be more precise if A is similar to B and B is similar to C, there is no need to compute the relevance between A and C any more. If A is similar to B but B is not similar to C, QB-Q still has to compute the relevance between A and B to avoid the quest interleaving.

QB-Q is efficient approach to bundle queries of related quests or same quest. Extracting such information regarding the user's objective of searching helps us to model a stronger and dynamic user profile. Thus we can develop a more accurate profile to reflect the user's requirements than the ones that are statically created at the time of registration.

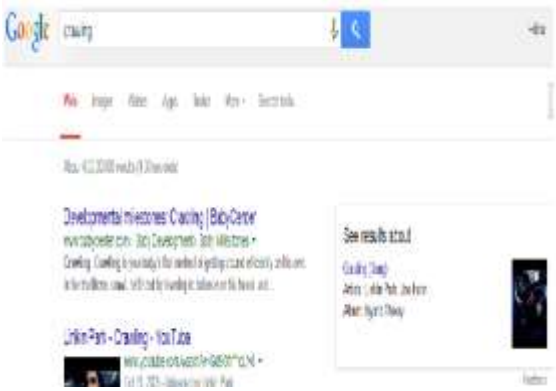
Algorithm 1: Query Bundle – QUEST

Input: Query set *Q*, cut-off threshold *b*;
Output: A set of Quest q;
Initialization: q= *null* ;
Query to Quest table *L [ ] =null* ;
1: **for** *len* = 1 **to** |*Q*| − 1 **do**
2: **for** *i* = 1 **to** |*Q*| − *len* **do**
3: **if** *L* [*Qi*] != *L* [*Qi+len*] **then** // if two queries are not in the same quest
4: *s* ← compare (*L* [*Qi*], *L* [*Qi+len*]); // compute similarity takes *O*(*k*)
5: **if** *s* ≥ *b* **then**
6: merge q (*Qi*) and q (*Qi+len*);
7: modify *L*;
8: **if** |q| = 1 break; // break if there is only one task
9: return q;

Let us now see a comparison of search result from our proposed work and a search result from Google. Assume the user is working in a cloud computing domain and he issues a search for the word "crawling". The same word will have different meaning in different context. That is the reason for the different results observed when the same word is searched in Google and our personalized metasearch engine "QUEST TRAIL".

In Google we observe results regarding, a baby's first movements, a Linken Park's song, and insects crawling and so on. The results do not seen to match the user's quest. Whereas the results of "QUEST TRAIL" are all relevant to the user's domain and hence is much more relevant to the user's quest.

/settings/ads/preferences%3Fhl%3Den
http://link.springer.com/chapter/10.1007%252F978-3-642-35864-7_53
http://link.springer.com/content/pdf/10.1007%252F978-3-642-35864-7_53.pdf
http://yourstory.com/2012/05/promptcloud-data-crawling-and-cloud-computing-solutions/
http://www.researchgate.net/publication/216965265_Service_Crawling_in_Cloud_Computing
http://www.connotate.com/4-strategies-everyone-web-crawling-industry-using-6414
https://www.promptcloud.com/
http://www.slideshare.net/ideseditor/an-efficient-cloud-based-approach-for-service-crawling



3. OTHER TECHNIQUES

Some of the other techniques that are currently being used for personalization are briefly listed below. Query rewriting, semantic content filtering, re-ranking [10], semantic celebrative filtering, user modelling or profiling and analysis of search logs are techniques that are used to improve the relevancy in the search results for the user, while reducing their effort[20].

3.1 Query Reformulation

Here the query is elaborated by the user to personalize the search result [30]. For example the query is to find the Thai restaurants located in the city of Chennai. Here in this technique Chennai is added to the search query and taken as "Thai restaurants Chennai "and the search results are given for this query. The search engine will now give the results for Thai restaurants that are in Chennai. The problem over here is there are chances where the user may not be clear about the location.

Let us say for example, a website organizer might use a word which he likes most but an individual looking for the same information might not go for the same word, instead he might use its synonym. Then tracing of such web pages will be quite a challenge for the search engines. Synonyms refer to many words expressing same meaning and poli-semis refer to one word with different meanings. Owing to this kind of language richness and the context sensitive sense a word assumes, the keyword method used by search engines faces quite a lot of issues. A user seeking for information is expected to keep reframing their query until it matches the form that is expected by the search engine.

3.2 URL Re-Ranking

Re-Ranking[10] the results for a user based on his profile is one of the conventional approaches for personalization. Page re-ranking is used mainly to take the advantage of user's profile. Initially some 'n' documents are taken that are reordered as per the preference from the user profile [22]. The re-ranking occurs by scores assigned to each SERP that checks with the user profile.

3.3 User Modeling in Personalized Systems

Traditional methods for modelling user profile were focused on creating a single static profile at the time of registration. It does not address the problem of different queries being needed to be handled differently. Collecting the user preferences and choices of the user at the time of registration helps in predicting the needs of the user.

3.4 Google's Approach

Google is taking several steps to improve the search results that are provided to the user. It provides the user with personalized results if the preferences are given by the user initially. For this enhancement it is required by the user to create a profile. The user is required to give the details of his preferences and based on it Google retrieves the personalized results to the user and also updates the user of any new information through mail.

4. CONCLUSION

In a software development organization there is a special need for quest – specific or domain – specific ranking. Applying QUEST level analysis of the search log for constructing a personalized search engine for software developers will make the searching process much easier. Our system proved an effective performance in personalizing the searching process especially for software developers. It is clearly seen that the combination of quest analysis and semantic analysis is an effective approach for personalization of searching process and it is better than any of the currently used techniques for personalization.

5. REFERENCES

[1] Boldi, P., Bonchi, F., Castillo, C., Donato, D., Gionis, A. And Vigna, S., "The query-flow graph: model and applications," ser. CIKM '08, 2008, pp. 609–618.

[2] Radlinski, F. and Craswell, N., "Comparing the sensitivity of information retrieval metrics," ser. SIGIR '10, 2010, pp. 667–674.

[3] Catledge, L.D. and Pitkow, J.E., "Characterizing browsing strategies in the world-wide web," Computer Networks and ISDN Systems, vol. 27, no. 6, pp. 1065–1073, 1995.

[4] Huang, C.K., Chien, L.F. and Oyang, Y.J., "Relevant term suggestion in interactive web search based on contextual information in query session logs," Journal of the American Society for Information Science and Technology, 2003.

[5] White, R., Bailey, P. and Chen, L., "Predicting user interests from contextual information," ser. SIGIR '09, 2009, pp. 363– 370

[6] Cao, H., Jiang, D., Pei, J., He, Q., Liao, Z., Chen, E. and Li, H., "Context-aware query suggestion by mining click-through and session data," in KDD '08, 2008, pp. 875–883.

[7] Song, Y., Zhou, D., and He, L.-w., "Query suggestion by constructing term-transition graphs," ser. WSDM '12, 2012, pp. 353–362.

[8] White, R. and Huang, J., "Assessing the scenic route: measuring the value of search trails in web logs," ser. SIGIR '10. ACM, 2010, pp. 587–594.

[9] Donato, D., Bonchi, F., Chi, T. and Maarek, Y., "Do you want to take notes? identifying research missions in yahoo! Search pad," ser. WWW '10, 2010, pp. 321–330.

[10] Xiang, B., Jiang, D., Pei, J., Sun, X., Chen, E. and Li, H., "Context-aware ranking in web search," ser. SIGIR '10. ACM, 2010, pp. 451–458.

[11] Jain, A., Ozertem, U. and Velipasaoglu, E., "Synthesizing high utility suggestions for rare web search queries," ser. SIGIR '11, 2011, pp. 805–814.

[12] Craswell, N. and Szummer, M., "Random walks on the click graph," ser. SIGIR '07, 2007, pp. 239–246.

[13] Jones, R. and Klinkner, K.L., "Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs," ser. CIKM '08, 2008, pp. 699–708.

[14] Liu, Y., Gao, B., Liu, T.-Y., Zhang, Y., Ma, Z., He, S. and Li, H., "Browserank: letting web users vote for page importance," ser. SIGIR '08, 2008, pp. 451–458.

[15] Beeferman, D. and Berger, A., "Agglomerative clustering of a search engine query log," ser. KDD '00, New York, NY, USA,2000, pp. 407–416.

[16] White, R., Bilenko, M. and Cucerzan, S., "Studying the use of popular destinations to enhance web search interaction," ser. SIGIR '07, 2007, pp. 159–166.

[17] Fox, S., Karnawat, K., Mydland, M., Dumais, S. and White, T., "Evaluating implicit measures to improve web search," ACM Trans. Inf. Syst., vol. 23, pp. 147–168, 2005.

[18] Chapelle O., Joachims T., Radlinski F. and Yisong Yue, "Largescale validation and analysis of interleaved search evaluation," ACM Trans. Inf. Syst., vol. 30, no. 1, p. 6, 2012.

[19] Gao, J., Yuan, W., Li, X., Deng, K. and Nie, J.-Y., "Smoothing clickthrough data for web search ranking," ser. SIGIR '09. ACM, 2009, pp. 355–362.

[20] He, D., G¨oker, A. and Harper, D.J., "Combining evidence for automatic web session identification," Inf. Process. Manage., vol. 38, no. 5, pp. 727–742, 2002.

[21] Silverstein, C., Henzinger, M.R., Marais, H. and Moricz, M., "Analysis of a very large web search engine query log," SIGIR Forum, vol. 33, pp. 6–12, 1999.

[22] Hassan, A., Jones, R., and Klinkner, K., "Beyond dcg: user behavior as a predictor of a successful search," ser. WSDM '10, 2010, pp. 221–230.

[23] Jansen, B., Spink, A. and Kathuria, V., "How to define searching sessions on web search engines," ser. WebKDD '06, 2007, pp. 92–109.

[24] Kotov, A., Bennett, P., White, R., Dumais, S. and Teevan, J., "Modeling and analysis of cross-session search tasks," ser. SIGIR '11, 2011, pp. 5–14.

[25] Jones, R., Rey, B., Madani, O. and Greiner, W., "Generating query substitutions," ser. WWW '06. ACM, 2006, pp. 387–396.

[26] Lucchese, C., Orlando, S., Perego, R., Silvestri, F., and Tolomei, G., "Identifying task-based sessions in search engine query logs," ser. WSDM '11, 2011, pp. 277–286.

[27] Hassan, A., Song, Y. and He, L.-w., "A task level user satisfaction metric and its application on improving relevance estimation," ser. CIKM '11, 2011.

[28] Olston, C. and Chi, E.H., "Scenttrails: Integrating browsing and searching on the web," ACM Trans. Comput.-Hum. Interact., vol. 10, pp. 177–197, September 2003.

[29] Shen, X., Tan, B. and Zhai, ChengXiang, "Context-sensitive information retrieval using implicit feedback," ser. SIGIR '05, 2005, pp. 43–50.

[30] Mei, Q. and Zhou, D. and Church, K., "Query suggestion using hitting time," ser. CIKM '08. ACM, 2008, pp. 469–478.

[31] IEEE TRANSACTIONS O,NV OKLN.O26W, LNEOD.G2E, FAENBDR EUNAGRINYE 2E0R1I4NG,VOL.26, NO.12, APRIL 2014 "Task Trail: An Effective Segmentation of user search behaviour" Zhen Liao, Yang Song, Yalou Huang, Li-wei He, Qi He.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 235 - 241, 2015, ISSN:- 2319–8656

Project Scheduling: Survey and Research Potentials

Asmaa Atef
Department of Operations Research, Faculty of Computers and Informatics, Zagazig University, El-Zera Square, Zagazig, Sharqiyah, Egypt

Mohamed Abdel-Baset
Department of Operations Research, Faculty of Computers and Informatics, Zagazig University, El-Zera Square, Zagazig, Sharqiyah, Egypt

Ibrahim El-henawy
Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, El-Zera Square, Zagazig, Sharqiyah, Egypt.

Abstract: project scheduling is very critical topic in project management. Resource constrained project scheduling problem (RCPSP) consists of activities that must be scheduled based on dependencies relationships and priorities of activities. In the recent years there have been many survey papers around the area of project scheduling, as many researchers developed both exact and heuristic scheduling schemes. This paper give an over view around the resource constrained project scheduling problem (RCPSP).

Keywords: Resource allocation; constrained resource scheduling; scheduling techniques; Project management

1. INTRODUCTION

Scheduling is the way we actually manage a project. Without scheduling, nothing or nobody is managing the project and hence amounts to failure of a project. Scheduling describes guidance and pathway for a project to run. It defines certain milestones and deliverables which need to be achieved on a timely basis for successful completion of a project. Monitoring the schedule provides an idea of the impact the current problems are having on the project, and provides opportunities to enhance or reduce the scope of a milestone/phase in the project. It also provides a medium for continuous feedback on how the project is progressing and if there are issues that need to be dealt with or if the client needs to be told about a delay in delivery (et al *Shruti Gauri).*

To organize and complete your projects in a timely, quality and financially responsible manner, you need to schedule projects carefully. Effective project scheduling plays a crucial role in ensuring project success. To keep projects on track, set realistic time frames, assign resources appropriately and manage quality to decrease product errors. This typically results in reduced costs and increased customer satisfaction. Important factors include financial, documentation, management and quality assurance.

Project scheduling impacts the overall finances of a project. Time constraints require project managers to schedule resources effectively. This is particularly true when resources must have highly specialized skills and knowledge in order to complete a task or when costly materials are required. Completing a project in a short time frame typically costs more because additional resources or expedited materials are needed. With accurate project scheduling, realistic estimates and accurate projections prevent last-minute orders that drive up costs.

A well-managed project will require a proper schedule and related documents. Project schedule will help to organize all the related tasks related to project in a easy to manage way. Project managers can easily see and manage tasks. Without a schedule you will never know when you're going to finish your project. Some people may just start projects without a proper schedule. If you don't make a project schedule, it is very likely that you may have to spend sleepless nights rushing to finish the project on time. And at the end of the day, the quality of the finished project might be poor.

Estimating the duration and making a schedule is not a thing which is possible for the inexperienced. You will need some amount of experience before you will be able to estimate. If you have never worked on a software project or never implemented an application, you won't be able to draw a time-line very well. Project scheduling ensures one task gets completed in a quality manner before the next task in the process begins. By assuring that quality measures meet expectations at every step of the way, you ensure that managers and team members address problems as they arise and don't wait until the end.

No major issues should appear upon completion because you've established quality controls from the very beginning of the scheduling process. Effective project managers understand that ensuring quality control involves managing risks and exploiting opportunities to speed up the schedule when possible

to beat the competition and achieve or maintain a competitive edge with a more reliable product.

1.1. The Resource-Constrained Project Scheduling Problem

A combinatorial optimization problem informally, a resource-constrained project scheduling problem (RCPSP) considers resources of limited availability and activities of known durations and resource requests, linked by precedence relations. The problem consists of finding a schedule of minimal duration by assigning a start time to each activity such that the precedence relations and the resource availabilities are respected.

1.2. Project management and scheduling

Project management and scheduling are major issues for every company or organization facing rapid changes in its environment. Furthermore, project scheduling techniques have become indispensable for the attainment of effective-ness and efficiency of processes. This special issue collects 11 carefully selected papers which deal with optimization or decision analysis problems in the field of project management and scheduling. It covers a considerable range of topics, including solution methods for classical project scheduling problems (Bianco and Caramia, Kone´et al., Lim et al., Van Peteghem and Vanhoucke), mixed-integer programming (MIP) formulations in order to solve scheduling problems with commercial MIP-solvers (Bianco and Caramia, Kone et al.), models and solution algorithms for multi-project scheduling (Besikci et al.), extensions of the classical resource-constrained project scheduling problem (Hartmann), models and solution approaches integrating decisions at different managerial levels such as selection and scheduling, or scheduling and control (Gutjahr and Froeschl, Hazir and Schmidt), risk in project management and scheduling (Tian and Demeulemeester, Artigues et al., Hazir and Schmidt, Gutjahr and Froeschl), and the assignment of resources to different stakeholders (Fink and Homberger, Besikci et al.).

The operations research techniques employed include common approaches such as metaheuristics (Lim et al., Fink and Homberger, Van Peteghem and Vanhoucke, Besikci et al., Gutjah and Froeschl) and mixed-integer programming (Bianco and Caramia, Kone et al.), but also specific approaches such as scenario-relaxation algorithms (Artigues et al.), a Lagrangian-based algorithm (Beşikci et al.), and a Frank-Wolfe type algorithm (Gutjahr and Froeschl). This special issue thus covers a broad range of models, algorithms and applications. It demonstrates that 55 years after the invention of CPM and PERT, the field is more innovative and lively than ever. We encourage the readers to seek inspiration from this unique collection of papers and to continue work in this area, for instance by applying these results in different application areas, by extending these models further, or by developing even better algorithms for solving them efficiently.

2. LITERATURE REVIEW

The project scheduling problems of small sizes can be solved by the traditional optimization techniques. However, as the number of projects and size of the project in terms of number of activities increase, the problem becomes more complex .Further, the complexity increases when variety of resources is considered. In this context, it is not feasible to develop the projects' schedules by using the traditional optimization techniques. The benefit of traditional optimization techniques cannot be utilized for generating the schedule of multiple projects simultaneously. Accordingly, researchers have developed several heuristic algorithm and rules and Meta heuristic methods for multiple resource scheduling and multi- project scheduling. Scheduling and allocation of resources for multiple projects is a non-polynomial (NP) hard problem and more difficult than a single project. Traditional optimization methods have been used in the literature to solve the multi project scheduling problems. The categorization of solution procedures for scheduling problem are classified as cited in Shouman et al [4,5], Rainer Kolisch and Sonke Hartmann [6] into priority rule-based-X-pass methods, classical meta-heuristics, non-standard meta-heuristics, and other heuristics. In X-pass methods a priority order is given for each specified activity included in the project under either a serial and/or parallel scheduling generated strategy. X-pass methods have been summarized in Alvaerz-Valdes and Tamarit [7], Boctor [8], Cooper [9,10], Davies [11], Davies Atterson [12], Elsayed [13], Klein [14], Kolisch [15,16,17], Kolisch and Drexl [18,19] Lawrence [20], Li and Willis [21], Ozdamar and Ulusoy [22,23,243], Patterson [25,26], Schirmer [27],

Schirmer and Riesenberg [28], Thesen [29], Thomas and Salhi [30], Ulusoy and Ozdamar [31], Valls et al. [32], and Whitehouse and Brown [33]. Shouman et al [4,5] proposed fifty-five heuristic rules for scheduling projects. These heuristic rules are tested for single and multiple resource(s) constrained projects using fifty test problems. The projects are classified into different categories. The performances of the scheduling process using these heuristic rules for the considered project categories are discussed and evaluated. The performance of these scheduling rules are evaluated and appeared promising tendencies in scheduling process for single and multiple oriented critical resources(s). Also he proposed a genetic algorithm for scheduling the same test instances and compared the resulted schedules by those rendered from heuristic rules and proofs the robustness of heuristic rules performance. Lova et al [34] presents some heuristics based on priority rule for the MRCPSP with renewable resources are analyzed. These methods are very important in the construction of more sophisticated heuristics as random sampling techniques or meta-heuristics. Hence, additional efforts in order to obtain better heuristics based on priority rule are justified. Three components of this type of heuristics are analyzed: scheduling generation scheme, priority rule and mode selection rule. Based on a well-known set of 240 randomly generated project instances, it is shown that the single-pass and multi-pass heuristics based on priority rules developed in this work greatly outperforms the ones previously published. The serial schedule generation scheme greatly outperforms the parallel scheme with the majority of the priority rules tested justifying the higher computational effort required by the former. Finally, a multi-pass method that combines eight heuristics based on priority rule obtains the lowest average deviation with respect to the critical path length (32.0%), thus being the best deterministic heuristic for this problem. Myszkowski et al [35] presents some novel scheduling heuristics for Multi–Skill Resource–Constrained Project Scheduling Problem and compared to state-of-the-art priority rules, based on task duration, resource salaries and precedence relations. New heuristics stand an aggregation of known methods, but are enhanced by skills domain. The goal of the paper is to investigate, whether evaluated methods can be used as robustness enhancement tools in meta-heuristics, mostly evolutionary algorithms. Experiments have been performed using artificially created dataset instances, based on real–world instances. Obtained results prove that such methods stand interesting feature that can be included to more complex methods and increase their robustness. Colak et al [36] consider the multi-mode resource-constrained project scheduling problem (MRCPSP) with renewable resources. In MRCPSP, an activity can be executed in one of many possible modes; each mode having different resource requirements and accordingly different activity durations. We assume that all resources are renewable from period to period, such as labor and machines. A solution to this problem basically involves two decisions – (i) The start time for each activity and (ii) the mode for each activity. Given the NP-Hard nature of the problem, heuristics and meta-heuristics are used to solve larger instances of this problem. A heuristic for this type of problem involves a combination of two priority rules - one for each of the two decisions. Heuristics generally tend to be greedy in nature. In this study we propose two non-greedy heuristics for mode selection which perform better than greedy heuristics. In addition, we study the effect of double justification and backward/forward scheduling for the MRCPS. We also study the effect of serial vs. parallel scheduling. We found that all these elements improved the solution quality. Finally an adaptive meta-heuristic procedure based on neural networks is proposed which further improves the solution quality. The effectiveness of these proposed approaches, compared to existing approaches, is demonstrated through empirical testing on two well known sets of benchmark problems. Chen and Lo [37] present and evaluate a modified ant colony optimization (ACO) approach for the resource-constrained project scheduling problems. A modified ant colony system is proposed to solve the resource-constrained scheduling problems. A two-dimensional matrix is proposed in this study for scheduling activities with time, and it has a parallel scheme for solving project scheduling problems. There are two designed heuristic is proposed. The dynamic rule is designed to modify the latest starting time of activities and hence the heuristic function. In exploration of the search solution space, this investigation proposes a delay solution generation rule to escape the local optimal solution. Simulation results demonstrate that the proposed modified ant colony system algorithm

provides an effective and efficient approach for solving project scheduling problems with resource constraints. Bastani and Yakhchali [38] proposed OR model for multi-mode resource-constrained project scheduling problem (MRCPSP), in which multiple execution modes are available for each of the activities of the project and pre-emptive extension of the activities which allows activity splitting is introduced. Also two important constraints about the length of the split and the number of split in each activity are added. In this paper each activity split up to N division is allowed and the length of each activity up to X is limited. Setup time for each pre-emption, is considered and setup time to objective function is added. This model is compared to other models that have been presented until now and the advantages of this model in comparison with others are explained. At last the most useful methods to solve this model is presented. A new recombination operator is presented by Andreica and Chira [39] for permutation based encoding has been proposed in this paper. This operator is suitable to Resource-Constrained Project Scheduling Problem and Multi-Mode Resource-Constrained Project Scheduling Problem because it preserves the precedence constraints when obtaining the offspring from feasible parents. The main feature of the proposed operator is the use of genetic information from the best individual besides the two parents considered for recombination. Experimental results performed on ProGen project instances indicate a superior performance of the proposed operator, thus emphasizing the role that recombination has in accelerating the search in an evolutionary process. An extensive study of all recombination operators used for RCPSP and MRCPSP will be performed and the experiments will be extended to more instances with more activities. Buddhaklsomsiri and Kim [40] introduced a priority rule-based heuristic for the multi-mode resource-constrained project scheduling problem with the splitting of activities around unavailable resources allowed. All resources considered are renewable and each resource unit may not be available at all times due to resource vacations, which are known in advance. A new concept called moving resource strength is developed to help identify project situations where activity splitting is likely to be beneficial during scheduling. The moving resource strength concept is implemented in priority rule-based heuristics to control activity splitting when scheduling. Multiple comparisons of the performance of combination of activity–mode priority rules used in the heuristics are provided. Computational experiments demonstrate the effectiveness of the heuristic in reducing project makespan, and minimizing activity splitting. Browning and Yasine [41] address the *static* resource-constrained multi-project scheduling problem (RCMPSP) with two lateness objectives, project lateness and portfolio lateness. In this context, past research has reported conflicting results on the performance of activity priority rule heuristics and does not provide managers with clear guidance on which rule to use in various situations. Using recently improved measures for RCMPSP characteristics, they conducted a comprehensive analysis of 20 priority rules on 12,320 test problems generated to the specifications of project-, activity-, and resource-related characteristics—including network complexity and resource distribution and contention. They found several situations in which widely advocated priority rules perform poorly and confirmed that portfolio managers and project managers will prefer different priority rules depending on their local or global objectives. The results are presented in two decision tables, the practical use of which requires managers to do only a rough, qualitative characterization of their projects in terms of complexity, degree of resource contention, and resource distribution. He and Zhang [42] proposed a dynamic priority rule-based forward-backward heuristic algorithm (FBHA). The FBHA optimizes resource allocation by shifting non-critical activities within their forward free float (FFF), forward total float (FTF) and backward free float (BFF), successively. A project is divided into several phases during each forward/backward scheduling module. In each phase, the shifting sequence and days of non-critical activities depend on a dynamic priority rule set. The FBHA is integrated into the Microsoft Project 2007 commercial software package to improve the performance of the software and facilitate the project planners. Singh [43] introduces an attempt to integrate the project priorities with the project schedule development. A hybrid algorithm was developed to accomplish this task. The presented algorithm is a new method for generating the schedule of any multi-project resource constrained scheduling problem where, each project has a

defined criticality. The proposed method was validated with a case study under various scenarios. Experimental results were compared with existing priority dispatching rules. Experimental results showed the superiority of the proposed method with the existing priority dispatching rules under different operating conditions. In real project management environment, a penalty is imposed if a project completes after its due date. Some projects carry higher penalty than others. In this context, project manager can make a tradeoff among the projects penalties and can develop the cost effective project schedule, which satisfies the customer requirements. In this context, the proposed algorithm would be beneficial for project manager's to deal with these conditions. In the future, research could explore the possibility of integrating other knowledge areas including risk management and procurement management with the project schedule development. Lova and Tormos [44] presents an algorithm to study the effect of the schedule generation schemes – serial or parallel – and priority rules – MINLFT, MINSLK, MAXTWK, SASP or FCFS – with two approaches – multi-project and single-project. The time criteria considered are the mean project delay and the multi-project duration increase. Through an extensive computational study, results show that with the parallel scheduling generation scheme and the multi-project approach the project manager can obtain a good multi-project schedule with the time criterion selected: minimizing mean project delay or minimizing multi-project duration increase. New heuristics – based on priority rules with a two-phase approach – that outperform classical ones are proposed to minimize mean project delay with a multi-project approach. Finally, the best heuristics analyses are evaluated together with a representative sample of commercial project management software. Browning and Yassine [45] introduce the static resource-constrained multi-project scheduling problem (RCMPSP)with two lateness objectives,projectlatenessandportfoliolateness.Inthisc ontext,pastresearchhasreported conflicting results on the performance of activity priority rule heuristics and does not provide managers with clear guidance on which rule to use in various situations. Using recently improved measures for RCMPSP characteristics, they conducted a comprehensive analysis of 20 priority rules on 12,320 test problems generated to the specifications of project-, activity-, and resource-related characteristics including network complexity and resource distribution and contention. We found several situations in which widely advocated priority rules perform poorly. They also confirmed that portfolio managers and project managers will prefer different priority rules depending on their local or global objectives.

3. CONCLUSION AND FUTURE WORK

Project scheduling is very important topic in project management. In this survey we have attempted to review the resource constrained project scheduling problem. We give emphasis to the heuristic rules used as inputs to the scheduling process. No one rule can achieve the best schedules in all projects as each project has special characteristics. So we develop system of large number of heuristic rules for scheduling multiple resource constrained projects to adapt to the variation of projects [46].

We encourage researchers to develop flexible heuristic decision-making procedures to meet the needs of practitioners. Also encourage researchers to develop genetic systems in scheduling based on heuristic rules that can make results of scheduling are better than genetic system from scratch.

Develop heuristic procedures in multi-mode resource constrained projects scheduling problems [46].

REFERENCES

[1] A., A., Najafi, and F., Azimi, "A Priority Rule-Based Heuristic for Resource Investment Project Scheduling Problem with Discounted Cash Flows and Tardiness Penalties", Mathematical Problems in EngineeringVolume 2009 (2009), Article ID 106425, 10 pages doi:10.1155/2009/106425.

[2] J., M., Nicholas, and H., Steyn, "Project Management for Business, Engineering, and Technology "Business & Economics press, 2088.

[3] M.,Vanhoucke, "Optimizing Regular Scheduling Objectives: Priority Rule Based Scheduling," *PM Knowledge Center*, 2012.

[4] M., A. Shouman M.S. Ibrahim M. Khater A.A Elfrgany "Some Heuristic Rules for Scheduling Single and Multiple Resources Constrained Projects" , AEJ, Journal , Alex. 2007.

[5] M.A., Shouman, A.Z., Ghafagy, M.A., Zaghloul, and A., Bou-Shaala, "New Heuristics for Scheduling Single Constrained Resource Projects", AEJ, Journal, Vol. 38, No. 3, 1999.

[6] R., Kolisch, and S., Hartmann, "Experimental Investigation of Heuristics for Resource-constrained

Project Scheduling: Update". European *Journal of Operational Research*, 2005

[7] R., Alvarez-Valdes, and J.M., Tamarit, "Heuristic Algorithms for Resource-constrained Project Scheduling: A Review and an Empirical Analysis", In R.Slowinski and J. Weglarz, editors, Advances in project scheduling ,pages 113-134.Elsevier,Amsterdam, Netherlands,1989.

[8] F.F., Boctor, "Some Efficient Multi-heuristic Procedures for Resource-constrained Project Scheduling", *European Journal of Operational Research*, 49:3-13, 1990

[9] D.F., Cooper, "Heuristics for Scheduling Resource-constrained Projects: An Experimental Investigation" *Management Science, 22:1186*-1194, 1976.

[10] D.F., Cooper, "A Note on Serial and Parallel Heuristics for Resource-constrained Project Scheduling", *Foundations of Control Engineering*, 2:131-133, 1977.

[11] E.M., Davies, "An Experimental Investigation of Resource Allocation in Multi-activity Projects", *Operations Research Quarterly, 24:587*-591, 1973.

[12] E.W., Davis and J.H., Patterson, " A Comparison of Heuristic and Optimum Solutions in Resource-constrained Project Scheduling", Management *Science* , 21:944-955, 1975

[13] E.A., Elsayed, "Algorithms for Project Scheduling with Resource Constraints", *International Journal of Production Research*, 20:95-103, 1982

[14] R., Klein, "Bidirectional Planning: Improving Priority Rule Based Heuristics for Scheduling Resource –constrained Projects", *European Journal of Operational Research*, 127:619-638, 2000.

[15] R., Kolisch, "Project Scheduling under Resource Constraints–efficient Heuristics for Several Problem Classes", Physica, Hdeidelberg, 1995.

[16] R., Kolisch, "Efficient Priority Rules for the Resource-constrained Project Scheduling Problem", *European Journal of Operations management, 14:179*-192, 1996

[17] R., Kolisch, "Serial and Parallel Resource–constrained Project Scheduling Methods Revisited: Theory and Computations Management", *European Journal of Operational Research*, 90:320-333, 1996.

[18] R., Kolisch and A., Drexl, "Adaptive Search for Solving Hard Project Scheduling Problems", *Naval Research Logistics, 43:23*-40, 1996

[19] S.R., Lawrene, Resource Constrained Project Scheduling- A Computational Comparison of Heuristic Scheduling Techniques , Technical Report, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1985.

[20] K.Y., Li and R.J., Willis, "An Iterative Scheduling Technique for Resource–constrained Project Scheduling', *European Journal of Operational Research*, 56: 370-379, 1992

[21] L., Ozdamar and G., Ulusoy, "A Local Constraint Based Analysis Approach to Project Scheduling Under General Resource Constraints", *European Journal of Operational Research*, 79: 287-298, 1994.

[22] L., Ozdamar and G., Ulusoy, "An Iterative Local Constraint Based Analysis for Solving the Resource–constrained Project Scheduling Problem", *European Journal of Operations management*, 14: 193-208, 1996.

[23] L., Ozdamar and G., Ulusoy, "A Note on an Iterative Forward/Backward Scheduling Technique with Reference to a Procedure", by Li and Willis. *European Journal of Operational Research*, 89: 400-407, 1996

[24] J.H., Patterson, "Alternate Methods of Project Scheduling with Limited Resources", *Naval Research Logistics Quarterly*, 20:767-784, 1973

[25] J.H., Patterson, "Project Scheduling: The Effects of Problem Structure on Heuristic Performance", *Naval Research Logistics Quarterly*, 23: 95-123, 1976

[26] A., Schrimer, "Case-based Reasoning and Improved Adaptive Search for Project Scheduling ", .*Naval Research Logistics*, 47: 201-222, 2000

[27] A., Schrimer, and S., Riesenberg, "Parameterized Heuristics for Project Scheduling-biased Random Sampling Methods," Manuskripte aus den instituten fur betriebswirtschaftslehre, 456, universitat kiel, germany, 1997.

[28] A., Schrimer and S., Riesenberg, "Class-based Control Schemes for Parameterized Project Scheduling Heuristics", Manuskripte aus den instituten fur betriebswirtschaftslehre 471, universitat kiel, germany, 1998.

[29] A., Thesen, "Heuristic Scheduling of Activities Under Resource and Precedence Restrictions", *Management Science*, 23: 412-422, 1976.

[30] P., R.Thomas, and S., Salhi, "An Investigation into the Relationship of Heuristic Performance with Network-resource Characteristics", *Journal of Operational Research Society*, 48: 34-43, 1997.

[31] G., Uiusoy, and L., Ozdamar, "Heuristic Performance and Network/Resource Characteristics in Resource-constrained Project Scheduling", *Journal of Operational Research Society*,40: 1145-1152, 1989.

[32] V., Valls,, M.A., Perez, and M.S., Quintanilla, "Heuristic Performance in Large Resource-constrained Projects", Technical Report 92-2, Department of Statistics and Operations Research, University of Vaencia, 1992.

[33] G.E., Whitehouse, and J.R., Brown, "GENRES: An Extension of Brooks Algorithm for Project

Scheduling with Resource Constraints", *Computers& Industrial Engineering*, 3: 261-268, 1979.

[34] A., Lova, P., Tormos, and F., Barbe, "Multi-Mode Resource Constrained Project Scheduling: Scheduling Schemes, Priority Rules and Mode Selection Rules", Intelligence Artificial, V. 10, Nº 30, 2006.

[35] B., Paweł, E., M., Marek, S., Nski, and L., Podlodowski, "Novel heuristic solutions for Multi Skill Resource Constrained Project Scheduling Problem", Proceedings of the 2013 Federated Conference on Computer Science and Information Systems pp. 159–166

[36] S., Colak, A., Agrwal, and S.,Erenguc, " Multi-Mode Resource-Constrained Project Scheduling Problem With Renewable Resources: New Solution Approaches", Journal of Business & Economics Research – November 2013 Volume 11, Number 11.

[37] R., M., Chen, and S., T., LO, "Using an Enhanced Ant Colony System to Solve Resource-Constrained Project Scheduling Problem", IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.11, November 2006

[38] M., Bastani, and S., H., Yakhchali, "Multi-mode Resource-Constraint Project Scheduling Problem (MRCPSP) with pre-emptive activities", nd International Conference on Mechanical, Automobile and Robotics Engineering (ICMAR'2013) March 17-18, 2013 Dubai (UAE).

[39] A., Anderica, and C., Chira, "Best-Order Crossover in an Evolutionary Approach to Multi-Mode Resource-Constrained Project Scheduling", International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 6 (2014) pp., 364 – 372.

[40] J., Buddhakulsomsiri, and D., Kim, "Priority rule-based heuristic for multi-mode resource-constrained project scheduling problems with resource vacations and activity splitting", European Journal of Operational Research Volume 178, Issue 2, 16 April 2007, Pages 374–390.

[41] T., Browining, and A., A., Yassine, "Resource-constrained multi-project scheduling: Priority rule performance revisited", International Journal of Production Economics Volume 126, Issue 2, August 2010, Pages 212–228.

[42] L., He, and L. Zhang, "Dynamic priority rule-based forward-backward heuristic algorithm for resource levelling problem in construction project", Journal of the Operational Research Society (2013), Volume: 64, Pages: 1106–1117.

[43] A., Singh," Resource Constrained Multi-Project Scheduling with Priority Rules & Analytic Hierarchy Process", Journal of the Operational Research Society, Procedia Engineering 69 (2014) 725 – 734, Elsevier Ltd.

[44] A., Lova, and P., Tormos, "Analysis of Scheduling Schemes and Heuristic Rules Performance in Resource-Constrained Multiproject Scheduling", Annals of Operations Research, February 2001, Volume 102, Issue 1-4, pp 263-286

[45] T., R., Browining, and A., A."Resource constrained multi-projects Scheduling: Priority Performance Revisited", Int. J., Production Economics, 2014.

[46] A. Atef, M. Abdel-Baset and I. Elhenawy," Composite Heuristic Priority Rules –Based on Tie-Breakers for Scheduling Multiple-Constrained Resource Projects", *under publishing*.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 240 - 245, 2015, ISSN:- 2319–8656

Developing a Crime Mapping GIS System For Law Enforcement: A Case Study of Owerri Metropolis, Nigeria

Onyewuchi Chinedu Nkwachukwu
Department of Computer Science,
Faculty of Physical & Applied Sciences,
University of Port Harcourt,
Rivers State, Nigeria

Dr. Eke Bartholomew O.
Department of Computer Science,
Faculty of Physical & Applied Sciences,
University of Port Harcourt,
Rivers State, Nigeria

Abstract: This paper examines the use of GIS in the development of a crime analysis information system for the Nigeria police. In recent times, criminality has been on the increase with criminals using new and more sophisticated ways to commit crime; resulting to fear and restlessness among the citizens. They police have found it difficult to manage and control these crimes largely due to the obsolete methods and resources they employ in doing so. The purpose of this study is to see how the Nigerian Police Force can adopt the use of crime maps in its operations and reap the benefits. The system will help the police in the analysis of crimes which will lead to crime hotspots identification. Using ArcGIS Software 10.0, we created a digital land use map of crime hotspots in the area and a crime-geospatial database. The results of the spatial analysis and a 500m buffering done on the data shows that areas that are more vulnerable to crime, have no police stations situated around them. This study shows that a GIS based Information system will give the police better insights into crime mapping and analysis which will be a tool to help them effectively manage and combat crime. This study recommends full government involvement in the area of human personnel and infrastructure development for the police to effectively change from the traditional to GIS based ways of combating crime.

Keywords: Crime, Geographical Information System (GIS), Hotspots, Buffering.

1. INTRODUCTION

The incidence of kidnapping, armed robbery, assassinations, terrorism among other crimes has become a daily occurrence in Nigeria. This increase in crime is attributed to the rising number of unemployed youths and the poor economic situation of the country. The government has allocated billions of naira to curb this rising incidence of criminality without much success.

The police force in Owerri has received massive attention under the state governor's administration in an effort to tackle the rising wave of crime in the state. The government has spent billions of naira in the procurement and distribution of equipment such as utility vehicles, communication gadgets, bullet proof vests, as well and the building and renovation of dilapidated police stations. More so, there has been the introduction of Joint Task Force (JTF) operation; an integration of the army, navy and the police force to join hands with the police to root out crime in the state. Among all these efforts from the government, the police are still not efficient in the control and management of crime in the city. This is attributed to the old and manual ways they employ in crime fighting.

Irrespective of government's huge investment in the Nigerian police force by way of personnel training and crime fighting equipment, crime has remained the bane of social and economic wellbeing of the people of Owerri-city making the once peaceful city now a heaven for criminals. The means of getting offenders is very much limited and the police force as it appears is yet not fully exposed to modern technologies that will help them combat crime properly. It is therefore with this in mind that this research work is carried out. This research attempts to explore the analytical approach to crime using the GIS technology in Owerri-City. It is hoped that by adopting this innovative approach in combating crime the spate of crime will be drastically reduced.

In today's modern age where computers have become a way of life, it is therefore imperative for Nigeria Police to migrate from the manual system to a digital system in order to reap the associated benefits like crime mapping, crime hotspots identification and GIS analysis of crime. Case files go missing

in manual systems yet this can be eliminated by computerizing the data storage. Storing crime information in a database would lead to more efficient data sharing within the force. This would mean that investigating officers have access to up to date information from any location where there is a computer. This can only be achieved through the use of crime mapping information systems and geographic information system (GIS). Using questionnaire and qualitative data from the Nigeria police, the objectives of this study include 1) To identify the socio-economic factor responsible for crime, 2) To determine the spatial distribution of Crime type based on relevant data from Police stations. 3) To capture and display the location of police stations and crime hotspots using global positioning system (GPS)

A GIS is a system of hardware and software used for the storage, retrieval, mapping and analysis of geographical data. It is a tool for revealing what is otherwise invisible in geographical information [3]. GIS assisted crime mapping is often employed to understand the geographical distribution of crime, identify crime concentrated area, or hot spots, and facilitate deployment decisions regarding the duration and dosage of intervention programs[4]. The early applications of GIS in policing can be traced back to the 1960s, when it was constrained by the limitation of older computer systems lacking memory and speed [5]. The migration of GIS from mainframe to desktop computers provides the law enforcing agencies with a cost effective option for crime control since hardware and software prices have reduced drastically. Methods of data collection available to law enforcement agents include street investigations, informers and undercover operations. GIS would enhance the analysis of the collected data due to its ability to handle spatial data[6].

2. STUDY AREA

Owerri is the capital of Imo State in Nigeria, set in the heart of the Igbo land and is located within Coordinates: 5.485 °N North of Equator and longitude 7.035°E East of the meridian. Owerri consists of three Local Government Areas namely Owerri Municipal, Owerri North and Owerri West, it has an estimated population of about 400,000 (NPC 2006) and is approximately 40 square miles (100 km^2) in area. The Owerri Slogan is Heartland. It is currently referred to as the entertainment capital of Nigeria and is home to an annual beauty pageant called "Miss Heartland". There are five higher institution of learning in the city namely Federal University of Technology, Federal Polytechnic Nekede, Imo state Polytechnic Umuagwo, Alvan Ikoku Federal College of Education and Imo state University Owerri. Figure 1 shows the map of Owerri city

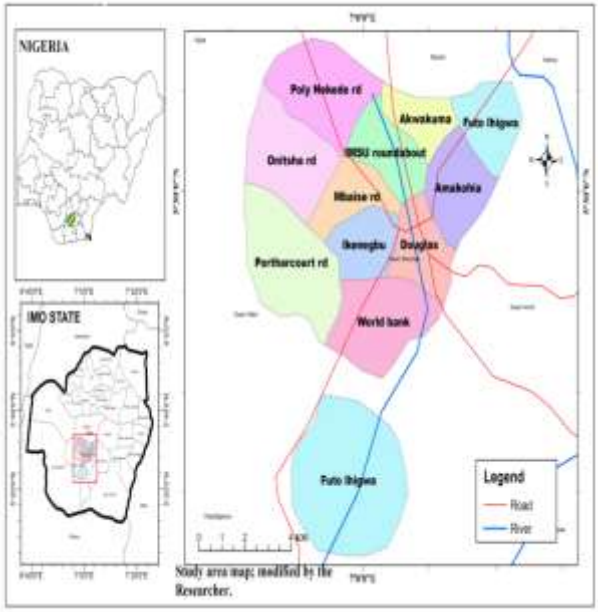


Figure 1 Map of Owerri municipal

3. CONCEPTUAL FRAMEWORK

A crime is an act against a person (for example, murder and sexual assault), or his/her property (for example, theft and property damage) and regulation (for example, traffic violations. Crime is a human phenomenon; therefore it occurs in a place and at a particular time. Crime analysis is important because it helps to identify the different geographic patterns in criminal behavior.

Crime mapping is a procedure using GIS to conduct spatial analysis of crime problems and police-related issues [1]. Crime mapping is a sub discipline of crime analysis which has three main functions. The first one is to facilitate visual and statistical analysis to unravel the spatial nature of crime. The second one is to provide a link to unlikely data sources on common geographic variables and the last one is to produce maps that help to communicate for analysis results [1].A geographic information system integrates hardware, software, and data for capturing, managing, analyzing, and displaying all forms of geographically referenced information [7]. Use of GIS in police departments has proliferated over the past decade. Crime mapping capabilities are useful for police officers patrolling

neighborhoods and investigators trying to solve cases. They can view the recent crime pattern of a neighborhood and query a GIS to search for particular types of crime patterns, such as the location of all recent burglaries within a mile radius of a given intersection[8].

Crime mapping implementation is relatively low in Nigeria. However, numerous examples exist in the developed world eg MAPS (Map-based Analytical Policing System) developed by Rick McKee of the new Zealand police department in the year 2000 to assist his police colleagues in tackling crime. MAPS was predominantly built for and used by Police intelligence analysts to assist in identifying crime patterns and trends, it allowed basic mapping queries to be compiled with a wizard-based formula that could be conducted by all operational police staff. MAPS allow users to build a query, select, display and explore crimes for any location in New Zealand [2]

4. METHODOLOGY AND DATA

Two types of methods were adopted in this study. The first method is the collection of data using questionnaires. Owerri Municipal has 17 wards of which 8 wards were selected randomly and 15 questions were distributed to each. A list of Police stations in Owerri municipal was also collected from the Police State Criminal investigation department. Ten (10) Police stations were sampled and 2 questions were distributed to each giving us a total of 140 respondents; this is to enable all police stations in the study area to be part of the information gathering and for accuracy purpose. The second method is a step by step process (Figure 2) of how GIS can be employed in the creation and analysis of Crime maps.

4.1 *Spatial data used in this study include;*

a) GPS coordinates of crime hotspots in Owerri City
b) GPS Coordinates of police stations in the study area
c) land use and road map of Owerri municipal gotten from Ministry of lands and survey, New-Owerri secretariat, Owerri.

4.1.1 *Attribute data include;*

- Records of types of crime
- Attributes of police stations in the study area
- Attribute of crime hotspot

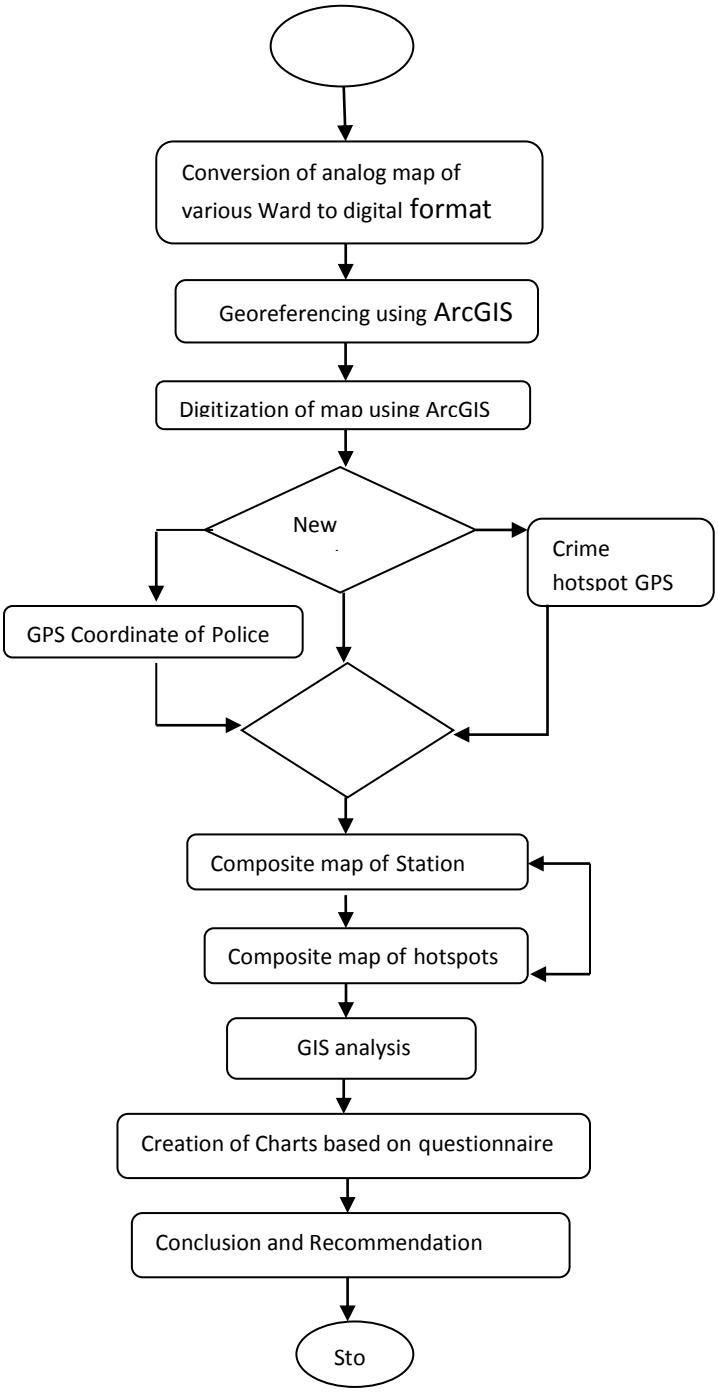


Figure 2 flow chart of GIS map preparation and analysis

5. RESULTS AND ANALYSIS

This study assessed the spatial distribution of crime in Owerri municipal. Data gotten from field work and other relevant sources were analyzed to answer the objective of the study. 140 questioners were distributed, 90% of the respondents are above 27 years of age. All the respondents from the public have lived In Owerri for more than 21 years and they are mainly indigenes which makes them aware of the happenings within the study area.

5.1 Methods employed by police in crime detection

Figure 3 show that the police still rely on the old method of crime detection in the area. 43 of the respondents opined that the use of patrol is commonly used by the police to combat crime, sometimes with patrol van or on foot without uniform, and followed by road block where peoples and car are checked to gather information that may lead to arrest. According to them informant also notify the police from time to time about criminal hide out and crime occurrence.

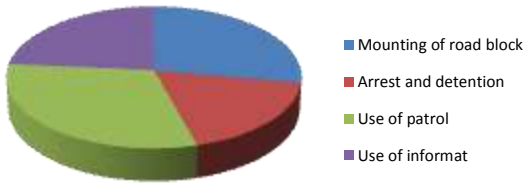


Figure 3 chart showing method employed by police in crime detection

5.2 Crime types and pattern

With GIS, it is possible to map crime by type and to show which crime is prevalent in a particular area. Figure 4 shows the crime record collected from the police headquarters which aided the analysis; From the chart it is very evident that Assault rank the highest type of crime been committed in the city, followed by rape, burglary etc. This is important for police officers because different types of crime need different strategies in controlling and preventing them. For instance, to reduce chances of murder being committed, assault cases must be controlled or prevented. This type of information guides police officers in making decisions on how to allocate specific resources for specific crimes.

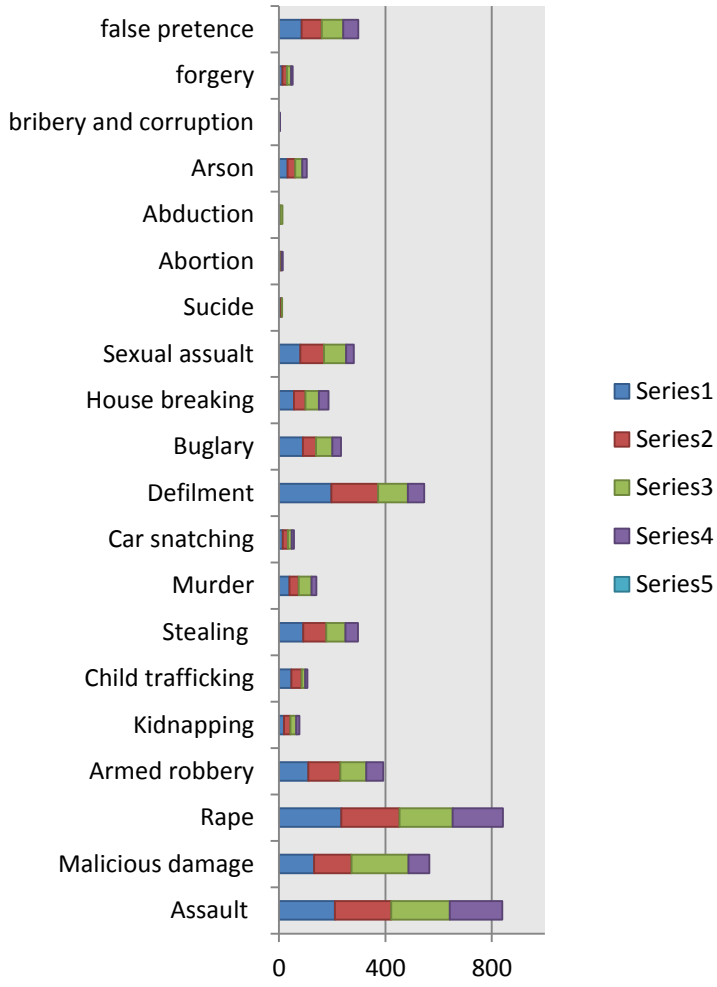


Figure 4 chart of different types of crime

5.3 Socio Economic factors responsible for crime in Owerri

From the public respondents, it was gathered that unemployment, poverty, use of illicit drugs etc are the most common factors responsible for crime in Owerri. Unemployment rank the highest, this is because the youth population which falls between 25-40 years of age is very high while employment is low, jobs are hard to come by, there are few firms in Owerri which are willing to hire the youth, while others who hire often retrench their staff due to poor economy. White collar jobs are readily available in the city, Therefore the only alternative is to indulge in crime to keep body and soul bus

5.4 Crime analysis

This analysis is done in relation to the four basic objectives of the research work. The crime record from police station was obtained from the police headquarter and list of police stations in the city was equally obtained which aided the capturing of the distribution of police station and crime hotspot with the use of global positioning system in the study area. The findings reveals that there are ten police stations in the study area, twelve crime hotspot and several types of crime been committed in the study area of which this will aid the police agency to spread their search and adopt the use of modern technology such as Geographic Information System and Goggle Maps to combat crime and apprehend criminals in the city.

5.4.2 Crime hotspots

GIS helps law enforcement agents plan for crime control. Police officers are able to know where crime is concentrated and focus resources in turn. With GIS it is not only possible to give statistical summaries of crime events per given area but also to visualize the location of crime on a map. Figure 5 shows crime hotspots location indicated with a red circle within the study area. These are the various places in which there are high rates of crime occurrences. Figure 6 shows the GIS analysis result of the crime hotspots. Figure 6, Imsu roundabout, futo Ihiagwa, okigwe road and douglas road has a higher incidence of crime. This is indicated by the yellow color on the map. Crimes are higher in these places because of the location of Owerri main market and the concentration of university students living around these areas. These students belong to cult groups and due to the non location of police station around this area, they find it easy to commit crime and go scout free.

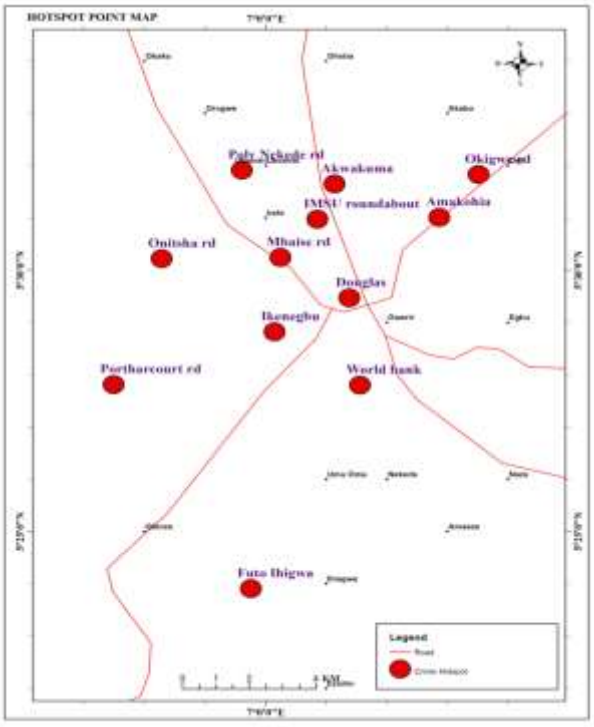


Figure 5 GPS coordinate of Crime hotspot was converted into a point Map and overlaid on the study area Map

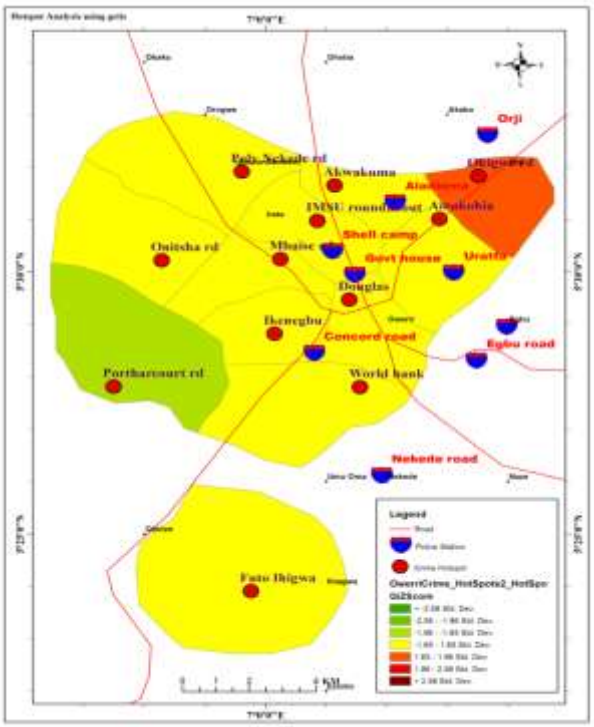


Figure 6 Crime hotspots analysis result

6. CONCLUSIONS

Based on the response from the field survey, it was gathered that the Police force has heard of geographic Information System but are yet to incorporate it into their daily routine, this was very evident from my visit to the state headquarter of police in the city of Owerri. There were few computer system and no single global positioning system (GPS) Tools seen anywhere in all the offices not to talk about the various divisions, all method of handling data were done manually, that is the reason why combating of crime has remained traditional over the years without improvement. There is need therefore to incorporate geographic information system tools into their structure for better performance. More so, crime hotspots are far from Police stations this is the reason why pattern can be predicted and some areas experience repeat victimization. Furthermore from the statistics of crime gathered assault, rape and robbery are the highest in the city. This is as a result of Institutional and commercial land use of which the youth are mostly involved, this is due to unavailability of jobs and as the adage goes an idle mind is the devils workshop. The solution to all these increasing problem lies in the hand of the government and various recommendations has been proffered below

6.1 Recommendations

The government should establish GIS/ICT sections in all the police quarters and procure all the needed tools, and annually train the police force on GIS usefulness in crime detection and how to use the various tools. Police stations should be cited in areas close to crime hotspot. A map of all the hotspots and photographs of suspects should be stored in GIS tools to increase the knowledge of the police in GIS and aid planners in decision making. More so the police force should have a website where citizens can easily alert them on crime scene, suspect and their photographs at will and an accurate GPS coordinates of all roads network should be stored in the GIS tools in order to provide routine instruction to direct patrol vans to crime scene. The use of GIS can assist the police to improve their traditional method by acquiring information on land use and the socio-economic factors that affects crime. Finally, government should make drastic effort to create jobs by empowering the youths who are more associated with crime from time to time.

7. REFERENCES

[1] Boba .R (2005). Role of GIS in the institutionalization of analysis in police departments, Corrections Technology center rocky mountain division pre-conference workshops, Department of Justice, Sage publication Inc Washington, U.S.A.

[2] Andy G. and Barclay J. (2007). Developing geographical information systems and crime mapping tools in New Zealand, Crime Mapping Case Studies: Practice and Research, John Wiley & Sons, Ltd, California, U.S.A.

[3] Longley, Goodchild .M, Maguire .D and Rhind .D (2005). Geographical Information Systems and Science, Wiley, West Sussex: John Wiley and Sons Ltd, California, U.S.A.

[4] Yunus F. (2006) .web based multi participant spatial data entry in crime mapping, Msc Thesis, Middle East Technical University, Department of Geodetic and Geographic Information Technologies, pp 1-31, Ankara, Turkey.

[5] Weisburd, Bernasco .D, Bruinsma .W (2009). Putting Crime in its Place: Units of Analysis in Geographic Criminology. New York: Springer, U.S.A.

[6] Mukumbira S. (2012). Development of a Crime Mapping, Analysis and PredictionTool for Windhoek, Msc Thesis, Polytechnic of Namibia, School Of Information Technology, Department Of Software Engineering, 1 – 41, Namibia.

[7] Escobar, F., Hunter, G., Bishop, I., Zerger, A.: Introduction to GIS. http://www.sli.unimelb.edu.au (2014)

[8] Karen U, Kenneth W, and Vivaswan V. (2003). Crime mapping applications for hawaii's Juvenile justice information system, Department of Urban and Regional Planning University of Hawaii at Mānoa

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 246 - 252, 2015, ISSN:- 2319–8656

THE CRITICAL ORGANIZATIONAL FACTORS OF E-GOVERNMENT IN KENYA

Godfrey Kyalo Makau,
Department of Business & Social Sciences,
Jomo Kenyatta University of Agriculture & Technology (JKUAT)
Nairobi, Kenya

Elijah, I. Omwenga
Department of computer Science, School of Computing and Informatics,
University of Nairobi,
Nairobi, Kenya

Nzomo Daudi N.
Department of Finance and Accounting, School of Business
University of Nairobi,
Nairobi, Kenya

Abstract: eGovernment focusses on the use of technology to achieve levels of improvement in various areas of government, transforming the nature of politics and relations between the government and citizens. However, in Kenya, just like in other developing nations, many eGovernment projects have either stalled or failed to meet their objectives due to some key organizational factors. This study therefore highlights critical organizational factors affecting eGovernment projects and the nature of their relationships with eGovernment performance. The study employed cross-sectional survey design. Targeting the entire 18 eGovernment projects implemented through the Information Communications Authority of Kenya since 2005. Both primary and secondary data was collected and analyzed based on response from 217 respondents out of the 300 who participated (72% response rate). At the end, it emerged that out of the various organizational factors hypothesized to predict eGovernment projects Performance, only organizational structure, prioritization of deliverables, and organizational culture are critical in Kenyan context. Others identified in previous studies such as future needs of the organization, power distribution, structure, information system strategy alignment, prioritization of deliverables, and training were also important but not critical.

Keywords: **eGovernment; Organizational; Critical Factors; Performance; Projects.**

1. INTRODUCTION

1.1 Background of the Study

Globalization has, in the past decade, perpetuated need for inventing and applying technological solutions to service delivery in governments (Cordella & Bonina, 2012). According to the recent UN e-government development index report (EGDI), among the 193 countries surveyed on online provision, 190 had online services. The report concluded that Information and Communication Technologies (ICTs) is continuously gaining recognition due to its ability to promote economic growth and development at government, business and citizens levels (UN, 2014).

In developing countries, the idea that implementation of eGovernment ensures modernization of the public sector, coupled by the current quest for citizenry empowerment and the demand for e-participation changes sweeping the modern world, many governments have continued developing, implementing and improving their strategies to transform government services using eGovernment (Cordella & Bonina, 2012; UN, 2014). However, despite a lot of interest being drawn into eGovernment, the field is criticized for not having a common definition of e-Government (Hu et al. 2009) and also generally lacking clarity and rigor about the factors contributing to its implementation alongside poor treatment of generalization (Heeks and Bailur, 2007). Furthermore, recently, researchers and other stakeholders affected by the e-Government innovation's impacts have started raising concerns on its high reported failure rates (Heeks and Bailur, 2007).

In an effort to comprehensively define eGovernment and address the above concerns, three perspectives have dominantly emerged: technological, organizational and environmental perspectives of eGovernment. Ahmad et al., (2012) observe that the different conceptualization indicate that eGovernment is a broad concept whose evaluation and measurement, demands exhaustive assessment of each of the perspectives. This study centers on the organizational perspective of eGovernment which focusses on the use of technology to achieve levels of improvement in various areas of government, transforming the nature of politics and relations between the government and citizens (Ahmad, et al., 2012).

The organizational context is a necessity for any eGovernment project as it helps to explain the tangible and non-tangible organizational factors (Ahmad et al., 2012). Researchers have confirmed complex relationships between the organizational factors and eGovernment Adoption, use, and hence eGovernment success. Organizational aspects such as the organizational structure, the presence of innovation-enabling processes such as informal communication and strategic behaviour of top management, quality of human resource, firm size, amount of slack resources of the organization are vital for eGovernment success (Tornatzky and Fleischer, 1990). Consequently, ensuring successful adoption of eGovernment initiative calls to assessment of the contribution each of these organizational aspects have of the performance of eGovernment projects.

1.2 Statement of the Problem

In Kenya, just like in other developing nations, many eGovernment initiatives fail or do not achieve their goals due to some key organizational factors. Currently, Kenya ranked number 119 globally and number seven in Africa, after Seychelles, Mauritius and South Africa who are ranked first, second and third respectively, in eGovernance. Despite their low ranking in eGovernance, African governments support

eGovernment initiatives and appreciate its contribution to the government agenda (Mutula, 2008).

However, after the recent realization that for eGovernment projects in developing and transitional countries, 35% were total failures, 50% were partial failures and only 15% were successful(Heeks, 2003;Schedler and Schmidt, 2004), more questions are raised on the factors affecting success of eGovernment projects. This study seeks to answer this question by assessing the organisational critical factors to the performance of eGovernment in Kenya

1.3 Objectives of the Study

i) Establish the organisational factors predicting eGovernment projects performance in Kenya;

ii) Examine the nature of relationships between the organisational factors and eGovernment performance

2. LITERATURE REVIEW

Research has helped to explain the tangible and non-tangible organisational factors influencing eGovenrment projects (Ahmad et al., 2012). These include: the organizational readiness in terms of business strategic planning; technical infrastructure; management systems and structures; top management; and quality of human resource (Ahmad et al., 2012). Each of these can either impede or promote eGovernment project performance (Bjorck, 2004). A bureaucratic organisation with a conservative culture raises the issue of resistance to change from new innovation (Ahmad et al., 2012). Therefore, transformation and re-engineering of government processes and activities must be embraced for successful eGovernment (Basu, 2004). Many studies attribute eGovernment failure to a variety of reasons, including: lack of executive and top managers' commitment; employees' resistance to change; lack of skills and training programs; lack of awareness and conceptual understanding; old and inflexible management systems (Karlson, et al., 2012). The success of e-government initiatives will also depend upon the developed legal and regulatory framework for their operations (Basu, 2004). In general, UN (2014) found out that institutional and organisational weaknesses in the design of policies, the organization of programs and stakeholder coordination jeopardize the long term development of e-government practices.

Favourable organizational structures lead to higher system usage and consequently successful eGovernment implementation outcome (Baker, 2011). The implementation of a new ICT environment may result in some employees losing their authority and power over traditional business processes hence triggering resistance (Doherty and King, 2005). Therefore, favourable power distribution practices would reduce resistance thereby resulting to higher system usage and consequently successful eGovernment implementation outcome. A comprehensive eGovernment strategy is also essential to effectively and efficiently deliver the successful implementation of online public services (Iran et al. 2006). The presence of a well synchronized information system strategy with clearly spelled out future organizational needs to be achieved through ICTs would lead to higher system user support and consequently successful eGovernment implementation outcome (Lee et al., 2008). The paradigm shift and change of culture that is introduced by eGovernment may result in some resistance and failure. There is therefore need for continued training in order to realize eGovernment success (Iran et al., 2006). According to the UN (2014) countries need to focus on building human capital, including ICT literacy and on bridging infrastructure gaps to provide an enabling environment for e-government development. Visionary strategies and practical implementation plans should follow for effective deployment of sustainable online services (UN, 2014). Oreste, et al. (2005), found that funding facilitates the infrastructure (such as building, technology, human resources) necessary for eGovernment implementation.

3. RESEARCH METHODOLOGY

3.1 Research Design

This study adopted a descriptive cross-sectional survey design. According to Cresswell and Clark (2007), a combined descriptive cross-sectional survey research design is used when seeking to gather information, summarize, present and interpret it for the purpose of clarification. This design was therefore chosen as the study sought personal views, opinions, attitudes, and perceptions about eGovernment critical factors and project performance status.

3.2 Target Population

The study targeted the entire 18 eGovernment projects that had been in place since 2005 and which were implemented through the Directorate of eGovernment (now Communications Authority of Kenya) in Kenya government. The respondents therefore included all the eGovernment project implementers and eGovernment service consumers of the eGovernment services in Kenya.

3.3 Data Collection

The study collected both primary and secondary data. Primary data were collected using survey questionnaires, although interviews and observations were also employed where necessary and possible. Secondary data sources included journals, books and articles addressing the objectives of this study.

3.4 Operationalization of Variables

This study employed quantitative measures using a 4-point likert scale on technological factors indicators as defined by Agresti (2002). The operationalization and measurements of the variables in this study is as shown in Table 1 below.

Table 1 Operationalization and measurements of the variables

THE ORGANISATIONAL FACTORS INFLUENCING EGOVERNMENT PROJECT IMPLEMENTATION, ADOPTION, & E-SERVICE USE		
Construct	**Construct Domains**	**Measures**
Organisational Factors	Status before delivery of required results; Organizational structure; Power distribution; System Structure; Information system strategy alignment; Prioritization of deliverables; Resistance to change; Human Capacity; Management skills; Future needs of the organization; Organizational culture; Training; Cooperation/Collaboration;	Nominal & 4-point likert

	Smooth Processes	

3.5 Data analysis

Data analysis was performed at both descriptive and inferential statistical analysis levels using a mixture of tools available in SPSS. Descriptive statistics involved use of frequency tables, percentages and charts and other measures of variable associations (De Vaus, 2001). Inferential statistical analyses included various correlation and regression tests (Saunders et al., 2003). Factor analysis test were used to group and detect opinions/perception disagreement or changes to help in assigning variables into TOE, failure and success groupings based on their agreements.

4. RESULTS AND DISCUSSIONS

The results are based on response from 217 respondents out of the 300 who participated (72% response rate). Of the 217 respondents 52 were eGovernment project implementers while 165 were eGovernment service consumers.

4.1 Organizational eGovernment Projects Success and Failure Factors

Nine statements on four point likert scale were used to assess the organizational factors affecting eGovernment project implementation. These statements were presented to eGovernment project implementers alone because they are the only lot that directly engaged with the projects implementation process.

a. **Overall Judgement on Project Delivery of Results**- Majority of the study participants (57%) disagreed with the statement that the project has taken too long to show any meaningful results.

b. **Organizational structure issues**-Majority of the respondents (79%) supported the statements there exist proper allocation of work roles and administrative mechanisms to conduct, coordinate, and control eGovernment implementation work activities.

c. **Power distribution**-Majority of the respondents (61%) agreed that there is resistance for fear of government employees losing their authority and power over traditional business processes.

d. **Structure**- Majority of the respondents (68%) agreed with the statement that due to the horizontal and vertical computerized linkages, there is a great level of flexibility in task-performing to accommodate the new eGovernment system procedures.

e. **Information system strategy alignment**- Majority of the respondents (57%) agreed that there is proper alignment of strategies between different players for the eGovernment systems success.

f. **Prioritization of deliverables**- Majority of the respondents (69%) agreed that there is proper prioritization of deliverables to ensure the most strategically significant services are managed and delivered appropriately in time.

g. **Future needs of the orga**nization-Majority of the respondents (37%) disagreed with the statement that there exists a strategic plan for eGovernment systems implementation that is being strictly followed to ensure the implementation process caters for the future needs of our organization.

h. **Organizational culture**-Majority of the respondents (59%) disagreed with the statement that there has been a good organizational environment that encourages smooth and total transformation from manual to eGovernment culture.

i. **Training issues**- Majority of the respondents (52%) agreed with the statements that there has been enough training for employees and managers to get familiar with working under new eGovernment system circumstances. The details are as shown in table 2 below.

Table 2 Organizational Factors Descriptive Analysis Results

	Total	S. D	D.	A.	S. A.
Overall judgment	The project has taken too long to show any meaningful results	19%	38%	28%	15%
Organizational structure	There exist proper allocation of work roles and administrative mechanisms to conduct, coordinate, and control eGovernment implementation work activities.	4%	18%	65%	14%
Power distribution	There is resistance for fear of government employees losing their authority and power over traditional business processes.	8%	31%	41%	20%
Structure	Due to the horizontal and vertical computerized linkages, there is a great level of flexibility in task-performing to accommodate the new eGovernment system procedures	12%	20%	60%	8%
Information system strategy alignment	There is proper alignment of strategies between different players for the eGovernment systems success	10%	33%	45%	12%
Prioritization of deliverables	There is proper prioritization of deliverables to ensure the most strategically significant services are managed and delivered appropriately in time.	6%	25%	56%	13%

Future needs of the organization	There exists a strategic plan for eGovernment systems implementation that is being strictly followed to ensure the implementation process caters for the future needs of our organization.	8%	29%	49%	14%
Organizational culture	There has been a good organizational environment that encourages smooth and total transformation from manual to eGovernment culture.	10%	49%	33%	8%
Training	There has been enough training for employees and managers to get familiar with working under new eGovernment system circumstances.	13%	35%	42%	10%

4.2 Test of Associations (Correlation) and Factor Analysis for the Organisational Factors

The study also sought to establish the specific factors predicting eGovernment projects performance from the collected data through tests of associations. This was achieved through correlations and factor analysis. The composite variables emerging from factors analysis were then used in regression analysis, presentation, interpretation and discussions of the outcomes.

The goal of factor analysis was to reduce "the dimensionality of the original space and to give an interpretation to the new space, spanned by a reduced number of factors (Darlington, 2004). Guttman-Kaiser rule was applied in retaining only the factors whose eigenvalues were larger than 1 and in total accounted for over 0.5 of the variance (Field 2000). Therefore, items with variance loadings of over 0.6 were retained for further analysis as recommended by Rietveld & Van Hout(1993).

Correlation results

Correlation was first done on all the data items under organizational factors and only those that significantly correlated to each other were further reduced into few principal components. In the end, the factor reduction split the data items into two significant factors considered to significantly affect eGovernment implementation, adoption and use in the research. Results from correlations showed that the overall judgment on the performance of eGovernment (success or failure of eGovernment projects) did not correlate significantly with most of other items apart from two variables, Power distribution and Organizational culture, and was therefore discarded at this stage while the rest were reserved for use in running the factor analysis. This is as shown in table 3 below.

Table 3 Organizational Factors Correlations Contingency Table Results

		Overall judgment	Organizational structure	Power distribution	Structure	Information system strategy alignment	Prioritization of deliverables	Future needs of the organization	Organizational culture	Training
Overall judgment	Pearson Correlation	1	.193	.390**	-.094	-.082	-.120	.125	.310*	-.135
	Sig. (2-tailed)		.174	.006	.516	.569	.398	.381	.027	.339
	N	53	51	49	50	51	52	51	51	52

Factor Analysis Results

The table 4 below shows the eigenvalues (variances of the principal components) associated with each linear component (factor) before extraction, after extraction and after rotation. The rotations converged in two iterations with two significant components with Eigenvalues accounting for 60.494% of the variance explained. Being above the threshold of 50% it indicated that the two-component factor model derived fitted the data appropriately.

Table 4 Organizational Factors Total Variance Explained Results

Component	Initial Eigen values			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.616	45.202	45.202	3.616	45.202	45.202	2.671	33.394	33.394
2	1.223	15.291	60.494	1.223	15.291	60.494	2.168	27.100	60.494
3	.968	12.104	72.597						
4	.739	9.238	81.835						
5	.516	6.450	88.285						
6	.451	5.637	93.922						
7	.305	3.809	97.731						
8	.182	2.269	100.000						

Extraction Method: Principal Component Analysis.

Items loading greater than 0.6 for each component combined to form the two principal components and the variables that

clustered into each are shown in table 5 below. Cronbach alpha analysis for reliability showed internal consistency. Therefore, the eight items were used in further analysis.

Table 5 Organizational Factors Rotated Component Matrix Results

	Component	
	1	2
Organizational structure	.374	.625
Power distribution	-.795	-.030
Structure	.775	.089
Information system strategy alignment	.514	.356
Prioritization of deliverables	.760	.325
Future needs of the organization	.067	.889
Organizational culture	.178	.800
Training	.651	.327
Cronbach's Alpha	.712	.657

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

4.3 Correlation Analysis between Organizational Factors and the Project Performance

The table 6 below displays the correlation between the individual factors measuring organizational factors and the project performance. From the results, it emerged that only Organizational structure, Prioritization of deliverables and Organizational culture had positive significant relationships with project performance hence reserved for entry into the logistic regression model while the rest (highlighted below) were not, and hence eliminated at this stage. These findings support the findings of the latest survey by the UN on eGovernment which concluded that institutional and organisational weaknesses in the design of policies, the organization of programs and stakeholder coordination jeopardize the success of e-government projects (UN, 2014). Additionally, the results of the study support Ahmad et al., (2012) findings that a bureaucratic organisation culture with a conservative culture raises the issue of resistance to change to new innovations. Therefore, transformation and re-engineering of government processes and activities must be embraced for successful eGovernment (Basu, 2004). It also supports Karlson et al., (2012) finding that old and inflexible management systems with resistance to change would cause eGovernment project failures. The findings also concur with those of Baker (2011) who found that favourable organizational structures lead to higher system usage and consequently successful eGovernment implementation outcome. However the results contradict Doherty and King (2005) findings that favorable power distribution practices would reduce resistance thereby resulting to higher system usage and consequently successful eGovernment implementation outcome. It also contradicts Iran et al. (2006) findings that the paradigm shift and change of culture that is introduced by eGovernment may result in some resistance and failure and therefore there is need for continued training in order to realize eGovernment success is also not supported.

Table 6 Results of Correlation between Organizational Factors and Project Performance

		Overall judgment	Organizational structure	Power distribution	Structure	Information system strategy alignment	Prioritization of deliverables	Future needs of the organization	Organizational culture	Training
Project performance	Pearson Correlation	.136	.419**	-.249	.218	.199	.370*	.161	.522**	.150
	Sig. (2-tailed)	.361	.004	.103	.150	.184	.010	.284	.000	.313
	N	47	46	44	45	46	47	46	46	47

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

4.4 Organizational Factors Logistic Regression

This procedure was conducted to predict the probability that a participant would give his/her eGovernement project a success performance judgment (rating) given presence/nature of critical organisational factors. Therefore, the outcome variable entered in the model was successful eGovernment project performance which was measured by a binary question Yes/No (0 = failure and 1 = success), and the predictor variable entered in the model was organizational Factors (X). A regression model predicting the logit, i.e, the natural log of the odds of success of e-government project or failure was then run. Table 7 below shows the SPSS output for the initial model which includes only the intercept (the constant). Given the base rates of the two success of e-government project performance options (1 and 0), the system correctly grouped 62.2% of the respondents cases as having reported success of e-government project with only 37.8% of the cases reporting failure of e-government project. No other information was printed out as there were no predictor variables at this stage of the logistic regression process. Therefore, in order to achieve more information details, the best strategy was to perform predictions for every case that the subject will report successful performance of the e-government projects. Using this strategy, a respondent chosen at random for any random project would be correct 62.2 % of the times in judgment translating to 0.62 chances of judging an eGovernment project as successful when it is correctly so.

Table 7 Classification for the Initial Model

Observed		Predicted		
		Project judgment		Percentage Correct
		Yes	No	
Step 0	Project judgment Yes	28	0	100.0
	No	17	0	.0
	Overall Percentage			62.2

a. Constant is included in the model.

b. The cut value is .500

Further, table 8 below shows that the 2 Log Likelihood function would drop by 4.963 if a single unit of X predictor was added to the model (which already has the intercept) and the drop is significant (Pvalue = 0.026).

Table 8 Variables not in the Equation

			-2 Log likelihood score	df	Sig.
Step 0	Variables	X_2	4.963	1	.026
	Overall Statistics		4.963	1	.026

Table 9 below shows the block1 outputs where the SPSS added the independent variables Organizational Factors (X) as the predictor. Omnibus Tests of Model Coefficients gives us a Chi-Square of 5.523 on 1 df which is significant as the P-value (.019) was less than 5% (.05). This is a test of the null hypothesis that adding the independent variable to the model did not significantly increase the likelihood of the respondents to give an eGovernment project a success outcome when it is correctly so. A positive and significant Chi-Square statistic indicates that there is a positive relationship between X and the eGovernment project success performance.

Table 9 Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	5.523	1	.019
	Block	5.523	1	.019
	Model	5.523	1	.019

Under Model Summary printed in table 4.59 below, the -2 Log Likelihood statistics is shown as 54.144. This statistic measures how poorly or well the model predicts the judgment decisions, the smaller the statistic the better the model. The Cox & Snell R^2 value of .116 implies that only 11.6% variation in the dependent variable is explained by the model. Alternatively, the Nagelkerke R^2 output of 0.157 indicates that a larger figure of 15.7% in the dependent variable is explained by the model.

Table 10 Model Goodness of Fit Tests Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	54.144[a]	.116	.157

Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Table 11 below shows the Hosmer-Lemeshow statistic, which tests the null hypothesis that there is a linear relationship between the predictor variable and the log odds of the outcome variable. A chi-square statistic was then computed comparing the observed frequencies with those expected under the linear model. A non-significant chi-square indicates that there exists a linear relationship and therefore the data fits the model well (Pvalue = 0.533).

Table 11 Hosmer and Lemeshow Linearity Test

Step	Chi-square	df	Sig.
1	5.088	6	.533

From table 12 results, it is noted that overall success rate in classification has improves from 62.2 – 64.4 percent after adding the independent variable.

Table 12 Classification for the Final model

Observed			Predicted		
			Project judgment		Percentage Correct
			No	Yes	
Step 1	Project judgment	No	24	4	85.7
		Yes	12	5	29.4
	Overall Percentage				64.4

a. The cut value is .500

Table below 13 below shows the Regression Coefficients and Odds Ratio. The Wald Chi-Square statistic, which tests the unique contribution of each predictor, holding other predictors constant is also given. The output indicates that the predictor X_2 relationship with the outcome meets the conventional .05 standard for statistical significance. It's 3.211 odds ratio statistic indicates that the chances of eGovernment project success judgment are increased by more than triple for each one point increase in respondent's exposure to or interaction with eGovernment project organizational Factors and the increase is significant (Pvalue =.037).

Table 13 Variables in the Model Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	X2	1.167	.561	4.331	1	.037	3.211
	Constant	-4.380	1.932	5.139	1	.023	.013

a. Variable(s) entered on step 1: X2.

5. CONCLUSION

The results of the study indicate that among the organizational factors hypothesized to influence the success of eGovernment projects, only organizational structure, prioritization of deliverables, and organizational culture emerged to have positive significant relationships with project performance in Kenya. Future needs of the organization, power distribution, structure, information system strategy alignment, prioritization of deliverables, and training had insignificant relationships with performance.

6. RECOMMENDATIONS

Based on the results from the study, eGovernment projects implementers should ensure that the critical success factors that include efficient and effective organisational structure, organisational culture and priorities of the deliverables are availed, to minimise the cases of failure in implementing eGovernment projects.

To researchers and academicians the study recommends that replica studies be done on Kenya's eGovernment projects with larger samples of project implementers for purposes of generalization and eGovernment projects critical success and failure factors theory building.

7. REFERENCES

[1] Agresti, A. (2002). *Categorical Data Analysis* (2nd Edition ed.). RW: Mee.

[2] Ahmad, M. O. Markkula, J. and Oivo, M. (2012). *Factors Influencing the Adoption of eGovernment Services in Pakistan.* Paper presented at the Proceedings of the 9th European, Mediterranean and Middle Eastern Conference on Information Systems, Munich, Germany.

[3] Barker, J. (2011). The technology–organization–environment framework. In Information Systems Theory: Explaining and Predicting our Digital Society (Dwivedi, Y., Wade, M. and Schneberger, S. Eds.), pp. 231-246, Springer, New York.

[4] Basu, S. (2004). E-government and developing countries: an overview. *International Review of Law, Computers & Technology, 18*, 109-132.

[5] Bjorck, F. (2004). Institutional theory: A new perspective for research into IS/IT security in organisations.

[6] Cordella, A. & Bonina C.M. (2012) A public value perspective for ICT enabled public sector reforms: A theoretical reflection. *Government Information Quarterly*. Volume 29, Issue 4, pp. 512–520.

[7] Cresswell, J., & Clark, P. (2007). *Designing and Conducting Mixed Methods Research*. Carlifonia: SAGE Publications.

[8] De Vaus, D. A. (2001). *Research design in social research* London: SAGE.

[9] Doherty, N. F., & King, M. (2005). From technical to socio-technical change: tackling the human and organizational aspects of systems development projects. *European Journal of Information Systems, 14*, 1-5.

[10] Field, A. (2009). *Discovering Statistics using SPSS* (3rd ed ed.). London: Sage

[11] Heeks, R. & Bailur, S. (2007) Analyzing e-government research: Perspectives, philosophies, theories, methods, and practice. *Government Information Quarterly*, 24, 243-265.

[12] Heeks. (2003). Most egovernment-for-development projects fail: how can risks be reduced? : Institute for Development Policy and Management.

[13] Hu, P., Brown, S.A., Thong, J., Chan, F., & Tam, K.Y. (2009). Determinants of service quality and continuance intention of online services: The case of eTax. *Journal of the American Society of Information Science and Technology*, 60(2), 292-306.

[14] Irani, Al-Sebie, M., & Elliman, T. (2006). *Transaction Stage of e-Government Systems: Identification of its Location & Importance.* Paper presented at the the 39th Hawaii International Conference on System Science, Hawaii

[15] Karlsson, F., Holgersson, J., Söderström, E., & Hedström, K. (2012). Exploring user participation approaches in public e-service development. *Government Information Quarterly, 29*(2), 158–168.

[16] Lee, Irani, Z., Osman, I. H., Balci, A., Ozkan, S., & Medeni, T. D. (2008). Toward a reference process model for citizen-oriented evaluation of e-Government services. *Transforming Government: People, Process and Policy, 2*(4), 297-310.

[17] Mutula, S. M. (2008). Africa's e-government status with developed and transitional nations. *Information Management & Computer Security 16* (3), 235-250.

[18] Oreste, S., Chesi, F., & Pallotti, M. (2005, 7–9 June). *E-Government: Challenges and Opportunities.* Paper presented at the Italy-XIX Annual Conference, Florence, Italy.

[19] Saunders, M., Lewis, P., & Thornhill, A. (2003). *Research methods for business students* (3rd edition ed.). Harlow: Prentice Hall.

[20] Schedler, K., & Schmidt, B. (2004). Managing the e-government organization. *International Public Management Review* 1-20.

[21] Tornatzky, L. G., & Fleischer, M. (1990). *The Process of Technology Innovation*. Lexington, MA: Lexington Books.

[22] UN. (2014). United Nations E-government Survey 2014. Retrieved Accessed: 10 January, 2014 http://unpan3.un.org/egovkb/Portals/egovkb/Documents/un/2014-Survey/E-Gov_Complete_Survey-2014.pdf

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 253 - 257, 2015, ISSN:- 2319–8656

Online Signature Authentication by Using Mouse Behavior

Jitender Kumar
Bharath University
Chennai ,India

S Goutham
Bharath University
Chennai,India

Amitabh Kumar
Bharath University
Chennai, India

Abstract: Several large-scale parole leakages exposed users to associate unprecedented risk of speech act and abuse of their data. associate inadequacy of password-based authentication mechanisms is turning into a serious concern for the complete data society. carries with it 3 major modules: (1) Mouse–Behavior dynamics Capture, (2) Feature Construction, and (3) coaching or Classification. the primary module serves to make a taking mouse behavior user signs. The second module is employed to extract holistic and procedural options to characterize mouse behavior and to map the raw options into distance-based options by exploitation numerous distance metrics. The third module, within the coaching section, applies neural network on the distance-based feature vectors to reckon the predominant feature elements, then builds the user's profile employing a one-class classifier. within the classification section, it determines the user's identity exploitation the trained classifier within the distance-based feature exploitation NN. A four Digit OTP is generated to the user's email ID. The user are going to be giving the '2' digit OTP and therefore the server are going to be giving balance '2' digit OTP. Users '2' digit OTP is verified by the server and contrariwise.
Keywords: mouse behaviour signatures; Biometric authentication; Verification; Template; String matching; Feature detection

1 INTRODUCTION

The quest for a reliable and convenient security mechanism to authenticate a computer user has existed since the inadequacy of conventional password mechanism was realized, first by the security community, and then gradually by the public.[3] As data are moved from traditional localized computing environments to the new Cloud Computing paradigm (e.g., Box.net and Drop box), the need for better authentication has become more pressing.[1]

Recently, several large-scale password leakages exposed users to an unprecedented risk of disclosure and abuse of their information. These incidents seriously shook public confidence in the security of the current information infrastructure; the inadequacy of password-based authentication mechanisms is becoming a major concern for the entire information society. Of various potential solutions to this problem, Mouse dynamics measures and assesses a user's mouse-behavior characteristics for use as a biometric.[2]

Compared with other biometrics such as face, fingerprint and voice, mouse dynamics is less intrusive, and requires no specialized hardware to capture biometric information. Hence it is suitable for the current Internet environment. When a user tries to log into a computer system, mouse dynamics only requires her to provide the login name and to perform a certain sequence of mouse operations. Extracted behavioral features, based on mouse movements and clicks, are compared to a legitimate user's profile.[4] A match authenticates the user; otherwise her access is denied. Furthermore, a user's mouse-behavior characteristics can be continually analyzed during her subsequent usage of a computer system for identity monitoring or intrusion detection.

2. EXISTING SYSTEM

In the existing system, there many password leakages exposed users to an unprecedented risk of disclosure and misuse their information. These types of password-based authentication mechanisms is becoming a major concern for varieties of Security based applications.[8] Also some attacks namely called, password guessing attacking has become more concern for the users, while accessing the some of the sensitive application like Bank transaction, Train Booking and Online Shopping.[6]

3. PROPOSED SYSTEM

We are implementing the proposed System which is consisting of three major modules: (1) signature dynamics, (2) fluctuate data define , and (3) design to different pattern **.** In the First Module, we'll create a user defining data, and to capture and information data. The second module is used to extract holistic and procedural features to characterize mouse behavior and to map the raw data manipulate by the user by using various distance metrics. The third module, in the defining the different section, applies on taking assign vectors to compute the predominant feature components, and then builds the user's profile using .

4. ARCHITECTURE DIAGRAM

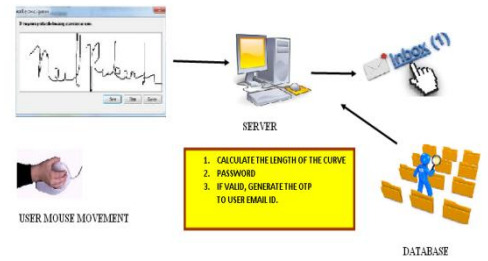


Fig 1. Architecture diagram

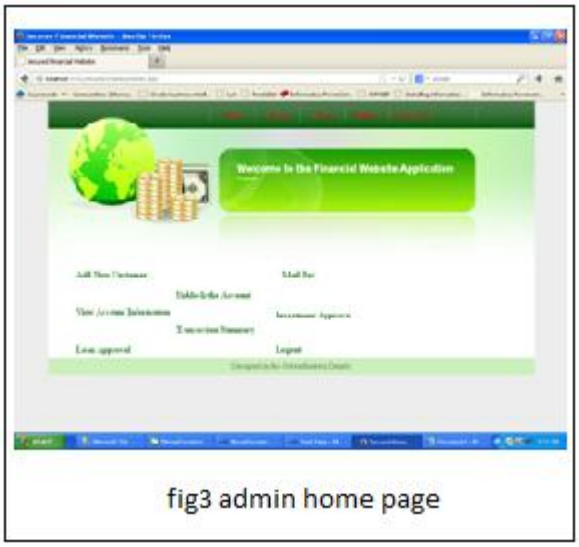


fig3 admin home page

5. MODULES

5.1 CLIENT OF THE NETWORK:

In this module we are implementing the Client interface by which the Client can interact with the Application. To access the Application, the Client want to the register their details with Application Server. They have to provide their information . This information will stored in the database of the Application Server. The User is allowed to the access the application only by their provided Interface.

5.2 SERVER:

The Server will monitor the entire User's information in their database and verify the if required. Also the Server will store the entire User's information in their database. Also the server localize itself. It be the data load in its database. they access the Application. So that the server are to taking by the user guide to server loaded data.

fig 2 admin login

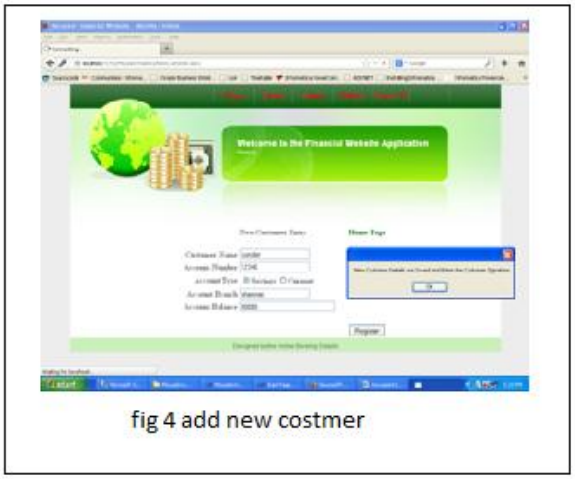


fig 4 add new costmer

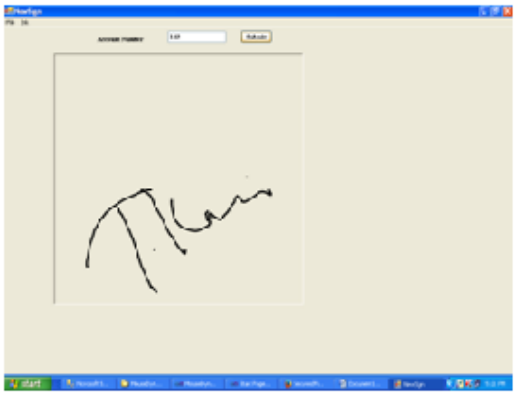


fig 5 costmer sign

fig 7 new user sign

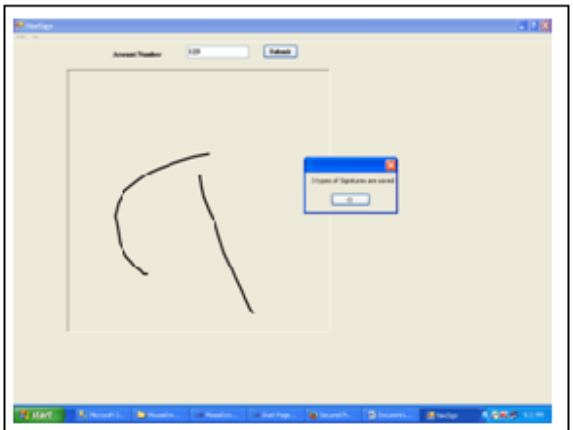


fig 6sign data to server

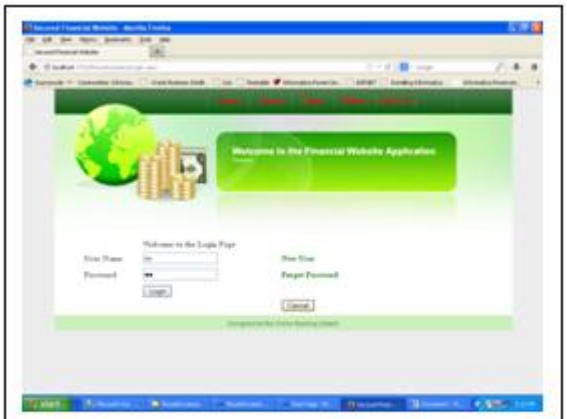


fig 8 user log in

5.3 LEARNING PHASE:

In this phase, the we'll train the system according to identify the User's Signature by using the following modules. (1) signature dynamics, (2) fluctuate data define , and (3) design to different pattern.. The first module serves to create a mouse-operation task, and to capture and user defining data. The second module is used to extract holistic and procedural features to characterize mouse behavior and to map the raw features into distance-based features by using various distance metrics. The third module, in the designing different phase, applies neural network on the data set to vectors compute.

5.4 VERIFICATON PHASE:

In the Verification Phase, the Server will verify the User when they are login into their account. The Servera will provided by the User while login with the Signature provided by the User when they provided during the Training Phase. If the signature is not matched, then the Server will not allow the User to access their account.

fig 9 verification sign

Fig9User authecate

fig10 verifigning user

5.5 CHECK MAIL/OTP VERIFICATION:

Once the User provided their signature correctly, the Server will generate the Session Key using Secure Random Number generation algorithm and send it to the User Email id. Once the User received their session key in their Email id, they have to provide the first '2' digits of the session key and the server will verify the next '2' digits of the session key. Once the Session key is verified by the Server, the User is allowed to access their account.

fig 11token verification

fig 12 user defing

6. CONCLUSION

Mouse dynamics is a newly emerging behavioral biometric, which offers a capability for identifying computer users on the basis of extracting and analyzing mouse click and movement features when users are interacting with a graphical user interface. Many prior studies have demonstrated that mouse dynamics has a rich potential as a biometric for user authentication. In this study, we highlighted the challenges faced by mouse-dynamics-based user authentication, and we developed a simple and efficient approach that can perform the user authentication task in a short time while maintaining high accuracy. Holistic features and procedural features are extracted from the fixed mouse-operation task to accurately characterize a user's unique behavior data. Then distance-based feature construction and parametric eigenspace transformation are applied to obtain the predominant feature components for efficiently representing the original mouse feature space. Finally, a one-class classification technique is used for performing the user authentication task.

7. ACKNOWLEDGMENT

We would like to express our sincere gratitude to our respected Chancellor Dr.J.SUNDEEP AANAND and Managing Director Dr.SWETHA AANAND for their valuable support and encouragement in technological upgrades and novel projects.

We take great pleasure in expressing our sincere thanks to our Pro-chancellor Dr.K.P.THOOYAMANI for backing us in this project. We take great pleasure in expressing our sincere thanks to our Vice-chancellor Dr.M.PONNAVAIKO for backing us in this project.

We thank our Dean Engineering Dr.J.HAMEED HUSSAIN , for providing sufficient facilities for the completion of this project.

We express our sincere thanks to our Dean-Research Dr.KATHIR VISWALINGAM and our Dean-CSE Dr.A.KUMARAVEL and Head of the Department Dr.K.P.KALIYAMURTHIE and Project Co-ordinator Dr.C.NALINI for their kind permission to carry out this project

8. REFERENCES

[1] A. A. E. Ahmed and I. Traore, "Anomaly intrusion detection based on biometrics," in *Proc. IEEE Information Assurance Workshop*, West Point, NY, 2005, pp. 452–453.

[2] Ahmed, A A E & Traore .I (2005) detection computer instruction using behavior biometric proc of 3rd ann conf on privacy security and trust Canada(pp91-98)

[3] Ahn,L v. Blum, M & Langford. J(2004)how lazy cryptography do Al communication of the ACM 47(2),56-60 doi 10 1109/TDSC 2007,70207

[4] Y. Aksari and H. Artuner, "Active authentication by mouse movements," in *Proc. 24th Int. Symp. Computer and Information Science*, Guzelyurt, 2009, pp. 571–574.

[5] S. Bengio and J. Mariethoz, "A statistical significance test for person authentication," in *Proc. Speaker and Language Recognition Workshop*, Toledo, Spain, 2004, pp. 237–244.

[6] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. Advance in Knowledge Discovery in Database: Papers From the 1994 AAAI Workshop*, Jul. 1994, pp. 359–37.

[7] R. Biddle, S. Chiasson, and P. C. van Oorschot, "Graphical passwords: Learning from the first twelve years," *ACM Computing Surveys*, vol. 44, no. 4, 2012, to be published.

[8] P. Bours and C. J. Fullu, "A login system using mouse dynamics," in *Proc. 5th Int. Conf. Intelligent Information Hiding and Multimedia Signal Processing*, Kyoto, Japan, 2009, pp. 1072–1077.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 258 - 261, 2015, ISSN:- 2319–8656

ENERGY EFFICIENCY IN FILE TRANSFER ACROSS WIRELESS COMMUNICATION

Snigdhamayee Biswal
Bharath University
Chennai, India

A.Agal Veena
Bharath University
Chennai, India

Shilpi Kumar
Bharath University
Chennai, India

G.Michael
Bharath University
Chennai, India

Abstract*:* The key idea of our Energy Efficiency management is to use the exchange between energy consumption vs the gain in responsibility, timeliness, and security to maximize the system helpful time period. we tend to formulate the exchange as Associate in Nursing optimization downside for dynamically crucial the most effective redundancy level to use to multipath routing for intrusion tolerance so the question response success likelihood is maximized whereas prolonging the helpful time period. Moreover, we think about this optimization downside for the case during which a voting-based distributed intrusion detection formula is applied to sight and evict malicious nodes during a HWSN. we over see to develop a novel likelihood model to investigate the most effective redundancy level in terms of path redundancy and supply redundancy, further because the best intrusion detection settings in terms of the amount of voters and the intrusion invocation interval below that the time period of a HWSN is maximized. we over see to then apply the analysis results obtained to the planning of a dynamic redundancy management formula to identify and apply the most effective style parameter settings at runtime in response to environmental changes, to maximize the HWSN lifetime.

Keywords: Intrusion detection system (IDS), multipath routing, fault tolerance, trust management, heterogeneous WSN (HWSN).

1.INTRODUCTION:

Advances in wireless communication and miniature electronics have enabled the Development of small, low-cost, low-power sensor nodes (SNs) with sensing and Communication capabilities [1],[2]. Therefore, the issues of Wireless Sensor Networks (WSNs) have become popular research subjects. WSN is infrastructure based network, and through the mass deployment of SNs, a WSN is formed. The major function of WSN is to collect and monitor the related information which about the specific environment. The SNs detect the surrounding environment or the given target and deliver the data to the sink using Wireless communication. The data is then analyzed to find out the state of the target. However, due to the design of their hardware, WSNs [3]-[7] suffer from many resources Constraints, such as low computation capability, limited memory and limited energy. Because WSNs are composed by numerous low-cost and small devices which are usually deploy to an open and unprotected area, they are vulnerable to various types of attacks. A prevention mechanism is used to counteract well-known attacks. However, interference mechanisms cannot resist overall attacks. Therefore, the attacks area unit needed to be detected. An Intrusion Detection System (IDS) is used frequently to detect the packets in a network, and confirm whether or square measure attackers. Additionally, IDS will facilitate to develop the hindrance system through acquired natures of attack. Many wireless sensor networks (WSNs) are deployed in an unattended environment in which energy replenishment is difficult. Due to limited resources, a WSN must not only satisfy the application specific QoS requirements such as reliability, minimum delay and security, but also minimize energy consumption to prolong the system useful lifetime. Recently, prior research efforts have been made to develop network architectures and sensor hardware in order to effectively deploy WSNs for a variety of applications. However, Due to a wide diversity of WSN application requirements, a general-purpose WSN design cannot fulfill the needs of all applications. In order to attain this, it is essential to capture the impacts of network parameters on network performance with respect to application specifications. Intrusion detection (i.e., object tracking) in a WSN can be regarded as a monitoring system for detecting the intruder that is invading the network domain. Thus, it is necessary to develop the intrusion detection system (IDS) which is capable of handling more extensive malicious attacks with energy conservation mechanism to increase system lifetime. If any system in the network gets affected by malicious behaviour, the request is given to the guardian system. The patch framework is given to the affected system by the guardian system and with the help the patch framework, the malicious in the affected system is cleared[8]-[10].

2. SYSTEM ANALYSIS

2.1.Existing System:

In Existing System, effective redundancy management of a clustered HWSN to prolong its lifetime operation in the presence of unreliable and malicious nodes. We tend to address the exchange between energy consumption vs. QoS gain in responsibility, timeliness and security with the goal

to maximize the lifetime of a clustered HWSN while satisfying application QoS requirements in the context of multipath routing. More specifically, we analyze the optimum quantity of redundancy through which data are routed to a remote sink in the presence of unreliable and malicious nodes, so that the question success probability is maximized where as maximizing the HWSN lifetime.

2.2. Proposed System:

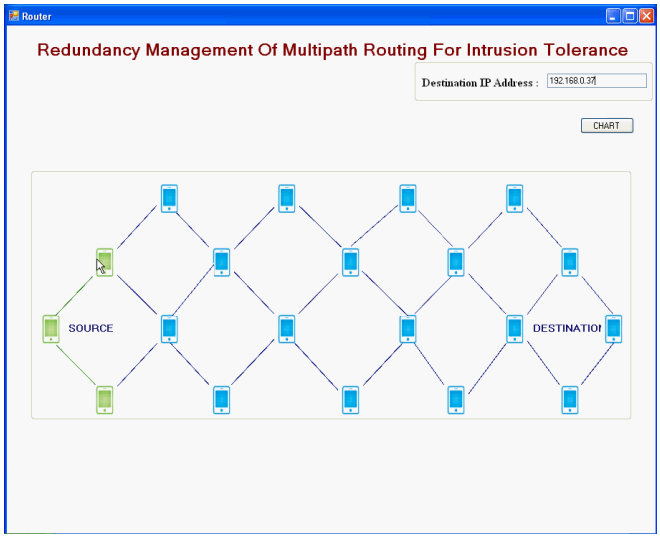
In planned System, the best communications vary and communication modes were derived to maximize the HWSN period of time. In intra-cluster planning and inter-cluster multi-hop routing schemes to maximize the network period of time. They thought of a hierarchic HWSN with CH nodes having larger energy and process capabilities than traditional SNs. Associates is developed as an improvement drawback to balance energy consumption across all nodes with their roles. In either work cited higher than, no thought was given to the existence of malicious nodes. A two-tier HWSN with the target of maximizing network period of time whereas fulfilling power management and coverage objectives. They determined the best density magnitude relation of the 2 tier's nodes to maximize the system period of time.

.

3. MODULES DESCRIPTION

3.1. Multipath routing:

One path reaching the sink node or base station will increase during this module, Multipath routing is taken into account an efficient mechanism for fault and intrusion tolerance to boost knowledge delivery in WSNs. the essential plan is that the likelihood of at least as we've a lot of methods doing knowledge delivery. Whereas most previous analysis targeted on exploitation multipath routing to boost responsibility, some attention has been paid to exploitation multipath routing to tolerate business executive attacks. These studies, however, mostly unnoticed the trade-off between QoS gain vs. energy consumption which may adversely shorten the system time period.



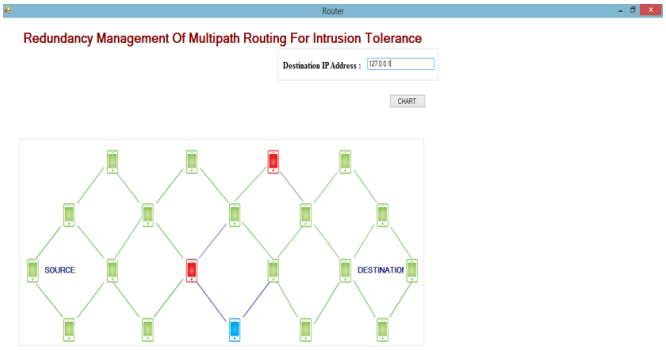
3.2.Intrusion Tolerance:

In these Modules, intrusion tolerance through multipath routing, there are two major issues to solve:

(1) How many paths to use and

(2) What paths to use.

Intrusion tolerance may be a fault-tolerant style approach to defensive data systems against malicious attack. Abandoning the traditional aim of preventing all intrusions, intrusion tolerance instead implies triggering mechanisms that forestall intrusions from resulting in a system security failure.



3.3.Energy Efficient:

In this module, there are two approaches by that energy economical IDS may be enforced in WSNs. One approach particularly applicable to flat WSNs is for AN intermediate node to feedback spite and energy standing of its neighbor nodes to the sender node (e.g., the supply or sink node) WHO will then utilize the data to route packets to avoid nodes with unacceptable spite or energy standing. Another

approach that we tend to adopt during this paper is to use native host-based IDS for energy conservation.

3.4.Simulation Process:

In this module, the value of execution the dynamic redundancy management rule delineated higher than, as well as periodic bunch, periodic intrusion detection, and question process through multipath routing, in terms of energy consumption.

4. IMPLEMENTATION IN ACTION

Fig. 1:Client a receive datas

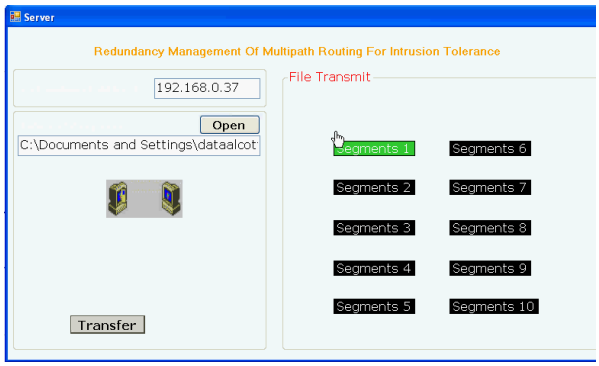


Fig. 2:Server transfers files

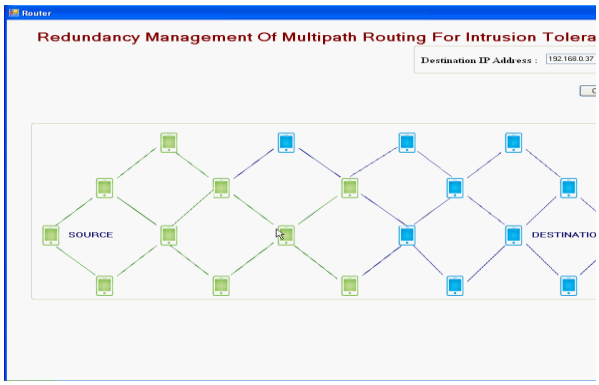


Fig. 3: Router a forward datas

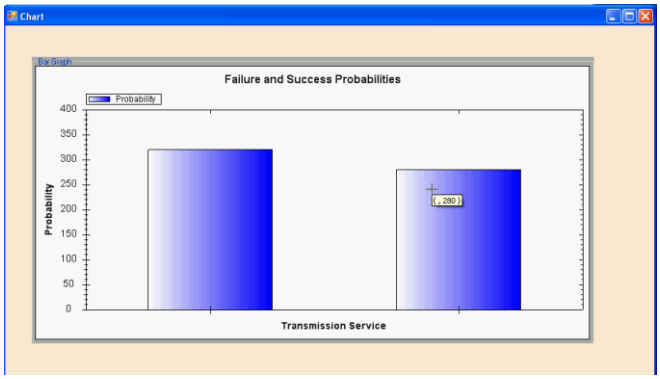


Fig.4: router transmission service

5. CONCLUSION

The proposed hierarchical dynamic trust management protocol for cluster-based wireless sensor networks, considering two aspects of truthfulness, namely, social truth and QoS truth. The research work will include the development of a probability model utilizing various techniques to analyze the protocol performance, and valid subjective trust against objective trust obtained based on ground truth node status. Based on the protocol the algorithm for truth-based intrusion detection will be developing using weighted choice. The algorithm will identify the best way to form trust out of social and QoS trust properties (i.e., identifying weights to assign to individual trust properties) and to assign the minimum trust threshold, in order that the performance of trust-based intrusion detection is maximized, i.e., each false positives and false negatives are minimized. Also, the research will deal with the challenging issue of providing fault tolerance in wireless device networks. Firstly a new multipath sample for heterogeneous wireless sensor networks will be define and analyzes upon various parameters. Then, propose a new fault tolerant multipath routing protocol which discovers an important number of energy node disjoint paths with the slightest overhead of one message per node. Intensive simulations will be conducted to evaluate our protocol with different scenarios, sensor nodes densities and deployment strategies.

6. REFERENCES

[1] O. Younis and S. Fahmy, “HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks,” *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 366–379, 2004.

[2] E. Felemban, L. Chang-Gun, and E. Ekici, “MMSPEED: multipath multi-SPEED protocol for QoS guarantee of reliability and timeliness in wireless sensor networks,” *IEEE Trans. Mobile Comput.*, vol. 5, no. 6, pp. 738–754, 2006.

[3] S. Bo, L. Osborne, X. Yang, and S. Guizani, “Intrusion detection techniques in mobile ad hoc and wireless sensor networks,” *IEEE Wireless Commun. Mag.*, vol. 14, no. 5, pp. 560–563, 2007.

[4] I. Krontiris, T. Dimitriou, and F. C. Freiling, "Towards intrusion detection in wireless sensor networks," in *Proc. 2007 European Wireless Conf.*

[5] J. H. Cho, I. R. Chen, and P. G. Feng, "Effect of intrusion detection on reliability of mission-oriented mobile group systems in mobile ad hoc networks," *IEEE Trans. Reliab.*, vol. 59, no. 1, pp. 231–241, 2010.

[6] A. P. R. da Silva, M. H. T. Martins, B. P. S. Rocha, A. A. F. Loureiro, L. B. Ruiz, and H. C. Wong, "Decentralized intrusion detection in wireless sensor networks," in *Proc. 2005 ACM Workshop Quality Service Security Wireless Mobile Netw.*

[7] Y. Zhou, Y. Fang, and Y. Zhang, "Securing wireless sensor networks: a survey," *IEEE Commun. Surveys & Tutorials*, vol. 10, no. 3, pp. 6–28, 2008.

[8] G. Micheal and A.R. Arunachalam "EAACK: Enhanced Adaptive Acknowledgment for MANET" Middle-East Journal of Scientific Research 19 (9), 2014.

[9] S.Parameswari , G.Michael" Intrusion Detection System in MANET : A Survey " IJETR, Vol-2, April 2014.

[10] G. Micheal "Detection of Malicious Behaviour in P2P Networks" IJETR, December 2014.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 262 - 273, 2015, ISSN:- 2319–8656

CONSEQUENCES OF ROAD TRAFFIC ACCIDENT IN NIGERIA: TIME SERIES APPROACH

F.B. Adebola
Department of Statistics
Federal University of Technology Akure.
Nigeria.

Ridwan A Sanusi
Department of Mathematics and Statistics
King Fahd University of Petroleum and Minerals,
Saudi Arabia.

N.A. Adegoke
Department of Statistics
Federal University of Technology Akure.
Nigeria.

Abstract: Road traffic accident in Nigeria is increasing at a worrying rate and has raised one of the country major concerns. We provided appropriate and suitable time series model for the consequences of road accident, the injured, killed and total casualty of the road accident in Nigeria. The most widely used conventional method, Autoregressive Integrated Moving Average (ARIMA) model of time series, also known as Box-Jenkins method is applied to yearly data on the consequences of road accident data in Nigeria from 1960-2013 to determine patterns of road traffic accident consequences; injured, killed and total casualty of the road accident along the Nigeria motorway. Appropriate models are developed for the accident consequences; injured, killed and total casualty. ARIMA (0; 2; 1) model is obtained for the injury and total casualty consequences, whilst ARIMA(1,2,2) model is obtained for the killed consequences, using the data from 1960-2011. The adequacy and the performance of the model are tested on the remaining data from 2012 to 2013. Seven years forecast are provided using the developed models and showed that road traffic accident consequences examined; injured, killed and total casualty would continue to increase on average.

Keywords: ARIMA; forecast; injured; killed; casualty

1. INTRODUCTION

Road Traffic Accident occurs when there is collision of vehicle with another vehicle, pedestrian and animals among other, which at times result in injury, loss of property and death. As mentioned in [11], road traffic accident leads to approximately two million killed and approximately ten million injuries annually. Also, an estimated value of 3000 people die in the world as a result of road traffic accidents daily. A prediction of global leading causes of killed from 2008 to 2030 by World Health Organization revealed that, if current trends and patterns continue, road traffic accidents will increase from ninth to fifth of world leading cause of killed 3.6% of global killed, up from 2.2% in 2004 [11]. While, disability-adjusted life years will rise from ninth with 2.7% of total disability-adjusted life in 2004 to third and 4.9% of total disability-adjusted life in 2030 [10].

Nigeria, the most populous black country, has the highest rate of mortality from road accidents in the world according to statistics compiled by the Federal Road Safety Commission **(FRSC)**. The country leads 43 other nations with killed in 10,000 vehicle crashes. Ethiopia ranked second with 219 killed per 10,000 vehicles while Malawi, took the third position and Ghana took the fourth position with 183 and 178 killed respectively [1].

Road traffic accidents is one of the leading causes of death among older children and economically active adults between the ages 30 and 49 years ([8];[9]; [6]). Considering the importance of the road and the increased level of road traffic accidents in recent years along the Nigeria roads, this study aimed at characterizing the road traffic accident in Nigeria by providing appropriate models that explain the consequences of killed, injured and the total casualty from road accident in the country so as to provide an enabling base for the development of countermeasures by the government and the traffic control agents to reduce incidences of road traffic accident on the road.

Time series analysis encompases methods for analyzing data ordered in time in order to develop appropriate model and other characteristics of the time ordered data. It is commonly used in the fields of business, economics, finance, agriculture among others, as appropriate tool for model building. It systematics examine the ordered data with the aim of studying dynamic regularities that may enable forecasting future or even controlling the variable, the forecast model will then be used to predict future values based on previously observed values. In theory, Auto-regressive Integrated Moving Averages ARIMA Models are the most universal class of models for forecasting a time series data. As proposed by Box and Jenkins, that in general, forecasting based on ARIMA models comprises of three different steps: Model Identification, Parameter estimation and Diagnostic checking. Until a desirable model for the data is identified, the three steps will be repeated [3]. The method of Box and Jenkins dictates an iterative process requiring a sound understanding of time series analysis technique, some degree of judgement and many rounds of trials [13].

Numerous works have been done on the analysis of Road accidents. [5] examined road accidents in Kuwait, he used an ARIMA model and compared it with ANN to predict killed in Kuwait, he concluded that ANN was better in case of long term series without seasonal fluctuations of accidents or autocorrelations' components. [4] used Bayesian Model for ranking hazardous road sites, their model made use of all relevant information per accident location, including the total number of accidents and the number of killed, as well as the number of slight and serious injuries. Moreover, the model included the use of a cost function to rank the sites with respect to their total expected cost to society.

A procedure of Road Traffic Injury (RTI) in China by using RTI data from 1951 to 2003 was established by [12]. A series of predictive equations on RTI were established based on ARIMA models. They concluded that time series models thus established proves to be of significant usefulness in RTI prediction. Two time series techniques; ARMA and Holt-Winters (HW) algorithm to predict annual motor vehicle crash killed were used by [7]. They concluded that the values predicted by ARMA models are a little bit higher than the ones obtained by HW algorithm. Intervention analysis with univariate Box-Jenkins method to identify whether a change in a particular policy had made an impact on the trends in killed and fatality rates in Illinois was used [2]. He developed ARIMA forecasting model for future trends in motorway killed in an effort to provide assistance to policy development in reducing fatality rates in Illinois.

Time series analysis have been used in many fields of research and road safety is no exception. The results of this research would also add to the many research works carried out in road safety.

2. MATERIALS AND METHODS

Data used for the study is a secondary data, it was collected on yearly basis from the office of the Federal Road Safety Corps of Nigeria for the period 1960 to 2013. The data represents the total number of registered consequences of injuries, killed and total casualty for the period under study. The **Box and Jenkins** approach for time series analysis was employed for data analysis. According to **Box and Jenkins**, as mentioned above, the steps include, the identification of appropriate model for the data under study, estimation of model parameters, model diagnostic and adequacy checking and lastly, the model, if found appropriate would be used for forecasting. Data from 1960 to 2011 are used for models building, while, data from 2012 to 2013 are used for models validation and forecast values of the best models for the variables under study are obtained from 2014 to 2020. Meanwhile, It is worth mentioning here that because of the volume of the work, the best models out of several competing models that explain the variables under study are only included in the work.

3. MODEL BUILDING

The first step in model building is to obtain the time plot of the data. This will give us an insight of the behaviour of the series. Figures (1a, 1b, and 1c) show the time series plot of injuries consequences, killed consequences and total casualty from the total number of road accident in Nigeria.

The plots exhibit upward and downward movement for all the three variables under study, with some significant upward and downward trends at some parts of the series. The mean and variance of the variables are not stable and varies with time.

The autocorrelation function of the studied variables has shown in Figure (2a), Figure (2b) and Figure (3) describe the correlation between values of the studied variables at different points in time, as a function of the time difference. The first several autocorrelations are persistently large and trailed off to zero rather slowly for all the three variables and their spikes also went of the autocorrelations limit at lag 13 the variables under study.

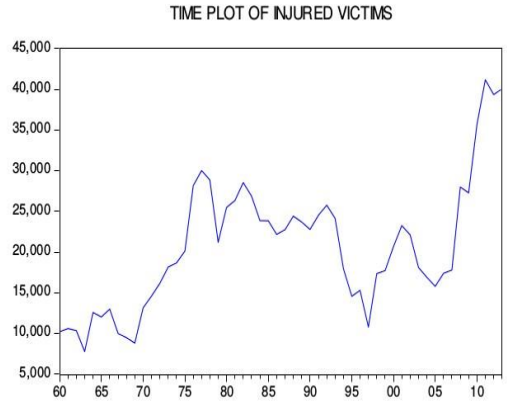


Figure 1a: Time Series Plot of Injured Victims from Road Accidents in Nigeria.

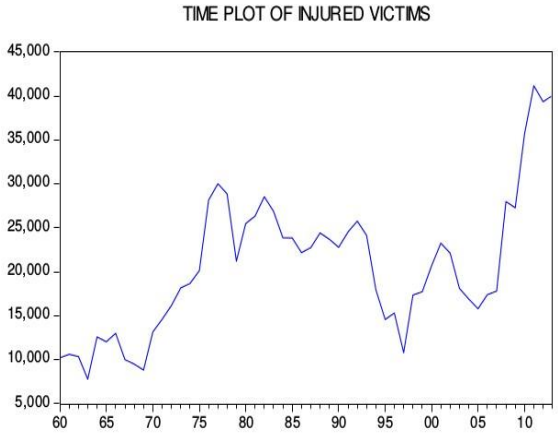
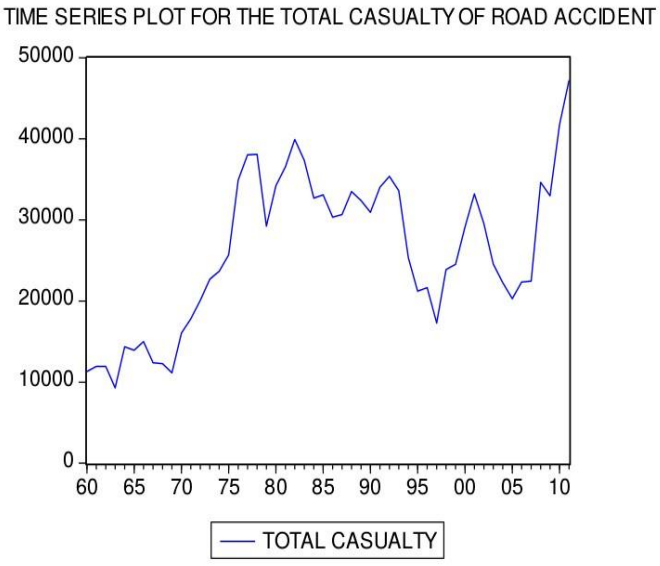
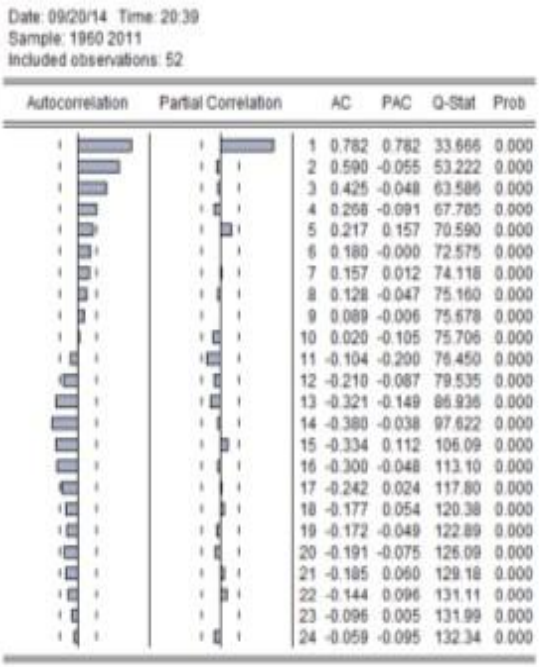


Figure 1b: Time Series Plots of killed from Road Accidents in Nigeria.



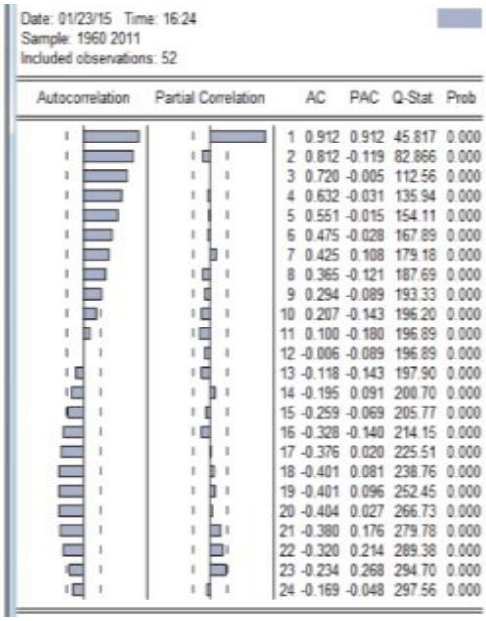
Figures 1c: Time Series Plot of Total Casualty from Road Accidents in Nigeria.

The Augmented Dickey Fuller test as given in Figures (4a and 4b) and Figure (5) give a p-value of 0.91 for the injured


Date: 09/20/14 Time: 20:39
Sample: 1960 2011
Included observations: 52

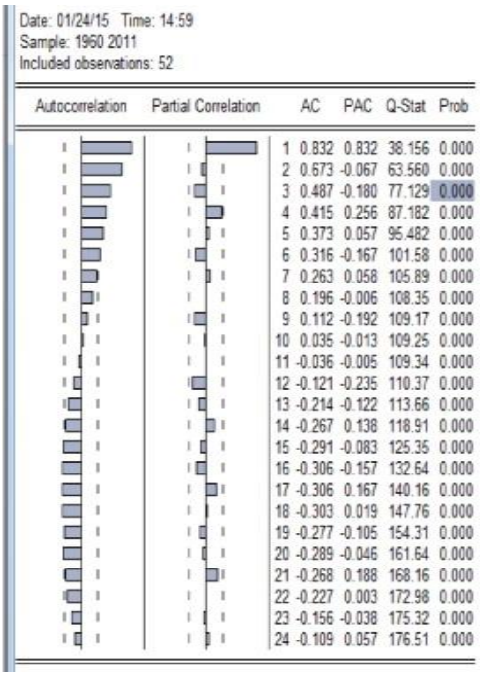
	Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
			1	0.782	0.782	33.666	0.000
			2	0.590	-0.055	53.222	0.000
			3	0.425	-0.048	63.586	0.000
			4	0.268	-0.091	67.785	0.000
			5	0.217	0.157	70.590	0.000
			6	0.180	-0.000	72.575	0.000
			7	0.157	0.012	74.118	0.000
			8	0.128	-0.047	75.160	0.000
			9	0.089	-0.006	75.678	0.000
			10	0.020	-0.105	75.706	0.000
			11	-0.104	-0.200	76.450	0.000
			12	-0.210	-0.087	79.535	0.000
			13	-0.321	-0.149	86.936	0.000
			14	-0.380	-0.038	97.622	0.000
			15	-0.334	0.112	106.09	0.000
			16	-0.300	-0.048	113.10	0.000
			17	-0.242	0.024	117.80	0.000
			18	-0.177	0.054	120.38	0.000
			19	-0.172	-0.049	122.89	0.000
			20	-0.191	-0.075	126.09	0.000
			21	-0.185	0.060	129.18	0.000
			22	-0.144	0.096	131.11	0.000
			23	-0.096	0.005	131.99	0.000
			24	-0.059	-0.095	132.34	0.000

Figure 2a: Correlogram Plot of the Injured Victims Nigeria


Date: 01/23/15 Time: 16:24
Sample: 1960 2011
Included observations: 52

	Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
			1	0.912	0.912	45.817	0.000
			2	0.812	-0.119	82.866	0.000
			3	0.720	-0.005	112.56	0.000
			4	0.632	-0.031	135.94	0.000
			5	0.551	-0.015	154.11	0.000
			6	0.475	-0.028	167.89	0.000
			7	0.425	0.108	179.18	0.000
			8	0.365	-0.121	187.69	0.000
			9	0.294	-0.089	193.33	0.000
			10	0.207	-0.143	196.20	0.000
			11	0.100	-0.180	196.89	0.000
			12	-0.006	-0.089	196.89	0.000
			13	-0.118	-0.143	197.90	0.000
			14	-0.195	0.091	200.70	0.000
			15	-0.259	-0.069	205.77	0.000
			16	-0.328	-0.140	214.15	0.000
			17	-0.376	0.020	225.51	0.000
			18	-0.401	0.081	238.76	0.000
			19	-0.401	0.096	252.45	0.000
			20	-0.404	0.027	266.73	0.000
			21	-0.380	0.176	279.78	0.000
			22	-0.320	0.214	289.38	0.000
			23	-0.234	0.268	294.70	0.000
			24	-0.169	-0.048	297.56	0.000

Figure 2b: Correlogram Plot of the killed from Road Accidents in Nigeria


Date: 01/24/15 Time: 14:59
Sample: 1960 2011
Included observations: 52

	Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
			1	0.832	0.832	38.156	0.000
			2	0.673	-0.067	63.560	0.000
			3	0.487	-0.180	77.129	0.000
			4	0.415	0.256	87.182	0.000
			5	0.373	0.057	95.482	0.000
			6	0.316	-0.167	101.58	0.000
			7	0.263	0.058	105.89	0.000
			8	0.196	-0.006	108.35	0.000
			9	0.112	-0.192	109.17	0.000
			10	0.035	-0.013	109.25	0.000
			11	-0.036	-0.005	109.34	0.000
			12	-0.121	-0.235	110.37	0.000
			13	-0.214	-0.122	113.66	0.000
			14	-0.267	0.138	118.91	0.000
			15	-0.291	-0.083	125.35	0.000
			16	-0.306	-0.157	132.64	0.000
			17	-0.306	0.167	140.16	0.000
			18	-0.303	0.019	147.76	0.000
			19	-0.277	-0.105	154.31	0.000
			20	-0.289	-0.046	161.64	0.000
			21	-0.268	0.188	168.16	0.000
			22	-0.227	0.003	172.98	0.000
			23	-0.156	-0.038	175.32	0.000
			24	-0.109	0.057	176.51	0.000

Figure 3: Correlogram Plot of the Total Casualty from Road Accidents in Nigeria.

Null Hypothesis: IJ has a unit root
Exogenous: None
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		0.970792	0.9100
Test critical values:	1% level	-2.611094	
	5% level	-1.947381	
	10% level	-1.612725	

*MacKinnon (1996) one-sided p-values.

Figure 4a: Unit Root Test of Injured consequences from Road Accidents in Nigeria.

Null Hypothesis: KL has a unit root
Exogenous: None
Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.109354	0.6412
Test critical values:	1% level	-2.611094	
	5% level	-1.947381	
	10% level	-1.612725	

*MacKinnon (1996) one-sided p-values.

Figure 4b: Unit Root Test of killed from Road Accidents in Nigeria.

victims, 0.6412 for the killed consequences and 0.8779 for the total casualty, these indicate the presence of unit roots for the series. All these aforementioned characteristics of the studied variables show that the series are not stationary, thus require differencing.

Null Hypothesis: TOTC has a unit root
Exogenous: None
Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		0.775684	0.8779
Test critical values:	1% level	-2.611094	
	5% level	-1.947381	
	10% level	-1.612725	

*MacKinnon (1996) one-sided p-values.

Figure (5): Unit Root Test of total casualty from Road Accidents in Nigeria.

Figures (6a, 6b, and 6c), show the second difference of the studied variables, the series look more stable around the mean, which shows that the variables are now stationary. All the three variables become stationary after taken second non-seasonal difference

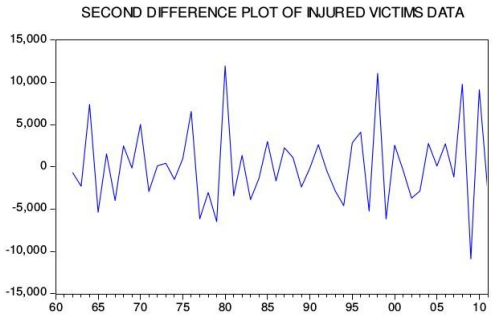


Figure 6a: Time Series Plot of the Second Difference for the Injured Victims consequences of Road Accidents.

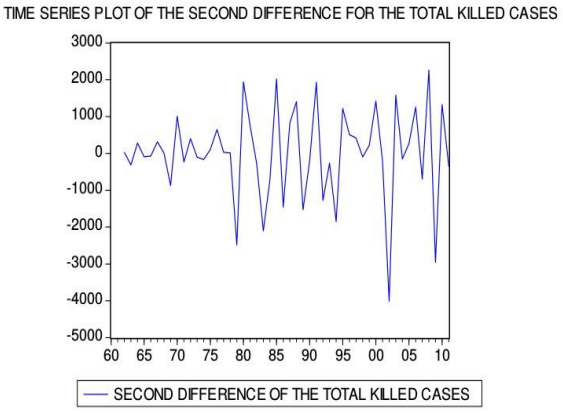


Figure 6b: Time Series Plot of the Second Difference for the killed consequences of Road Accidents.

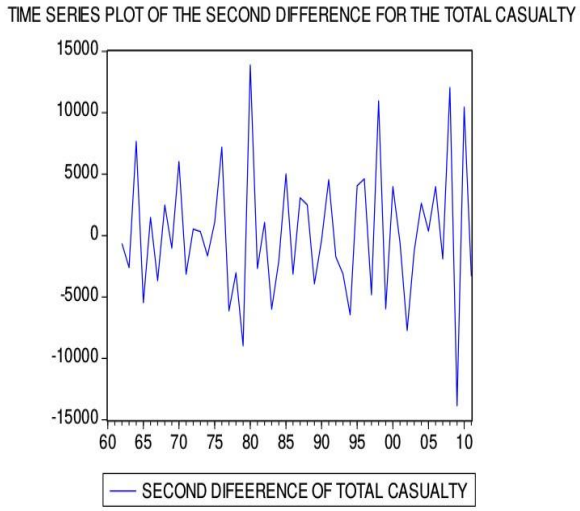


Figure 6c: Time Series Plot of the Second Difference for the Injured Victims, killed and Total Casualty consequences of Road Accidents.

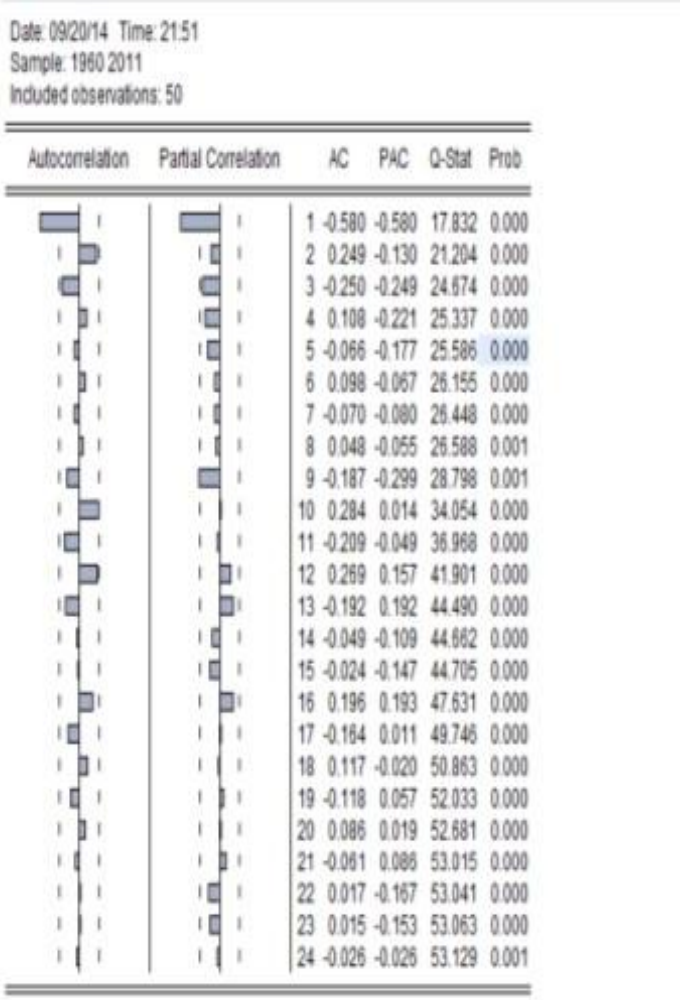


Figure 7a: Correlogram Plot of the Second Difference for the Injured Victims consequences of Road Accidents.

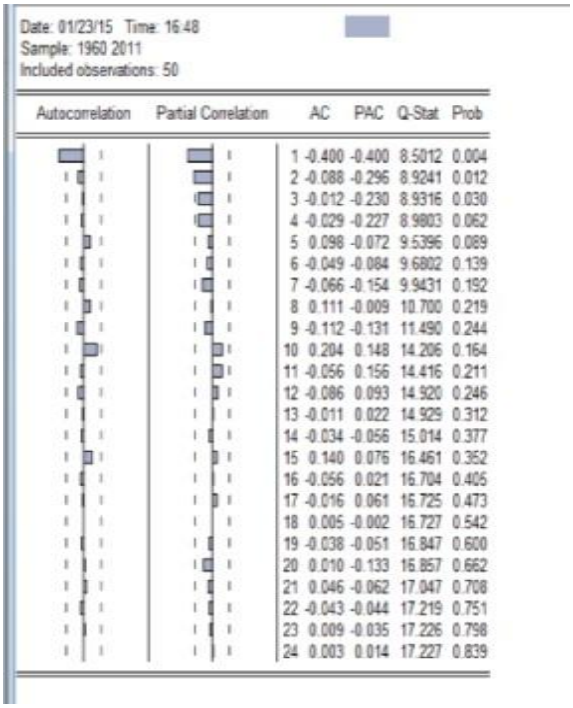


Figure 7b: Correlogram Plots of the Second Difference for the killed consequences of Road Accidents.

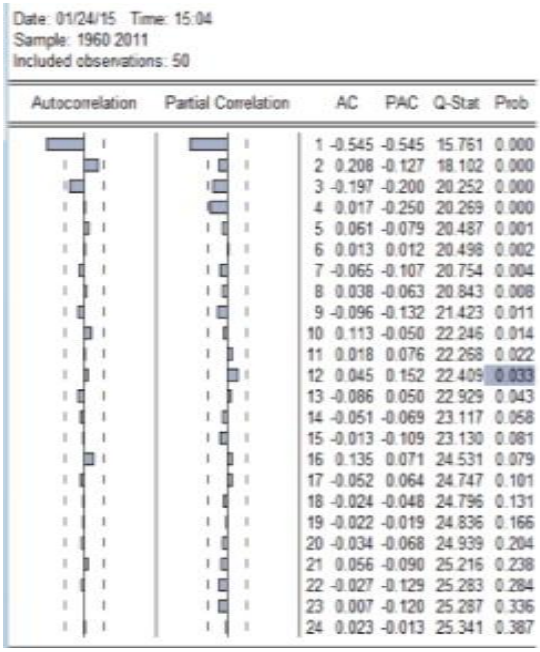


Figure 7c: Correlogram Plots of the Second Difference for the Total Casualty consequences of Road Accidents.

The autocorrelation functions of the second difference for the studied variables, has shown in Figures (7a, 7b, and 7c), also confirm that the second difference are now stationary. Also, the Augmented Dickey Fuller test as given in Figures (8a, 8b, and 8c) gave a p-value of 0.000 for the Injured victims, 0.0004 for the killed consequences and 0.000 for the total casualty, these also indicate the absence of unit roots in the series, which confirm that the second differenced series are stationary.

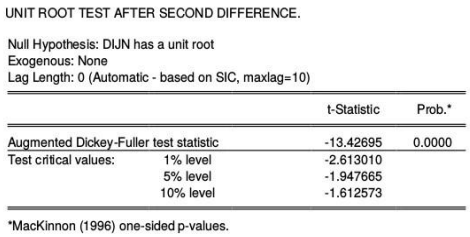


Figure 8a: Unit Root Test for the Second Difference for the Injured Victims consequences of Road Accidents.

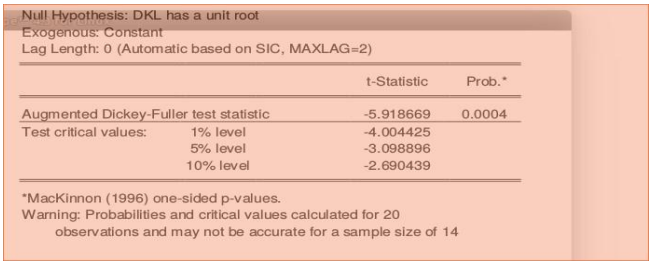


Figure 8b: Unit Root Tests for the Second Difference for the killed consequences of Road Accidents.

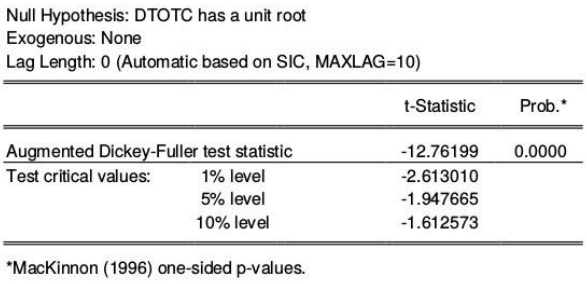


Figure 8c: Unit Root Tests for the Second Difference for the Total Casualty consequences of Road Accidents.

By comparing the autocorrelations functions with their error limits, the only significant autocorrelations are at lag 1 for all the three variables, that is, the autocorrelations cut off after lag one which shows the existence of MA(1) behavior. Similarly, the partial autocorrelations also cut off after lag one for the injured consequences and total casualty, this indicates the existence of AR(1) for the two variables (that is, injured consequences and total casualty). Meanwhile, the partial autocorrelation cuts off after lag two for the killed consequences, which shows the existence of AR(1) and AR(2) for the variable. Based on the features of the correlogram plots of the stationary series, the following model in Figure (1), are suggested.

Injured Victims	killed	Total Casualty
ARIMA(0,2,1)	ARIMA(0,2,2)	ARIMA(0,2,1)
ARIMA(1,2,0)	ARIMA(1,2,2)	ARIMA(1,2,0)
ARIMA(1,2,1)	ARIMA(1,2,3)	ARIMA(1,2,1)

Table 1: Suggested Models Based on the Correlogram Plots

Each of the model is assessed based on its parameter estimates, the corresponding diagnostics of the residuals, the AIC and SIC in order to select the best model for forecasting into the future. Meanwhile, out of all the competing models that explain the variable of interest, the best models are; ARIMA(0,2,1) for the Injured Victims consequences, ARIMA(1,2,2) for killed consequences and ARIMA(0,2,1) for the total casualty. The models are given in Figures (9a, 9b, and 10).

Time Series Models for the Injured Victims, killed and Total Casualty consequences of Road Accidents are given in Figures (9a, 9b, and 10), the models coefficients are significant and all the inverted AR roots satisfy the minimum stationarity condition, the invertibility condition of MA is satisfied and also. Also, the Durbin-Watson statistics is not far from 2, which implies that there is no serial correlation in the model residual, that is the model residual is not forecastable.

MODEL OUTPUT OF INJURED CASES, ARIMA(0,2,1)

Dependent Variable: D(IJ,2)
Method: Least Squares
Date: 09/20/14 Time: 16:27
Sample (adjusted): 1962 2011
Included observations: 50 after adjustments
Convergence achieved after 8 iterations
MA Backcast: 1961

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.954388	0.031563	-30.23748	0.0000

R-squared	0.438091	Mean dependent var	101.5200
Adjusted R-squared	0.438091	S.D. dependent var	4692.813
S.E. of regression	3517.759	Akaike info criterion	19.18883
Sum squared resid	6.06E+08	Schwarz criterion	19.22707
Log likelihood	-478.7208	Hannan-Quinn criter.	19.20340
Durbin-Watson stat	1.861128		

Inverted MA Roots	.95

Figure 9a: Time Series Models for the Injured Victims consequences of Road Accidents.

Dependent Variable: D(KL,2)
Method: Least Squares
Date: 01/23/15 Time: 16:50
Sample (adjusted): 1963 2011
Included observations: 49 after adjustments
Convergence achieved after 20 iterations
Backcast: 1961 1962

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.832466	0.089079	-9.345282	0.0000
MA(2)	-0.968707	0.040191	-24.10259	0.0000

R-squared	0.459802	Mean dependent var	-5.367347
Adjusted R-squared	0.448308	S.D. dependent var	1281.902
S.E. of regression	952.1448	Akaike info criterion	16.59527
Sum squared resid	42609245	Schwarz criterion	16.67249
Log likelihood	-404.5841	Durbin-Watson stat	1.925381

Inverted AR Roots	-.83	
Inverted MA Roots	.98	-.98

Figure 9b: Time Series Models for the killed consequences of Road Accidents.

Date: 01/24/15 Time: 14:57
Sample (adjusted): 1962 2011
Included observations: 50 after adjustments
Convergence achieved after 9 iterations
Backcast: 1961

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.959777	0.028209	-34.02330	0.0000

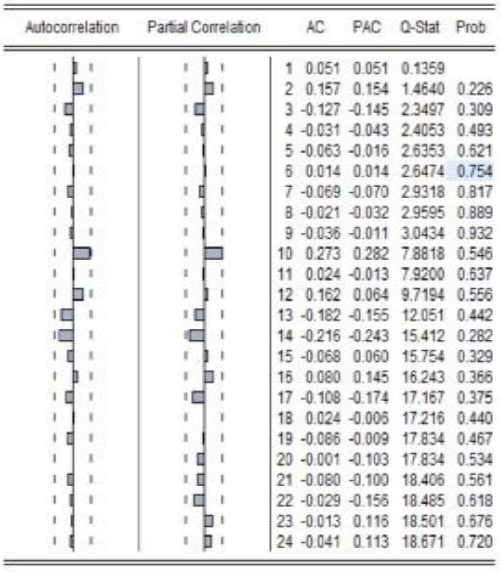
R-squared	0.427033	Mean dependent var	96.96000
Adjusted R-squared	0.427033	S.D. dependent var	5545.392
S.E. of regression	4197.560	Akaike info criterion	19.54219
Sum squared resid	8.63E+08	Schwarz criterion	19.58043
Log likelihood	-487.5548	Durbin-Watson stat	1.815913

Inverted MA Roots	.96

Figure 10: Time Series Models for the consequences of Road Accidents

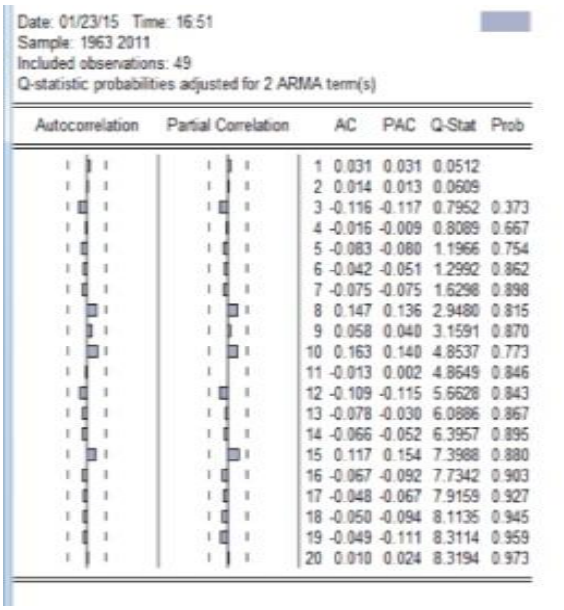
Also, all the Q-Stat of the correlogram plot of models residuals are greater than 0.05 for the lags as given in Figures (11a and 11b) and Figure (12), these imply that the model residuals are White-Noise, that is adjacent observations are not related (random) and which support the fact that the models may be the appropriate models for the observed time series.

Date: 09/20/14 Time: 17:56
Sample: 1960 2011
Included observations: 50
Q-statistic probabilities adjusted for 1 ARMA term



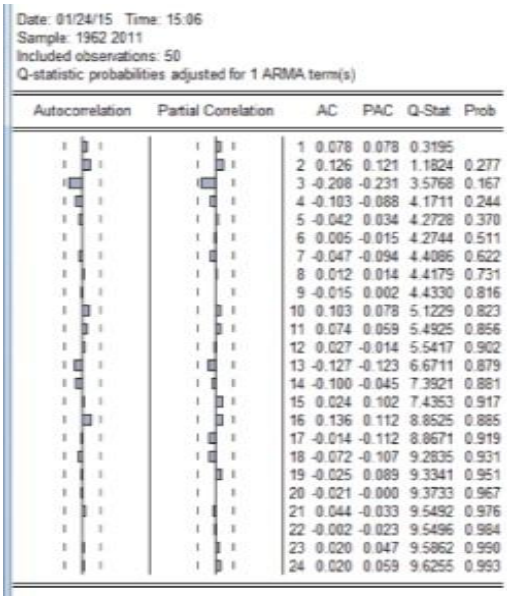
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.051	0.051	0.1359	
		2	0.157	0.154	1.4640	0.226
		3	-0.127	-0.145	2.3497	0.309
		4	-0.031	-0.043	2.4053	0.493
		5	-0.063	-0.016	2.6353	0.621
		6	0.014	0.014	2.6474	0.754
		7	-0.069	-0.070	2.9318	0.817
		8	-0.021	-0.032	2.9595	0.889
		9	-0.036	-0.011	3.0434	0.932
		10	0.273	0.282	7.8818	0.546
		11	0.024	-0.013	7.9200	0.637
		12	0.162	0.064	9.7194	0.556
		13	-0.182	-0.155	12.051	0.442
		14	-0.216	-0.243	15.412	0.282
		15	-0.068	0.060	15.754	0.329
		16	0.080	0.145	16.243	0.366
		17	-0.108	-0.174	17.167	0.375
		18	0.024	-0.006	17.216	0.440
		19	-0.086	-0.009	17.834	0.467
		20	-0.001	-0.103	17.834	0.534
		21	-0.080	-0.100	18.406	0.561
		22	-0.029	-0.156	18.485	0.618
		23	-0.013	0.116	18.501	0.676
		24	-0.041	0.113	18.671	0.720

Figure 11a: Correlogram Plot of the Residuals for the Injured Victims killed of Road Accidents.


Date: 01/23/15 Time: 16:51
Sample: 1963 2011
Included observations: 49
Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.031	0.031	0.0512	
		2	0.014	0.013	0.0609	
		3	-0.116	-0.117	0.7952	0.373
		4	-0.016	-0.009	0.8089	0.667
		5	-0.083	-0.080	1.1966	0.754
		6	-0.042	-0.051	1.2992	0.862
		7	-0.075	-0.075	1.6296	0.898
		8	0.147	0.136	2.9480	0.815
		9	0.058	0.040	3.1591	0.870
		10	0.163	0.140	4.8537	0.773
		11	-0.013	0.002	4.8649	0.846
		12	-0.109	-0.115	5.6628	0.843
		13	-0.078	-0.030	6.0886	0.867
		14	-0.066	-0.052	6.3957	0.895
		15	0.117	0.154	7.3988	0.880
		16	-0.067	-0.092	7.7342	0.903
		17	-0.048	-0.067	7.9159	0.927
		18	-0.050	-0.094	8.1135	0.945
		19	-0.049	-0.111	8.3114	0.959
		20	0.010	0.024	8.3194	0.973

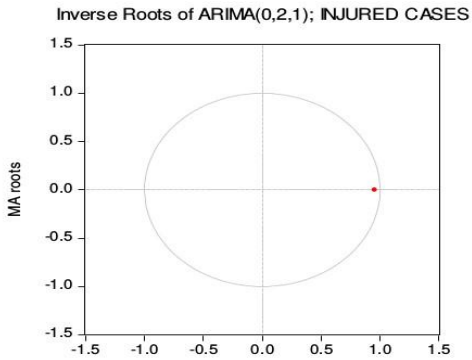
Figure 11b: Correlogram Plot of the Residuals for the killed of Road Accidents.


Date: 01/24/15 Time: 15:06
Sample: 1962 2011
Included observations: 50
Q-statistic probabilities adjusted for 1 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.078	0.078	0.3195	
		2	0.126	0.121	1.1824	0.277
		3	-0.208	-0.231	3.5768	0.167
		4	-0.103	-0.088	4.1711	0.244
		5	-0.042	0.034	4.2728	0.370
		6	0.005	-0.015	4.2744	0.511
		7	-0.047	-0.094	4.4086	0.622
		8	0.012	0.014	4.4179	0.731
		9	-0.015	0.002	4.4330	0.816
		10	0.103	0.078	5.1229	0.823
		11	0.074	0.059	5.4925	0.856
		12	0.027	-0.014	5.5417	0.902
		13	-0.127	-0.123	6.6711	0.879
		14	-0.100	-0.045	7.3921	0.881
		15	0.024	0.102	7.4353	0.917
		16	0.136	0.112	8.8525	0.885
		17	-0.014	-0.112	8.8671	0.919
		18	-0.072	-0.107	9.2835	0.931
		19	-0.025	0.089	9.3341	0.951
		20	-0.021	-0.000	9.3733	0.967
		21	0.044	-0.033	9.5492	0.976
		22	-0.002	-0.023	9.5496	0.984
		23	0.020	0.047	9.5862	0.990
		24	0.020	0.059	9.6255	0.993

Figure 12: Correlogram Plot of the Total Casualty consequences of Road Accidents.

Figure 13a: Unit Root Test for the Injured Victims consequences of Road Accidents.



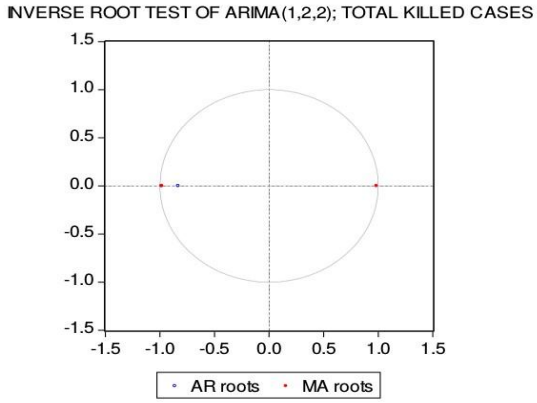


Figure 13b: Unit Root Test for the killed consequences of Road Accidents.

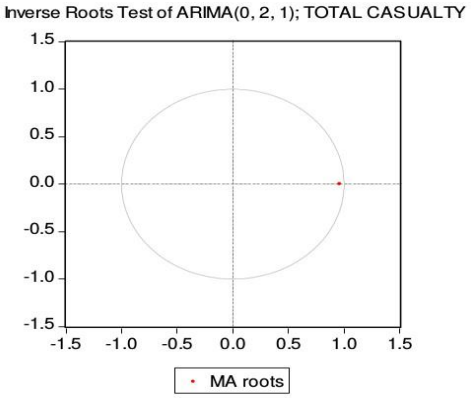


Figure 13c: Unit Root Test for the Total Casualty consequences of Road Accidents.

The unit roots tests of the models as given in Figures (13a, 13b, and 13c), show that the inverse roots of the models are within a unit circle, which confirmed that the models in Figures (9a, and 9b) and Figure (10) are stationary and invertible. Thus, the models can be written as general linear form

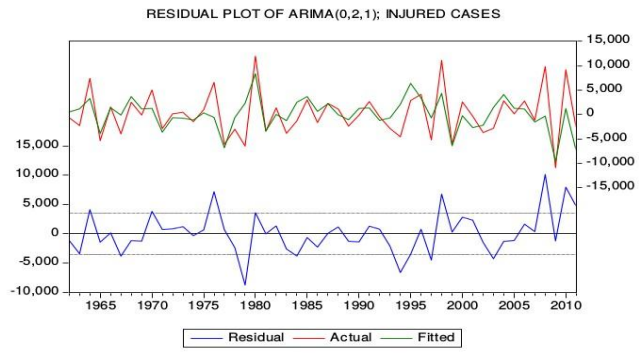


Figure 14a: Residual Plot for the Injured Victims consequences of Road Accidents.

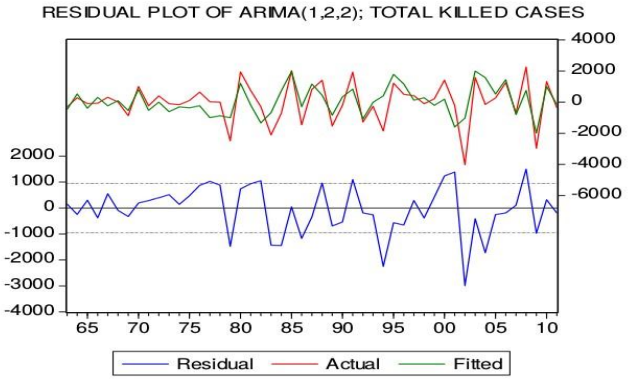


Figure 14b: Residual Plot for the killed consequences of Road Accidents.

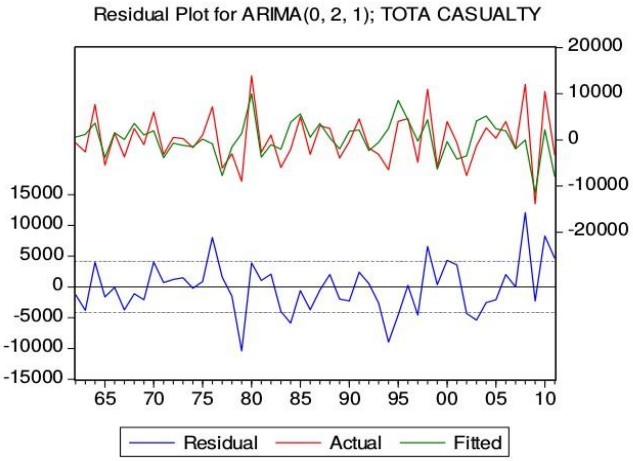
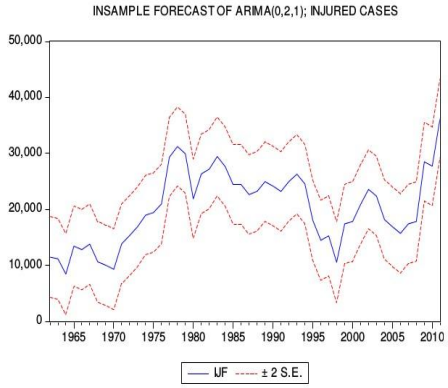


Figure 14c: Residual Plot for the Total Casualty consequences of Road Accidents.

The residual plots of the models as shown in Figures (14a, 14b, and 14c), also confirm that the models residuals are random and non-forecastable, which implies that the models are good.

Figures (15a, 15b, and 15c) gives the visual representation of the original Injured consequences, killed consequences and the Total casualty consequences, the data (blue line) and confidence interval (red

Figure 15a: In-sample Forecast Graph for the Injured cases.



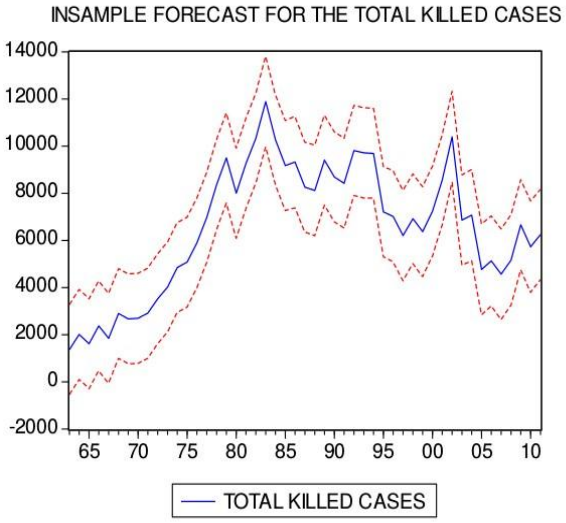


Figure 15b: In-sample Forecast Graph for the killed cases.

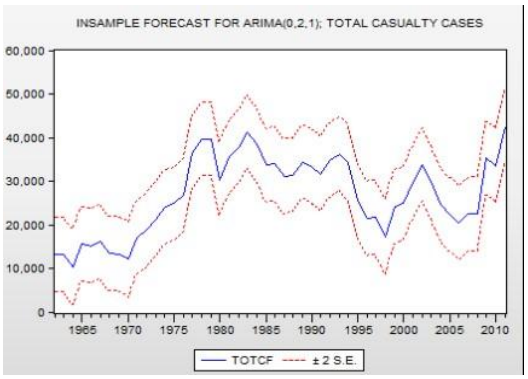


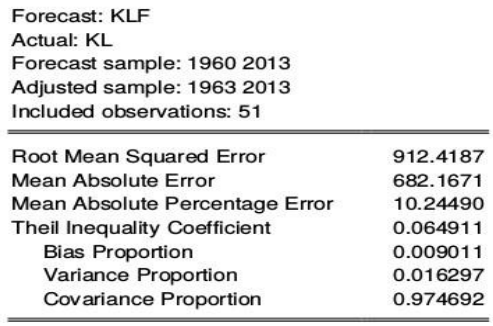
Figure 15c: In-sample Forecast Graph for the Total Casualties.

lines). The in-sample forecasts for the models fall within the 95% confidence Interval. Figures (16a, 16b, and 17) give the in-sample models evaluations, the bias proportion and variance proportion, which are used to check how far is the forecast mean from the mean of the actual series and how far is the forecast variance from the variance of the actual series respectively are very close to zero and comparatively much lower than the covariance proportion which measure the remaining systematic forecast error. Note, the sum of the bias proportion, variance proportion and the covariance proportion is 1.

Forecast: IJF
Actual: IJ
Forecast sample: 1960 2011
Adjusted sample: 1962 2011
Included observations: 50

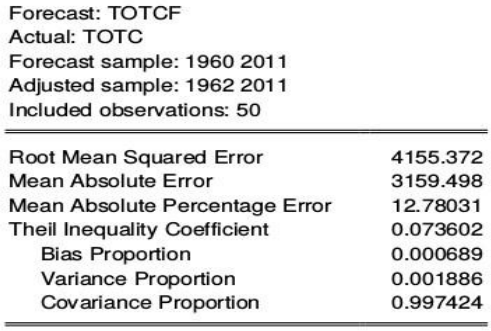
Root Mean Squared Error	3482.404
Mean Absolute Error	2545.251
Mean Absolute Percentage Error	13.77604
Theil Inequality Coefficient	0.081381
Bias Proportion	0.000286
Variance Proportion	0.015474
Covariance Proportion	0.984240

Figure 16a: In-sample Forecast Evaluation for the Injured cases.


Forecast: KLF
Actual: KL
Forecast sample: 1960 2013
Adjusted sample: 1963 2013
Included observations: 51

Root Mean Squared Error	912.4187
Mean Absolute Error	682.1671
Mean Absolute Percentage Error	10.24490
Theil Inequality Coefficient	0.064911
Bias Proportion	0.009011
Variance Proportion	0.016297
Covariance Proportion	0.974692

Figure 16b: In-sample Forecast Evaluation for the killed cases.


Forecast: TOTCF
Actual: TOTC
Forecast sample: 1960 2011
Adjusted sample: 1962 2011
Included observations: 50

Root Mean Squared Error	4155.372
Mean Absolute Error	3159.498
Mean Absolute Percentage Error	12.78031
Theil Inequality Coefficient	0.073602
Bias Proportion	0.000689
Variance Proportion	0.001886
Covariance Proportion	0.997424

Figure 17: In-sample Forecast Evaluation for the Total Casualties

3.1 MODEL VALIDATION

Table 2: Validation Table for ARIMA(0,2,1) Model of Injured consequences.

Year	Injured consequences	Forecast	% Variation
2012	39348	42213.26	7.28%
2013	40057	43261.51	7.99%

After determining the best-fit model for the series and estimating related parameters, the third phase of Box-Jenkins fitting model was evaluated for series prediction. Using the ARIMA (0,2,1) model, the model predicted that in 2012 an approximately 42213.26 Injure consequences, this gives 7.28% percentage increase when compared with the real value of 39348 Injured consequences. Also, the model predicted that in 2013 an approximately 43261.51 Injure consequences, this gives 7.99% percentage increament when compared with the real value of 40057 Injured consequences as given in Table (2).

Table 3: Validation Table for ARIMA(1,2,2) Model of Killed consequences.

Year	killed consequences	Forecast	% Variation
2012	6092	6046.28	-0.75%
2013	6544	6236.27	-4.702%

Also, Table (3) gives the model validation for ARIMA (1,2,2) model. The model predicted that in 2012 an approximately 6046.28 killed consequences of accident, this gives 0.75% percentage decrease when compared with the real value of 6092 killed consequences. Also, the model predicted that in 2013 an approximately 6236.27 killed consequences, this gives 4.702% percentage decrease when compared with the real value of 6544 killed consequences.

Table 4: Validation Table for ARIMA(0,2,1) Model of Total Casualty.

Lastly, Table (4) gives the model validation for ARIMA (0,2,1) model. The model predicted that in 2012 an approximately 46504.31 Total casualty consequences of accident, this gives 2.34% percentage increase when compared with the real value of 4544 total casualty consequences. Also, the model predicted that in 2013 an approximately 46838.61 total casualty consequences, this gives 0.51% percentage increase when compared with the real value of 46601 killed consequences.

Year	Total Casualy	Forecast	% Variation
2012	45440	46504.31	2.34%
2013	46601	46838.61	0.51%

3.2 Models Forecasting

Table 5: Forecast Table for ARIMA(0,2,1) Model of Injured consequences.

Year	Lower Control Limit	Forecast	Upper Control Limit
2014	31660.6	44309.8	56959.2
2015	30375.7	45358.1	60340.4
2016	29235.8	46406.3	63576.8
2017	28186.1	47454.5	66722.9
2018	27195.4	48502.8	69810.2
2019	26243.8	49551.1	72858.3
2020	25318.2	50599.3	75880.4

Table 6: Forecast Table for ARIMA(1,2,2) Model of killed consequences.

Year	Lower Control Limit	Forecast	Upper Control Limit
2014	2775.5	6261.7	9747.9
2015	2299.2	6424.1	10549.0
2016	1853.8	6472.4	11091.0
2017	1481.3	6615.8	11750.2
2018	1107.5	6680.0	12252.5
2019	785.5	6810.1	12834.6
2020	455.6	6885.4	13315.2

Table 7: Forecast Table for ARIMA(0,2,1) Model of Total Casualty consequences.

Year	Lower Control Limit	Forecast	Upper Control Limit
2014	32833.4	47415.9	61998.3
2015	30053.2	47145.8	64238.5
2016	27649.7	46679.6	65709.6
2017	26135.3	46897.7	67660.1
2018	24745.0	47274.3	69803.6
2019	22911.4	47098.1	71284.9
2020	21177.0	46794.0	72410.9

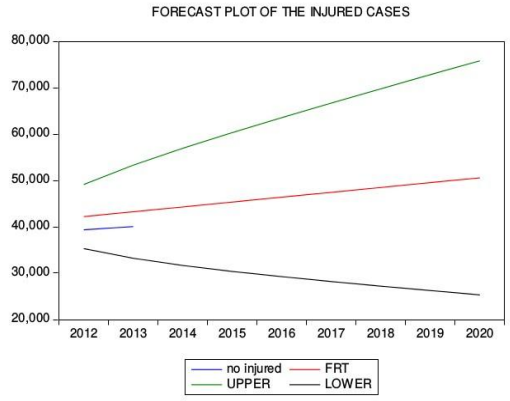


Figure 18a: Forecast Plot for the Injured cases.

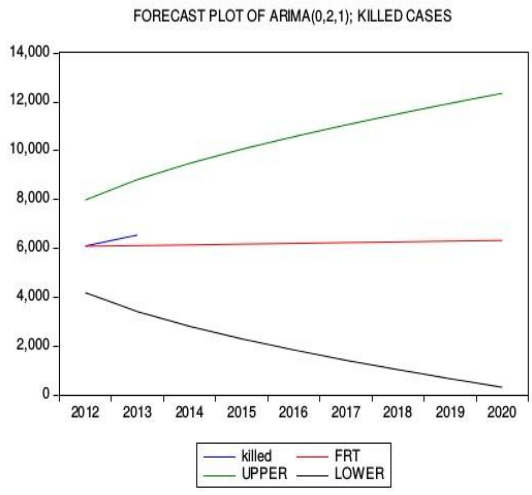


Figure 18b: Forecast Plot for the killed cases.

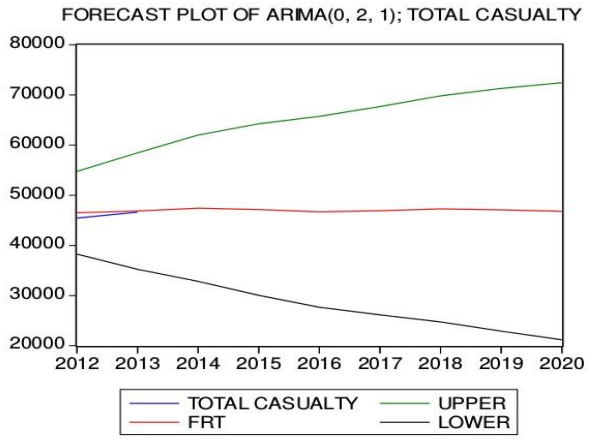


Figure 18c: Forecast Plot for the Total Casualties.

3.3 General Difference Form of the Models.

The general difference of ARIMA (0,2,1); Injured consequences is given as,

$Yt = 2Yt-1 - Yt-2 + et - \theta 1et-1,$

$Yt = 2Yt-1 - Yt-2 + et - \theta 1et-1.$

Substituting the value θ as given in Figure (9a), then the model for the Injured consequences becomes,

$Yt = 2Yt-1 - Yt-2 + et + 0.954388et-1.$

Also, the general difference of ARIMA (1,2,2); killed consequences is given as,

$Yt = 2Yt-1 - Yt-2 + \psi 1(Yt-1 - 2Yt-2 + Yt-3) + et - \theta 1et-1 - \theta 2et-2,$

but $\theta_1 = 0,$

$Yt = (2 + \psi 1)Yt-1 - (1 + 2\psi 1)Yt-2 + \psi 1Yt-3 + et - \theta 2et-2.$

Substituting the values of ψ and θ as given in Figure (9b), then the model for the killed consequences becomes,

$Y_t = 1.167534Y_{t-1} + 0.66492Y_{t-2} - 0.832466Y_{t-3} + e_t + 0.968707e_{t-2}.$

Lastly, the general difference of ARIMA (0,2,1); total casualty consequences is given as,

$$Yt = 2Yt-1 - Yt-2 + et - \theta 1et-1,$$

$$Yt = 2Yt-1 - Yt-2 + et - \theta 1et-1,$$

Substituting the value θ as given in Figure (10), then the model for the total casualty consequences becomes,

$$Yt = 2Yt-1 - Yt-2 + et + 0.959777et-1.$$

4. Discussion

Road traffic accident in Nigeria is increasing at a worrying and alarming rate and has raised one of the country major concerns. Federal Road Safety Corps of Nigeria recognizes the negative impacts of road safety accident and has commended the positive contribution of road safety researches as necessary tools to have significant accident initiatives. The paper was carried out in order to identify the patterns of road traffic accident consequences; injured, killed and total casualty by developing appropriate time series ARIMA models and predict 7 years consequences of road traffic accident; injured, killed and total casualty along the Nigeria motorway.

Time series analysis of the data from the years 1960-2013 showed that patterns of road traffic accident consequence; injured; killed and total casualty are increasing along the Nigeria motorway. The most widely used conventional method of time series known as Autoregressive Integrated Moving Average (ARIMA) model was applied to the annual-consequence of road accident data in Nigeria from 1960-2013 to determine patterns of road traffic accident consequences; injured, killed and total casualty of the road accident along the Nigeria motorway. After identifying various tentative models the appropriate models for the accident consequences; injured, killed and total casualty. ARIMA (0,2,1) model was found to be suitable model for the injury and total casualty consequences, whilst ARIMA(1,2,2) model was found to be suitable model for the killed consequences using the data from 1960-2011. The adequacy and performance of the model were tested on the remaining data from 2012 to 2013.

We provided 7 years forecasts of the consequences of road accident using the models developed and they showed that, road traffic accident consequences examined; injured, killed and total casualty will continue to increase. The study also revealed that road traffic accident cases; injured and killed along the motorway would continue to increase over the next 7 years. This study has provided reliable and genuine information that could be useful for determining road accident rate on Nigeria motorway and provide necessary prevention for the unwanted act. The study will also be used for providing important information in raising the level of awareness among stakeholders in road safety, since the problem has become a growing rife in Nigeria and also, be useful in setting priorities when planning road traffic accident interventions. Most Importantly, this study will provide expected benefit to the road users, Federal Road Safety Corps, researchers and other stakeholders in understanding the future rate of the consequences of road accident.

5. RECOMMENDATION

We have derived appropriate ARIMA Models that explain the behaviour and also the future patterns of the consequences of Road Accident along motor highway in Nigeria. Meanwhile, caution should be exercise in using the model, as it should not be used beyond the forecasted period, this is mainly because long time forecast may give arbitrary large forecast. Also, appropriate laws should be made to caution drivers that over-speed beyond the standard. Strict laws should be made to enforce the use of seat-belt among the driver and also, the passenger sitting in the front seat. This if enforced may reduce the critical state of the accident.

The Federal Road Safety Corp (FRSC) and all the stakeholders in charge of motorway in Nigeria should ensure proper maintenance of the motorway, it should be maintained in terms of the use of appropriate materials for patching pot holes, provision of street lights to aid visibility in the night, installation of traffic lights at new intersections created along the road. Also, proper education should be made known to the drivers on how to overtake on the motorway.

Appropriate training and retraining of drivers should be encourage towards reducing the carnage on over roads this will greatly reduce the rate of road traffic accident in the country. Road signals and signs that guide and instruct the drivers on what is happening in some kilometers ahead should always be made available on the motorway. Drivers should be discourage from receiving or making calls while driving.

6. REFERENCES

[1] Augustus O Atubi. Road traffic accident variations in lagos state, nigeria: A synopsis of variance spectra. *African Research Review*, 4(2), 2010.

[2] Cemal AYVALIK. Determinants of motor vehicle fatalities and fatality rates: Some preliminary findings for illinois.

[3] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley, 2013.

[4] Tom Brijs, Dimitris Karlis, Filip Van den Bossche, and Geert Wets. A bayesian model for ranking hazardous road sites. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(4):1001–1017, 2007.

[5] MOHAMMED A Hajeeh. Analysis of traffic problems in kuwait. 12:85–90, 2012.

[6] Goff Jacobs, Amy Aeron-Thomas, and Angela Astrop. *Estimating global road fatalities*. TRL, 2000.

[7] Cejun Liu and Chou-Lin Chen. Time series analysis and forecast of annual crash fatalities. *parameters*, 41876(37501):46251, 2004.

[8] CJL Murray, AD Lopez, World Health Organization, et al. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. *The global burden of disease. Cambridge (MA): Harvard School of Public Health*, 1996.

[9] Alan Ross, Chris Baguley, and Brian Hills. *Towards Safer Roads in Developing Countries. A Guide for Planners and Engineers. Prepared by the Ross Silcock Partnership on Behalf of and in Association With, the Overseas Unit of the Transport and Road Research Laboratory*. Ross Silcock Partnership, 1991.

[10] Wim Van Lerberghe. *The world health report 2008: primary health care: now more than ever*. World Health Organization, 2004.

[11] Wim Van Lerberghe. *The world health report 2008: primary health care: now more than ever*. World Health Organization, 2008.

[12] Jin Wen, P Yuan, ZH Deng, KL Liu, Yue-Kang Zhang, Li-Ke Liu, Bin Kong, and SX Huang. [time-series analysis on road traffic injury in china]. *Sichuan da xue xue bao. Yi xue ban= Journal of Sichuan University. Medical science edition*, 36(6):866–869, 2005.

[13] Chien-Ho Wu. Arima models are clicks away. *Applied mechanics and Materials*, 411- 414Publisher Trans Tech Publications, Switzerland.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 274 - 279, 2015, ISSN:- 2319–8656

Anonymizing and Confidential Databases for Privacy Protection Using Suppression and Generalization Based Protocols

K.Sathyamoorthy
Department of CSE
Panimalar Institute of Technology
Chennai,India

S.Venkata Lakshmi
Department of CSE
Panimalar Institute of Technology
Chennai,India

Tina Belinda Miranda
Department of CSE
Panimalar Institute of Technology
Chennai,India

Abstract—The technique of k-anonymization has been proposed in the literature as an alternative way to release public information, while ensuring both data privacy and data confidentiality. "X" owns a k-anonymous database and needs to determine whether "X" database, when inserted with a tuple owned by "Y", is still k-anonymous. Clearly, allowing "X" to directly read the contents of the tuple breaks the privacy of "Y". In this place,"Y" not get the privacy of own information because the information of "Y" can be accessed by "X" without the prior knowledge of "Y". On the other hand, the confidentiality of the database managed by "X" is violated once "Y" has access to the contents of database. Thus, the problem is to check whether the database inserted with the tuple is still k-anonymous, without letting "X" and "Y" knows the contents of the tuple and database respectively. In this paper, we propose two protocols solving this problem that is suppression-Based & Generalization-Based k-anonymous and confidential databases using through prototype architecture. And also those two protocols maintain privacy and confidential information in k-anonymous database.

Keywords — Confidentiality, Anonymity, Privacy, Secure Computation.

1. INTRODUCTION

TODAY'S globally networked society places great demand on the collection and sharing of person-specific data for many new uses [1]. This happens at a time when more and more historically public information is also electronically available. When these data are linked together, they provide an electronic image of a person that is as identifying and personal as a fingerprint even when the information contains no explicit identifiers, such as name and phone number. Other distinctive data, such as birth date and postal code, often combine uniquely[2] and can be linked to publicly available information to re-identify individuals. Data confidentiality is particularly relevant because of the value, often not only monetary, that data have. For example, medical data collected by following the history of patients over several years may represent an invaluable asset that needs to be adequately protected. Such a requirement has motivated a large variety of approaches aiming at better protecting data confidentiality and data ownership. Relevant approaches include query processing techniques for encrypted data and data watermarking techniques.

Data confidentiality is not, however, the only requirement that needs to be addressed. Today there is an increased concern for privacy. The availability of huge numbers of databases recording a large variety of information about individuals makes it possible to discover information about specific individuals by simply correlating all the available databases. Although confidentiality and privacy are often used as synonyms, they are different concepts: data confidentiality is about the difficulty by an unauthorized user to learn anything about data stored in the database. Usually, confidentiality is achieved by enforcing an access policy, or possibly by using some cryptographic tools. Privacy relates to what data can be safely disclosed without leaking sensitive information regarding the legitimate owner [5].

To better understand the difference between confidentiality and anonymity, consider the case of a medical facility connected with a research institution. Suppose that all patients treated at the facility are asked before leaving the facility to donate their personal health care records and medical histories to the research institution, which collects the records in a research database. To guarantee the maximum privacy to each patient, the medical facility only sends to the research database an anonymized version of the patient record. Once this anonymized record is stored in the research database, the non anonymized version of the record is removed from the system of the medical facility.

Thus, the research database used by the researchers is anonymous. Suppose that certain data concerning patients are related to the use of a drug over a period of four years and certain side effects have been observed and recorded by the researchers in the research database. It is clear that these data (even if anonymized) need to be kept confidential and accessible only to the few researchers of the institution working on this project, until further evidence is found about the drug. If these anonymous data were to be disclosed, privacy of the patients would not be at risk; however the company manufacturing the drug may be adversely affected. Recently, techniques addressing the problem of privacy via data anonymization have been developed, thus making it more difficult to link sensitive information to specific individuals.

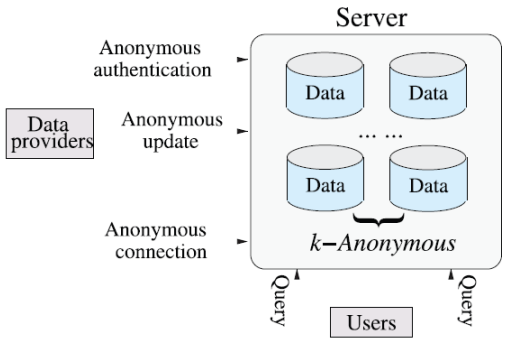


Figure 1: Anonymous Database

One well-known technique is k-anonymization. Such technique protects privacy by modifying the data so that the probability of linking a given data value, for example a given disease, to a specific individual is very small. So far, the problems of data confidentiality and anonymization have been considered separately. However, a relevant problem arises when data stored in a confidential, anonymity-preserving database need to be updated. The operation of updating such a database, e.g., by inserting a tuple containing information about a given individual, introduces two problems concerning both the anonymity and confidentiality of the data stored in the database and the privacy of the individual to whom the data to be inserted are related: 1) Is the updated database still privacy preserving? and 2) Does the database owner need to know the data to be inserted? Clearly, the two problems are related in the sense that they can be combined into the following problem: can the database owner decide if the updated database still preserves privacy of individuals without directly knowing the new data to be inserted? The answer we give in this work is affirmative. It is important to note that assuring that a database maintains the privacy of individuals to whom data are referred is often of interest not only to these individuals, but also to the organization owning the database. Because of current regulations, like HIPA organizations collecting data about individuals are under the obligation of assuring individual privacy. It is thus, in their interest to check the data that are entered in their databases do not violate privacy, and to perform such verification without seeing any sensitive data of an individual.

1.1 Problem Statement

Figure.1 captures the main participating parties in our application domain. We assume that the information concerning a single patient (or data provider) is stored in a single tuple, and DB is kept confidentially at the server. The users in Figure.1 can be treated as medical researchers who have the access to DB. Since DB is anonymous, the data provider's privacy is protected from these researchers. As mentioned before, since DB contains privacy-sensitive data, one main concern is to protect the privacy of patients. Such task is guaranteed through the use of anonymization. Intuitively, if the database DB is anonymous, it is not possible to infer the patients' identities from the information contained in DB. This is achieved by blending information about patients. Suppose now that a new patient has to be treated. Obviously, this means that the database has to be updated in order to store the tuple t containing the medical data of this patient.

The modification of the anonymous database DB can be naively performed as follows: the party who is managing the database or the server simply checks whether the updated database DB is still anonymous. Under this approach, the entire tuple t has to be revealed to the party managing the database server, thus violating the privacy of the patient. Another possibility would be to make available the entire database to the patient so that the patient can verify by himself/herself if the insertion of his/her data violates his/her own privacy. This approach however, requires making available the entire database to the patient thus violating data confidentiality. In order to devise a suitable solution, several problems need to be addressed: Problem 1: without revealing the contents of t and DB, how to preserve data integrity by establishing the anonymity of DB. Problem 2: once such anonymity is established, how to perform this update? Problem3: what can be done if database anonymity is not preserved? Finally, problem 4: what is the initial content of the database, when no data about users has been inserted yet? In this paper, we propose two protocols solving Problem 1, which is the central problem addressed by our paper. However, because the other problems are crucial from a more practical point of view. An approach that can be used is based on techniques for user anonymous authentication and credential verification [20]. The above discussion illustrates that the problem of anonymous updates to confidential databases is complex and requires the combination of several techniques, some of which are proposed for the first time in this paper. Figure.1 summarizes the various phases of a comprehensive approach to the problem of anonymous updates to confidential databases.

1.2 Proposed Solutions

All protocols we propose to solve Problem 1 rely on the fact that the anonymity of DB is not affected by inserting t if the information contained in t, properly anonymized, and is already contained in DB. Then, Problem 1 is equivalent to privately checking whether there is a match between (a properly anonymized version of) t and (at least) one tuple contained in DB. The first protocol is aimed at suppression-based anonymous databases, and it allows the owner of DB to properly anonymize the tuple t, without gaining any useful knowledge on its contents and without having to send to t's owner newly generated data. To achieve such goal, the parties secure their messages by encrypting them. In order to perform the privacy-preserving verification of the database anonymity upon the insertion, the parties use a commutative and homomorphic encryption scheme. The second protocol is aimed at generalization-based anonymous databases, and it relies on a secure set intersection protocol, such as the one found in [3], to support privacy-preserving updates on a generalization-based k-anonymous DB.

2. RELATED WORK

A preliminary approach to this problem was investigated in [33]. However, these protocols have some serious limitations, in that they do not support generalization-based updates, which is the main strategy adopted for data anonymization. Therefore, if the database is not anonymous with respect to a tuple to be inserted, the insertion cannot be performed. In addition one of the protocols is extremely inefficient. In the current paper, we present two efficient protocols, one of which also supports the private update of a generalization

based anonymous database. We also provide security proofs and experimental results for both protocols. So far no experimental results had been reported concerning such type of protocols; our results show that both protocols perform very efficiently.

The first research direction deals with algorithms for database anonymization. The idea of protecting databases through data suppression has been extensively investigated in the area of statistical databases [1]. The problem of protecting the privacy of time varying data have recently spurred an intense research activity which can be roughly divided into two broad groups depending on whether data are continuously released in a stream and anonymized in an online fashion, or data are produced in different releases and subsequently anonymized in order to prevent correlations among different releases. The second research direction is related to Secure Multiparty Computation (SMC) techniques. SMC represents an important class of techniques widely investigated in the area of cryptography. The third research direction is related to the area of private information retrieval, which can be seen as an application of the secure multiparty computation techniques to the area of data management. Here, the focus is to devise efficient techniques for posing expressive queries over a database without letting the database know the actual queries [10]. Thus, the goal is to protect data confidentiality from the external entities managing the data; however, data are fully available to the clients, which is not the case under our approach.

3. BASIC DEFINITIONS AND PRIMITIVES

3.1 Anonymity Definition: When using a suppression-based anonymization method, we mask with the special value ∂, the value deployed by Alice for the anonymization. When using a generalization-based anonymization method, original values are replaced by more general ones, according to a priori established value generalization hierarchies The information age has witnessed a huge growth in the amount of personal data that can be collected and analyzed. This has led to an increasing use of data mining tools with the basic goal of inferring trends in order to predict the future. However, this goal conflicts with the desire for privacy of personal data. In many scenarios, access to large amounts of personal data is essential in order for accurate inferences to be drawn. We adopt the following notations thereafter:

AREA	POSITION	SALARY
Data Mining	Associate Professor	$90,000
Intrusion Detection	Assistant Professor	$78,000
Handheld Systems	Research Assistant	$17,000
Handheld Systems	Research Assistant	$15,500
Query Processing	Associate Professor	$100,000
Digital Forensics	Assistant Professor	$78,000

TABLE.1. Original Data Set

AREA	POSITION	SALARY
*	Associate Professor	*
*	Assistant Professor	*
Handheld Systems	Research Assistant	*
Handheld Systems	Research Assistant	*
*	Associate Professor	*
*	Assistant Professor	*

TABLE 2.Suppressed with K=2

. Quasi-Identifier (QI): A set of attributes that can be used with certain external information to identify a specific individual.

. T½QI_: T½QI_ is the projection of T to the set of attributes contained in QI.

Definition 3.1. T½QI_ satisfies k-anonymity if and only if each record in it appears at least k times [32].

3.2.Cryptographic Primitives

A commutative, product-homomorphic encryption scheme ensures that the order in which encryptions are performed is irrelevant (commutativity) and it allows to consistently performing arithmetic operations over encrypted data (homomorphic property). Further, for the security proofs we require that the encryption scheme E satisfies the in distinguishability property. We extend the definition of commutative, indistinguishable encryption scheme presented in [3], in order to obtain an encryption scheme which also product-homomorphic. Given a finite set K of keys and a finite domain D, a commutative, product homomorphic encryption scheme E is a polynomial time computable function E : K X D→D satisfying the following properties:

1.Commutativity.

For all key pairs K_1,K_2€K andvalue d € D, the following equality holds:

$$E_{K1}(E_{K2}(d)) = E_{K2}(E_{K1}(d)) \quad (1)$$

2. Product-homomorphism.

For every K € K and everyvalue pairs d1,d2 € D, the following equality holds:

$$E_k(d_1) \cdot E_K(d_2) = E_K(d_1 \cdot d_2) \quad (2)$$

3. Indistinguishability

It is infeasible to distinguish an encryption from a randomly chosen value in the same domain and having the same length. In otherwords, it is infeasible for an adversary, with finite computational capability, to extract information about a plain text from the cipher text.

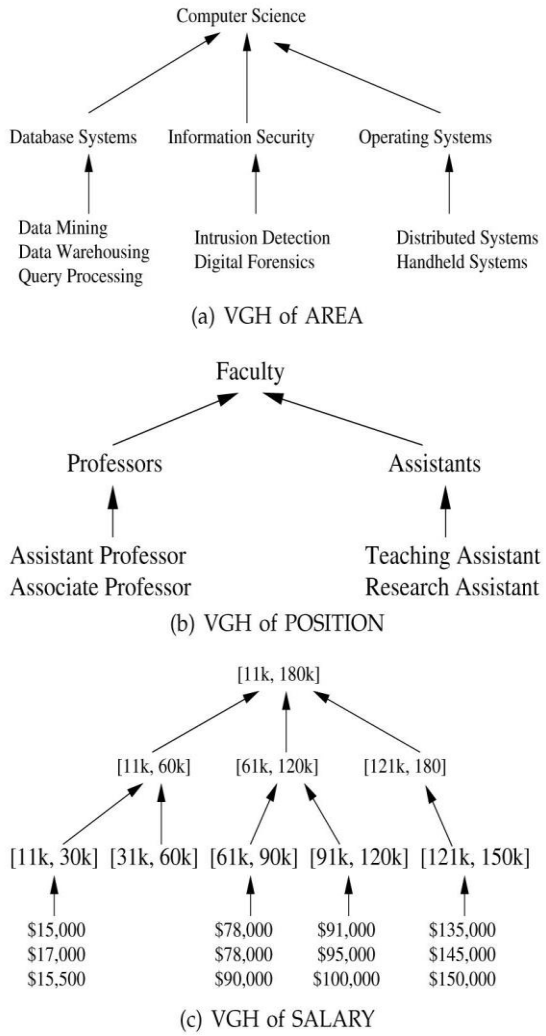


Figure. 2. Value Generalization Hierarchies.

AREA	POSITION	SALARY
Database Systems	Associate Professor	[61k, 120k]
Information Security	Assistant Professor	[61k, 120k]
Operating Systems	Research Assistant	[11k, 30k]
Operation Systems	Research Assistant	[11k, 30k]
Database Systems	Associate Professor	[61k, 120k]
Information Security	Assistant Professor	[61k, 120k]

TABLE 3: Generalized with K=2

4. PRIVATE UPDATE BASED ON SUPPRESSION BASED PROTOCOLS

The idea of suppressing an attribute is a simple concept. A value is replaced by a less specific, more general value that is faithful to the original. In a classical relational database system, domains are used to describe the set of values that attributes assume. For example, there might be a ZIP domain, a *number* domain and a *string* domain. In the original database, where every value is as specific as possible, every attribute is considered to be in a ground domain. In this section, we assume that the database is anonymized using a suppression-based method. Note that our protocols are not required to further improve the privacy of users other than that provided by the fact that the updated database is still k-anonymous.It allows the owner of DB to properly anonymized the tuple t, without gaining any useful knowledge on its contents and without having to send to t's owner newly generated data. To achieve such goal, the parties secure their messages by encrypting them. In order to perform the privacy-preserving verification of the database anonymity upon the insertion, the parties use a commutative and homomorphic encryption scheme.

Suppose that X owns a k-anonymous table T over the QI attributes. X has to decide whether T U t, where t is a tuple owned by Y is still k-anonymous, without directly knowing the values in t(assuming t and T have the same schema). This problem amounts to decide whether t matches any tuple in T on thenon-suppressed QI attributes. If this is the case, then t,properly anonymized, can be inserted into T. Otherwise, the insertion of t into T is rejected. A solution that addresses such drawback is based on the following protocol. Assume, X and Y agree on acommutative and product-homomorphic encryption scheme. Unless otherwise stated, the term *data* refers to person-specific information that is conceptually organized as a table of rows (or records) and columns (orfields). Each row is termed a *tuple*. Tuples within a table are not necessarily unique. Each column is called an *attribute* and denotes a semantic category of information that is a set of possible values; therefore, an attribute is also a domain. Attributes within a table are unique. So by observing a table, each row is an ordered *n*-tuple of values <*d1*, *d2*, …, *dn*> such that each value *dj*is in the domain of the *j*-th column, for *j*=1, 2, …, *n* where *n* is the number of columns.

Protocol 4.1

1. X codes his tuple ∂_i into c([v'$_1$…….v'$_{t]}$, denoted asC(∂_i). Then, X encrypts c(∂_i)with his private key and sends $E_A(c(\partial_i))$to Y.

2. Y individually codes each attribute value in t to get the tuple of coded values [c(v_1) . . . c(v_u)] encrypt search coding and $E_A(c(\partial_i))$ with his key B and sends (i)

[$E_B(c(v_1))$ . . . $E_B(c(v_u))$], and (ii) $E_B(E_A(c(\partial_i)))$ to X.

3. Since E is a commutative encryption scheme,

$E_B(E_A(c(\partial_i))) = E_A(E_B(c(\partial i)))$,

X decrypts $E_A(E_B(c(\partial_{i)}))$ to obtain $E_B(c(\partial_i))$

4. Since the encrypted values sent by Y are ordered according to the ordering of the attributes in T(assume this is a public information known to both X and Y), X knows which are, among the encrypted values sent by Y, the one corresponding to the suppressed and non suppressed QI attributes. Thus, X computes

$E_B(c(v_1))$ X . . . X $E_B(c(v_s))$ (5)

where v_1 . . . v_s are the values of nonsuppressed attributes contained in tuple t. As already mentioned, E is a product-homomorphic encryption scheme. Based also on the definition of function c(∂), this implies that Expression 5 is equal to

$E_B(c([v_1 \ldots v_s]))$ (6)

5. X checks whether

$E_B(c([v_1 \ldots v_s])) = E_B(c([v_1 \ldots v_0]))$

If true, t (properly anonymized) can be inserted to table T. Otherwise, when inserted to T, t breaks k-anonymity.

5. PRIVATE UPDATE USING GENERALIZATION BASED

CONFIDENTIAL AND ANONYMOUS DATA

The idea is that there are some numbers of records in *T* that can be considered outliers. For this reason, up to a certain number of records (the *maximum suppression threshold*) may be completely excluded from *V*. Under this combined scheme, *V* is obtained through full-domain generalization, with selected outlier tuples removed entirely. For any anonymization mechanism, it is desirable to some notion of minimality. Intuitively, a k-anonymization should not generalize, suppress, or distort the data more than is necessary to achieve k-anonymity. Indeed, there are a number of ways to minimality. One notion of minimal full-domain generalization using the distance vector of the domain generalization. Informally, this definition says that a full-domain generalization *V* is minimal if *V* is k-anonymous, and the height of the resulting generalization is less than or equal to that of any other k-anonymous full-domain generalization.

In this section, we assume that the table T is anonymized using a generalization-based method; let $T_1. . . . .T_u$ be udisjoint VGHs corresponding to $A_1…..A_u€ A_t^{anon}$ known to X. Let ∂ € T, and let GetSpec(∂[$A_1 . . .A_u$],$T_1 . . . T_u$) (GetSpec(∂) for short) denote a function which returns a set $_\gamma$ of specific values (values at the bottom of a VGH)related to each attribute $A_i€A_t^{anon}$ such that every value in $_\gamma$ can be generalized to ∂[Ai] for some i according to T_i. For example, let T refer to Table 4 and A_t^{anon}={AREA,POSITION, SALARY}. If T= [Operating Systems, Research Assistant, [11k, 30k]], then based on the VGHs (presented in Figure. 2) GetSpec(T) = {Distributed Systems, Handheld Systems, Research Assistant, $15,000, $17,000, $15,500}.

Protocol 5.1

1. X randomly chooses a ∂ € T_w.
2. Y compute $^\gamma$ = GetSpec(∂).
3. X and Y collaboratively compute s = SSI(γ, T).
4. If s = u then t's generalized form can be safely insertedto T.
5. Otherwise, X computes T_w←T_w– {∂} and repeatthe above procedures until either s = u or $T_{u=¥}$

5.1 Security Analysis

The security of Protocol 5.1 depends on that of the SSI protocol, and detailed security analyses of SSI can be found in [3], [13]. The SSI protocol presented in [3] is easy to implement and efficient to perform. Although the protocol leaks the intersection size between $^\gamma$ and T to the participating parties, it does provide sufficient privacy protection in our application. In case this linkage of intersection sizes is not acceptable, we can adopt one variation of the SSI protocol presented in [13]. We can make the protocol only return whether or not acceptable without disclosing the intersection size. Under the context of Secure Multiparty Computation, this variation of SSI does not leak any information that cannot be inferred from the final result and the private input data. Thus, using SSI proposed in [13], Protocol 5.1 can achieve very high security.

6. ARCHITECTURAL DESIGN

Our prototype of a Private Checker (that is, X) is composed by the following modules: a crypto module that is in charge of encrypting all the tuples exchanged between an user (that is, Y) and the Private Updater, using the techniques and a checker module that performs all the controls, as prescribed by Protocols 4.1 and 5.1; a loader module that reads chunks of anonymized tuples from the k-anonymous DB. The chunk size is fixed in order to minimize the network overload. In Figure. 3 such modules are represented along with labelled arrows denoting what information are exchanged among them. Note that the functionality provided by the Private Checker prototype regards the check on whether the tuple insertion into the k-anonymous DB is possible. We do not address the issue of actually inserting a properly anonymized version of the tuple.

The information flow across the above mentioned modules is as follows: after an initial setup phase in which the user and the Private Checker prototype exchange public values for correctly performing the subsequent cryptographic operations, the user sends the encryption E(c(∂i)) of her/his tuple to the Private Checker; the loader module reads from the k-anonymous DB the first chunk of tuples to be checked with E(c(∂i)). Such tuples are then encrypted by the crypto module.

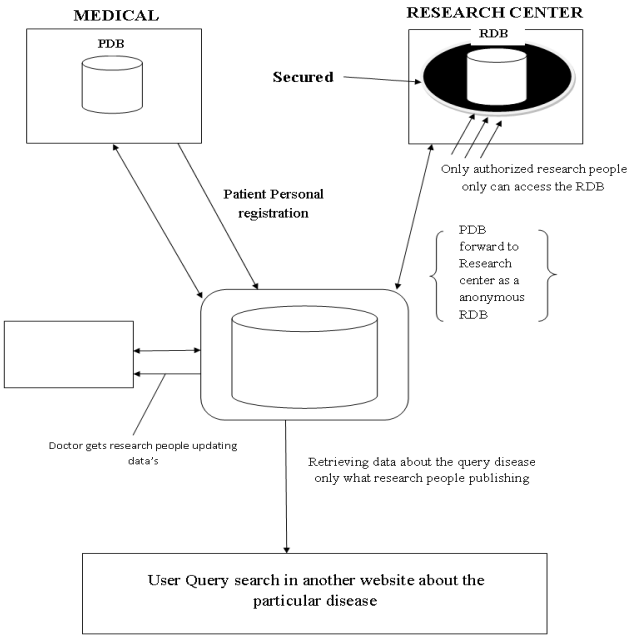


Figure. 3. Prototype architecture overview.

7. CONCLUSION / FUTURE WORK

In this paper, we have presented two secure protocols for privately checking whether k-anonymous database retains its anonymity once a new tuple is being inserted to it. Since the proposed protocols ensure the updated database remains k-anonymous, the results returned from a user's (or a medical researcher's) query are also k-anonymous. Thus, the patient or the data provider's privacy cannot be violated from any query. As long as the database is updated properly using the proposed protocols, the user queries under our application domain are always privacy-preserving. In order for a database system to effectively perform privacy preserving updates to a k-anonymous table, Protocols 4.1 and 5.1 are necessary but clearly not sufficient. As already mentioned in Section 1, other important issues are to be addressed:

1. The definition of a mechanism for actually performing the update, once k-anonymity has been verified.

2. The specification of the actions to take in case Protocols 4.1 or 5.1 yield a negative answer.

3. How to initially populate an empty table.

4. The integration with a privacy-preserving query system.

In addition to the problem of falling insertion, there are other interesting and relevant issues that remain to be addressed:

- Devising private update techniques to database systems that supports notions of anonymity different than k-anonymity.
- Dealing with the case of malicious parties by the introduction of an untrusted, non colluding third party .
- Implementing a real-world anonymous database system.
- Improving the efficiency of protocols, in terms of number of messages exchanged and in terms of their sizes, as well.

We believe that all these issues are very important and worthwhile to be pursued in the future.

8. REFERENCES

[1] N.R. Adam and J.C. Wortmann, "Security-Control Methods for Statistical Databases: A Comparative study," ACM Computing Surveys, vol. 21, no. 4, pp. 515-556, 1989.

[2] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy,D. Thomas, and A. Zhu, "Anonymizing Tables," Proc. Int'l Conf .Database Theory (ICDT), 2005.

[3] R. Agrawal, A. Evfimievski, and R. Srikant, "Information Sharingacross Private Databases," Proc. ACM SIGMOD Int'l Conf .Management of Data, 2003.

[4] C. Blake and C. Merz, "UCI Repository of Machine MLRepository.html, 1998.

[5] E. Bertino and R. Sandhu, "Database Security—Concepts, Approaches and Challenges," IEEE Trans. Dependable and SecureComputing, vol. 2, no. 1, pp. 2-19, Jan.-Mar. 2005.

[6] D. Boneh, "The Decision Diffie-Hellman Problem," Proc. Int'lAlgorithmic Number Theory Symp., pp. 48-63, 1998.

[7] D. Boneh, G. di Crescenzo, R. Ostrowsky, and G. Persiano, "PublicKey Encryption with Keyword Search," Proc. Eurocrypt Conf., 2004.

[8] S. Brands, "Untraceable Offline Cash in Wallets with Observers,"Proc. CRYPTO Int'l Conf., pp. 302-318, 1994.

[9] J.W. Byun, T. Li, E. Bertino, N. Li, and Y. Sohn, "Privacy-Preserving Incremental Data Dissemination," J. Computer Security,vol. 17, no. 1, pp. 43-68, 2009.

[10] R. Canetti, Y. Ishai, R. Kumar, M.K. Reiter, R. Rubinfeld, and R.N.Wright, "Selective Private Function Evaluation with Application to Private Statistics," Proc. ACM Symp.Principles of DistributedComputing (PODC), 2001.

[11] S. Chawla, C. Dwork, F. McSherry, A. Smith, and H. Wee,"Towards Privacy in Public Databases," Proc. Theory of CryptographyConf. (TCC), 2005.

[12] U. Feige, J. Kilian, and M. Naor, "A Minimal Model for SecureComputation," Proc. ACM Symp. Theory of Computing (STOC),1994.

[13] M.J. Freedman, M. Naor, and B. Pinkas, "Efficient PrivateMatching and Set Intersection," Proc. Eurocrypt Conf., 2004.

Mr.K.Sathyamoorthy is currently working as a Assistant Professor, Computer Science and Engineering, at Panimalar Institute of Technology, Poonamalle, Chennai 600123. He has completed his graduate in Anna University and Post graduate in Sathyabama University. His research interest includes image processing and soft computing.

Ms.S.Venkata Lakshmi is currently working as a Assistant Professor Grade 1, Computer Science and Engineering, at Panimalar Institute of Technology, Poonamalle, Chennai 600123.She has completed her graduate in M.S.University and Post graduate in DR.MGR university where she is currently working toward the Ph.D degree. She has published research papers in International, National conference proceedings. Her research interest includes image processing and medical imaging.

Ms.Tina Belinda Miranda is currently working as an Lecturer, Computer Science and Engineering at Panimalar Institute of Technology, Poonamalle, Chennai 600123.She has completed her graduate in Anna university and Post graduate in Anna university. Her research interest includes networks and Image processing.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 280 - 282, 2015, ISSN:- 2319–8656

Location Provider with Privacy Using Localized Server and GPS

Pintu.R
Department of CSE
Bharath University
Chennai, India

Rahul Gupta
Department of CSE
Bharath University
Chennai, India

G.Michael
Department of CSE
Bharath University
Chennai, India

Abstract:
Maps are an essential part of any handheld device and use constantly used for navigation and other resources by application for providing location based data which can be used for customized examination outcomes, however these data are conservatively stowed on a L.B.S Server which are susceptible to attacks and misuse as these data's are not usually have any significant security so these data's can be sold or misused by some other parties. We try to eliminate the problem as well as provided added functionality to the conventional maps by providing a customizable map which has the added functionality of offline use mode in addition to the online mode.
Keywords: GPS, MySQL Server, Maps, Android, SQLite

1. INTRODUCTION

Information agents are software products for assisting and guiding users to reach the goal of information retrieval. Up to now, however, most of Web information agent systems are closely related to the traditional information equipment's that cannot directly apply to the modern mobile equipment resulting from the core part of information agent in ubiquitous environments [1], [2]. This study exactly focuses on how to design a ubiquitous interface agent with mobile equipment's in ubiquitous environments.[3], [4] Ubiquitous computing is a post-desktop model of human computer interaction in which information processing has been thoroughly integrated into everyday objects and activities. Cloud computing is a technique of Internet- ("cloud-") based development and use of computer technology. Furthermore, how to construct an interaction diagram of cloud computing for extensively and seamlessly entering related web information agent systems through modern mobile equipment's in ubiquitous environments is under our investigation.

2. RELATED WORKS

In this paper,A new ubiquitous informational agent system with the GPS and Bluetooth techniques in the Google Android platform and related interaction diagrams with Onto IAS in cloud computing environments was proposed in this paper. It also explains how GPS

Devices access location based services by connecting to the google cloud server. With the propagation of portable expedients, impulsive connections between co-located devices that do not know each other a priori will become commonplace.

In this paper[5],[6] we postulate that mobile devices that are positioned in close proximity may be able to derive a shared secret to secure their communication by monitoring fluctuations in the signal strength of existing ambient radio sources (GSM cell towers or Wi-Fi access points) in their general environment. We explore the probability of deriving location-based secrets and describe two approaches for how such a secret could be used to secure spontaneous communication. Derive location-based secrets is a hard problem because while the radio environment perceived by various devices in close proximity is comparable, it is not undistinguishable. Global Positioning System (GPS) technology is changing the way we work and play.

In this paper you can use GPS technology when you are driving, flying fishing, marine, climbing, one after the other, biking, running, or exploring. Here are just a few examples of how you can use GPS technology Know precisely how far you have run and at what pace while tracking your path so you can find your way home Get the closest location of your favorite restaurant when you are out-of-town Find the nearest airport or identify the type of airspace in which you are flying., In existing system though we have many web sites gateway, and the user can get data about the searched location in web sites The user can be able to access the location only in online these systems are not available to not all time. And also, this information does not reach people at the time of emergencies

In this paper, in existing system if the user search some location the high level data information will be displayed on the screen. Whenever user want to search the places user need to specify the location of that particular area. in existing system user can be able to access the information which is stored in the server. The problem is to implement security to the location based services for the GPS devices on android platform. This is done by using a server implemented along with the main server to authenticate users. By implementing server additional features like real time location services are also made available to the user.

3. PROPOSED WORK

In the Proposed System of completion, user makes the query to the main record through our application. In proposed system Ontology is also implemented in this Project for the Relative Key word Search. If user search some location through our application it's take the current location of the application user using GPS. As the user makes the query along with his GPS values to find out the Exact Location from the data base, this process the query and authenticate the user and then passes the query to the main database The main system maps the query with its database along with the Location, auxiliary this location based in turn is passed back to the user. In existing system only contain the high level data but in our project

contain the deep level data because those data are collect from locally.

.

4. EXPERIMENTAL SETUP AND DESIGN

Using the GPS we locate the exact location of the user and the data is sent to the user. This data accessed by the user can be downloaded by the user and saved onto his device which can be used in offline mode after which search can be made in offline mode and LBS is not contacted thus location privacy is maintained. Using Wi-Fi and Dead Reckoning the locations in indoor location is predicted, where Wi-Fi based location prediction is used for indoor and dead reckoning is used to predict the location of user based on previously determined location. Now this downloaded data can be edited by user and user can add information

And shared via the app and people using the application can view the information instead of LBS open availability..

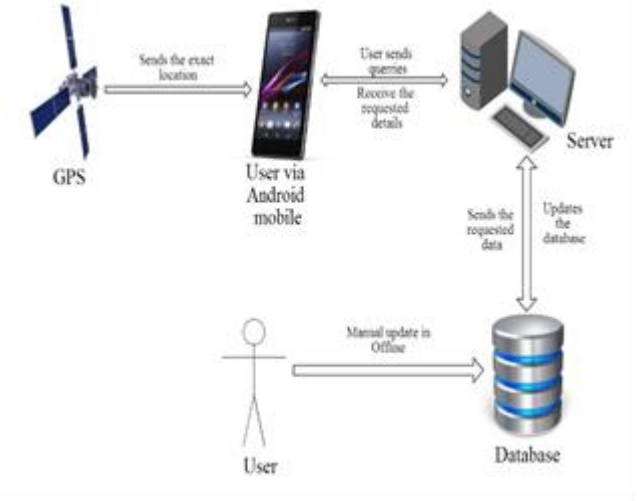


Figure. 1 System Architecture

5 MODULE DESCRIPTION

5.4.1 Server Modules

This module is mainly used to search the location of particular place in online from our server. The user can able to download the location to the local database consuming the server. After that user can be able to Examine the data in offline. In this server only the user can be able to upload the location to the server

5.4.2 Searching in Offline and Online using Ontology

In this module user can be able to search the location in online and in offline mode The current location of the user can be obtained using Indoor tracking and dead reckoning techniques Indoor tracking is a network of device used to wirelessly(WIFI,GPS) locate the user current location who is located inside the house . Dead reckoning is a procedure that is designed to guess current location based on the previously determined location

5.4.3 Add and Share location

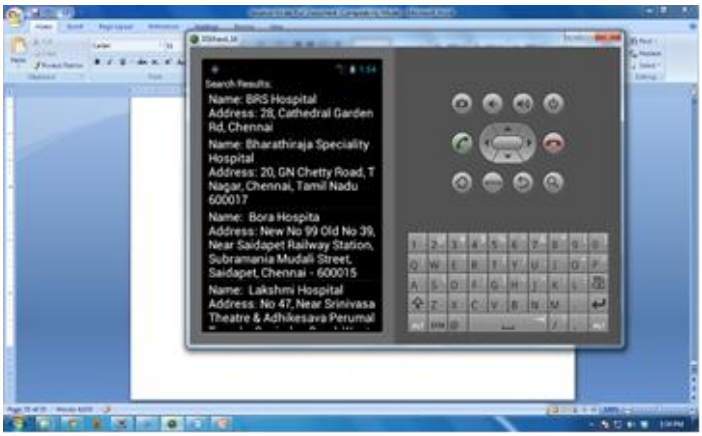
In this module the user can be able to add the location details of particular place to the server by using data offloading method Data offloading method is used to deliver the original data's to the targeted server from the mobile device After adding the location to the server and server share the location to the other user who are all using this application

5.4.4 Updating offline database

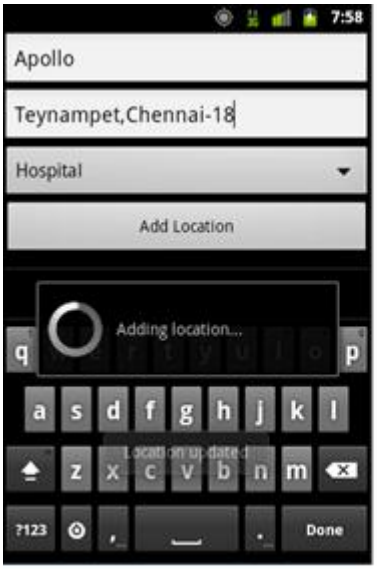
In this modules the user can be able to update the location details from the server to the offline database In our application we are using SQLite as an offline database the user get the data from the centralized server and data will be send to the SQLite.

6. IMPLEMENTATION AND WORKING

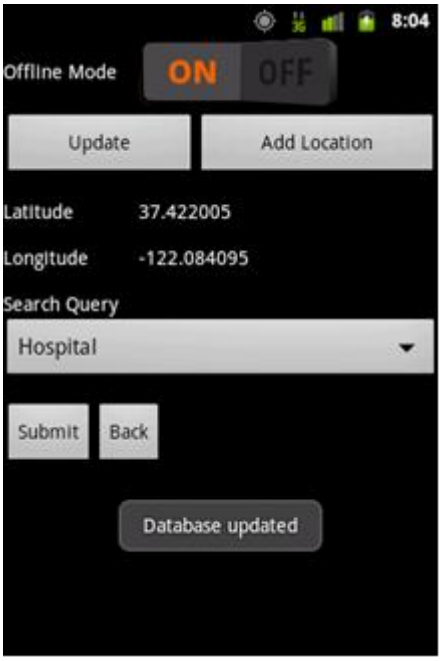
This is a screenshot of the android emulator running the map application on the system and shows the first module where we can make search on the map



We see here in the second screenshot the search results for a search for hospitals made



In this next screenshot we see the add location module we can edit the downloaded maps and customize and add locations to be shared



In this next screenshot we see how the offline mode is enabled to access the search in offline mode i.e. when not connected to internet still the locations can be accessed.

7. ACKNOWLEDGMENTS

We would like to extend our thanks to our Guide Prof Michael for his invaluable suggestions and guidance for the duration of the project and also to the all the people who have contributed for the project in any way .

8. CONCLUSION

We propose here a location provider which aims to provide the Location of user through a customizable map service for android devices which can be used in online as well as offline mode.

The request avoids the outdated LBS server based approach which is susceptible to attacks and misuse of data also using a customizable map a user can add functionality to maps and let other people see the updates made to the maps.

9. REFERENCES

[1] Alex Varshavsky, Antony Lamarca, Eyal Lara (2007)" Enabling secure and spontaneous communication between mobile devices using common radio environment" .IEEE Workshop on Mobile Computing Systems and Applications (Hot Mobile), Tucson, AZ, February 2007.

[2] Garmin International, Inc., KS, USA (2000)"GPS guide for beginner"..

[3] Patridge.K and P.Golle (2008)"On using existing time-use study data for ubiquitous computing applications".UbiComp '08 Proceedings of the 10th international conference on Ubiquitous computing

[4] Yang.S.Y (2009) "A Study on Developing an Ontology-Supported Information Agent Shell". National Science council project, NSC-98-2221-E-129-012(NT$:654,000), Taipei, Taiwan, 2009.

[5] Yang, Sheng-Yuan Lee, Dong-Liang(2011)"A new ubiquitous information agent system for cloud computing - example on gps techniques in google android platform". 01/2011; DOI:10.1109/COMPSYM.2010.5685438 In proceeding of: Computer Symposium (ICS), 2010 International.

[6] Yang.S.Y, K.W.Wu, C.M.Ni, Y.T.Lin and P.S.Kao (2010)" An ontology-supported ubiquitous interface agent for cloud computing – example on Bluetooth wireless technique with java programming". Proc. Of 2010 Conference on Information Technology and Applications in Outlying Islands, Kaohsiung, Taiwan, 2010

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 283 - 286, 2015, ISSN:- 2319–8656

Secure Personal Health Records Using Encryption

D. Devikannniga
Dept. of Information Technology
Sri Ramakrishna Engineering College
Coimbatore, India

S.N. Gomathi Balan
Dept. of Information Technology
Sri Ramakrishna Engineering college
Coimbatore, India

Abstract-In the dispersed world, health information is exchanged based on the patients Personal Health Records (PHRs). Due to this reason, the construction and maintenance are focused by data centers, which are used for persons to gain high cost. The cloud providers are used in most of the PHR services to outsource the PHRs, which are stored by third party. The privacy is main anxiety because the PHRs information is shared to third party servers and illegal parties. To avoid this problem and to provide the guarantee security for PHRs, the encryption is applied for all PHRs before it is outsourcing. After encryption is applied still few major issues are present such as, flexible access, scalability in key organizations and well organized user revocation. These are the residual important challenges. In this proposed system, a patient-centric model has been generated with appropriate mechanisms for accessing PHR which are stored in semi confidential servers. Here the Attribute Based Encryption technique is used to encrypt every patients PHR's. To support on demand, user revocations are also enabled dynamically based on the variations of access policies or file attributes to improve the process.

Keywords-cloud computing, personal health records.

1. INTRODUCTION

Personal Health Record (PHR) is an upcoming concept. Network Security can be preventing an unauthorized access. The network is controlled by Network administrator. Through access control policies authorized data can be handled by network security. The authorized people, who are accessing the authorized information, can be identified by their ID and password. Network security having different computer networks, such as public and private that

are used in everyday's transaction and communication of business. Networks can be private but it is used by public to access the information.

Our approach is to encrypt the data before outsourcing. Our approach consist of two modules the doctor and the lab technician. The doctor can give certain information of the patient to the lab technician and he may send the report to the doctor after test.

In our approach we used ABE encryption and Advanced Encryption Standard to encrypt the files of the patient

2.RELATED WORKS

The traditional encryption techniques were applied to the personal health record at the early stages of the cloud computing and personal health record. Which is not secure nowadays and so attribute based encryption[1][3] with various variations is used

2.1 Symmetric Key Cryptography (SKC) based Solutions

Symmetric key algorithms in cryptography use the same cryptographic keys for both encryption and decryption of text. Based on the symmetric key derivation methods, various solution for securing outsourced data on semi-trusted servers has been proposed,

2.2 Public Key Cryptography (PKC) based Solutions

The most traditional method applied to the PHR for the security of data was public key encryption method. It is very less scalable in high key management. In one–to-one encryption techniques, in which break glass access is not possible during emergencies.

2.3 Attribute Based Encryption based Solutions

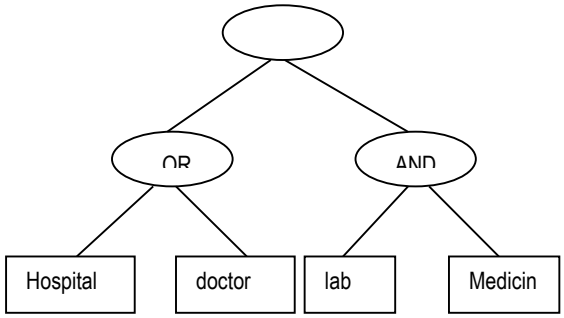


Fig.1. Attribute Based Encryption

The attributes can define an object very efficiently just as the identity of an object works In ABE system both the cipher text and secret key will be depended on the attributes. In such a system , the decryption of a cipher text is possible only if the set of attributes of the user key matches the attributes of the cipher text. So while applying this method the owner doesn't want to know about the entire list of users instead of that they can encrypt the data according to some attribute only.

3. PROBLEM DEFINITION

The problem is being extended to a wide range of PHR system, where there are multiple PHR owners and users. The patients whose health data are being controlled. There exist semi trusted servers where patients store their health details and the users have access to those data. The access rights various according to the users such as some can have read access alone and some can have both read and write access. These access rights are provided by the owner of the corresponding PHR. The PHR document can be handled by multiple owners and so Multi-Authority Attribute Based Encryption (MA-ABE) is .

4. PROPOSED SOLUTION

4.1 Architecture

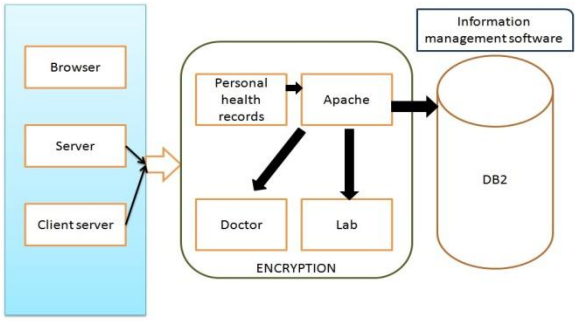


Fig.2. System Architecture

PHRs are stored and shared on semi trusted server. The patient health report is based on ABE algorithm. The owner allows multiple users to access patient health report with multiple authorities. Access rights are given based on authority. For this purpose MA-ABE techniques are proposed. An attribute wise encryption is done on health records.

Patient medical record is stored on server in encryspted format. Attribute wise encryption is done using ABE. PHR owners only knew who can access the health record. The owners distribute the secret keys for read and write data to authorized persons. The owner previously provides the temporary key to the emergency department. Breaking the normal procedure in an emergency period to access the whole data is called "Break-glass". This is clearly explained in Figure 2.

4.2 Design of Modules

This detailed insight of the design and working model of the proposed attribute based encryption for fine-grained access control is described below. The operations of proposed medical record sharing system combine KP-ABE traditional cryptography, allowing patients to share their medical records. These operations can be classified into following modules: In this section we discuss main module design concept for sharing of medical records using attribute based encryption.

Modules of the system are:

1. System Setup and Secret Key Generation
2. Encryption of Medical Records

3. View Medical Records (Decryption)

4.2.1 System Set-Up and Key-Generation

The KP-ABE algorithm takes security parameter as input. Using this input it provides public key and master secret key as output. Public key is specified as Pk and master secret key is specifies as Mk. Pk is used for encryption purpose by message senders. Mk is used for generating secret keys which is known by the authority.
user obtains secret key from the data owner through secure email by sending a request for the keys or data owner send the secret key to personal domain user via secure email.

4.2.2 Encryption

In attribute encryption the Sender runs randomized algorithm. It takes a given attribute for each user, the generated public key and a message as input. It provides cipher text as outputs.

The files can be requested by the user or the authority and can be viewed.

4.2.3 View Medical Record File /Decryption

The decryption uses deterministic algorithm which is run by patient or the doctor. It takes cipher text as input, which was encrypted and decrypted under the given set of attributes. The output is the patient health records.

5. IMPLEMENTATION

5.1 Algorithm for Attribute Key Setup

Step 1: KP-ABE Setup: Outputs public key and Master key for A as Set of attributes.

Step 2: Associate for each attribute in A with attribute universe as U= {1, 2, 3, …, n}.

Step 3: Associate each attributes i U with a number t_i and also chose y uniformly at random in public parameter (Zp*) and y.

Step 4: The public key is: PK = (T1= gt1,…, T1 = gt | U | , Y = e (g, g) y).

Step 5: The master key is: MK = (t1,…, t | U | y).

6. RESULT ANALYSIS

6.1 Analysis of Fine-Grained Access Control

User access structure is able to describe sophisticated logics over attributes. Each patients secret key has a unique secret sharing scheme which don't "match" each other[2][5]. Different types of users need access to different types of data in different phases by giving read and write permission to the public domain shown as fig.3.a and fig.3.b

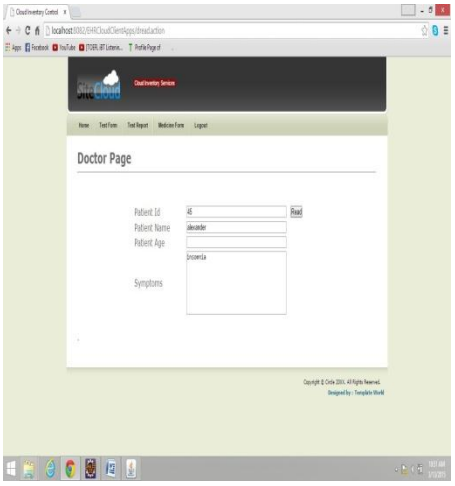


Fig.3.a. Fine-Grained Access Control Doctor page

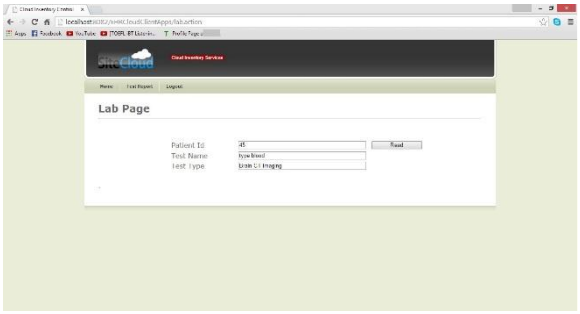


Fig.3.b: Fine-Grained Access Control lab page

In this access control method, the users access is limited. Based on the attribute the users defined access will be provided.. The policy updating is possible by updating the attribute or access policy in the system[4]. In emergency break glass access is provided.

6.2 ABE over Analysis of Fine-Grained Access Control

In the PHR system the users will be from different domain like the doctors, lab technician, and patient. Each user will be having different access control mechanism over the record.

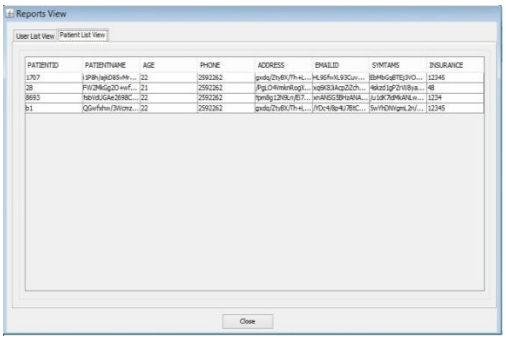


Fig.4. View of Health Records

The MA-ABE scheme will highly reduce the key-management issues. Users selects the attributes and the medical file wants to encrypt , Uploaded medical record can view by the other user as shown in fig.4. The users can view the record by providing the secret key matching the encrypted file.

7. APPLICATION

Medical centre ,organizations which try to secure their/employee health records can use this application

8. CONCLUSION

The personal health record system needs to be secure our application provides basic securities to protect the information from unauthorized access and loss. Our approach provides more security than traditional security which is easily hackable.

9. REFERENCES

[1] Ming Li and Shucheng Yu,"Scalable and Secure Sharing of PersonalHealth Records in Cloud Computing Using Attribute-Based Encryption," IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 1, January 2013

[2] M. Li, S. Yu, K. Ren, and W. Lou, "Securing Personal Health Records in Cloud Computing: Patient-Centric and Fine-Grained Data Access Control in Multi-Owner Settings," Proc. Sixth Int'lICST Conf. Security and Privacy in Comm. Networks (SecureComm '10), pp. 89-106, Sept. 2010

[3] L. Ibraimi, M. Asim, and M. Petkovic, "Secure Management of Personal Health Records by Applying Attribute-Based Encryption," technical report, Univ. of Twente, 2009

[4] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data," Proc. 13th ACM Conf. Computer and Comm. Security (CCS '06), pp. 89-98, 2006

[5] M. Li, S. Yu, K. Ren, and W. Lou, "Securing Personal Health Records in Cloud Computing: Patient-Centric and Fine-Grained Data Access Control in Multi-Owner Settings," Proc. Sixth Int' ICST Conf. Security and Privacy in Comm. Networks (SecureComm '10), pp. 89- 106, Sept. 2010

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 287 - 290, 2015, ISSN:- 2319–8656

A Formal Machine Learning or Multi Objective Decision Making System for Determination of Weights

Ganesan.R
Department of C.S.E
Bharath University
Chennai, India

Shanmuga Priyan.P
Department of C.S.E
Bharath University
Chennai, India

Kerana Hanirex.D
Dept. of C.S.E
Bharath University
Chennai, India

Abstract: Decision-making typically needs the mechanisms to compromise among opposing norms. Once multiple objectives square measure is concerned of machine learning, a vital step is to check the weights of individual objectives to the system-level performance. Determinant, the weights of multi-objectives is associate in analysis method, associated it's been typically treated as a drawback. However, our preliminary investigation has shown that existing methodologies in managing the weights of multi-objectives have some obvious limitations like the determination of weights is treated as one drawback, a result supporting such associate improvement is limited, if associated it will even be unreliable, once knowledge concerning multiple objectives is incomplete like an integrity caused by poor data. The constraints of weights are also mentioned. Variable weights square measure is natural in decision-making processes. Here, we'd like to develop a scientific methodology in determinant variable weights of multi-objectives. The roles of weights in a creative multi-objective decision-making or machine-learning of square measure analyzed, and therefore the weights square measure determined with the help of a standard neural network.
Keywords: Decision Making, Machine Learning, Data Integrity, Optimization, Constraints.

1. INTRODUCTION

System Science typically considers the knowledge domain interactions among a group of subsystems in individual disciplines as a result of associate rising trend and therefore the desires of finding out ever increasingly advanced systems in engineering, society, setting, and science, intensive analysis works are performed within the field of system science. The careful discussions and basic ideas during this field will be found in several literatures. As an example, Warfield stressed that system science could be a important conception with the potential to serve several functions and comprehend several things in its view. Bailey mentioned the history of the event of system science and known 10 goals together with 10 challenges; among these challenges, 3 square measure associated with the contradictions caused by multiple disciplines. Jamshidi created his remarks that system optimizations, reliable and strength among a group of subsystems to realize a standard system goal became the main focus of the many applications like region, automotive, military, environmental, and repair business. It's fascinating to realize activity from subsystems for associate optimized performance at a high system level. It's potential to elucidate, predict, and formulate the decision-making issues of advanced systems

2. LITERATURE SURVEY

Modular Neural Networks (MNNs) could be a quickly growing field in artificial Neural Networks (NNs) analysis.[8] This paper surveys the various motivations for making MNNs: biological, psychological, hardware, and machine. Then, the final stages of MNN style square measure printed and surveyed likewise, viz., task disintegration techniques, learning methods and multi-module decision-making ways. Benefits and downsides of the surveyed strategies square size known, secondary degreed an calculation with position to sensible possible is delivered. Finally, some general recommendations for future styles square measure conferred.

Option evaluation with standard Neural Networks investigates a statistic standard neural network (MNN) model to cost the European decision choices. The modules square measure supported time to maturity and moneyless of the choices. The choice value perform of interest is homogenized of degree one with reference to the underlying index value and therefore the strike value. Compared to associate degree array of constant quantity and statistic models, the MNN methodology systematically exerts superior out-of-sample evaluation performance. We tend to conclude that modularity improves the generalization properties of ordinary feed forward neural network choice evaluation models

State of the Art and Future Trends in Distributed Systems and present Computing summarizes trends in communication paradigms for distributed systems, mentions established and new computer code infrastructures for distributed systems (such as CORBA and Jini), and provides an summary of mobile code and mobile agent principles. It then explains the vision of present Computing and future networked sensible devices together with techniques like RFID tags (or "smart labels"). It conjointly discusses problems in spontaneous networking, service discovery, and connected ideas. The report quotes munificently from variety of on-line resources found on the net.

Pattern classification has been with success applied in several drawback domains, like biometric recognition, document classification or diagnosing. Missing or unknown knowledge square measure a typical disadvantage that pattern recognition techniques have to be compelled to agitate once determination real-life classification tasks. Machine learning approaches and strategies foreign from applied math learning theory are most intensively studied and employed in this subject. The aim of this work is to research the missing knowledge drawback in pattern classification tasks, and to summarize and compare a number of the well-known strategies used for handling missing values.

[8][1]We propose a unique 2-stage soft computing approach for knowledge imputation, involving native learning and world approximation in bicycle, whereas within the literature just one of them is employed. In stage 1, K-means formula is employed

to switch the missing values with cluster center. Stage two refines the resultant approximate values victimization multilayer perception (MLP). MLP is trained on the whole knowledge with the attribute having missing values because the target variable one at a time. The hybrid is tested on two benchmark issues every in classification and regression victimization 10-fold cross validation. Altogether datasets, some values, that square measure indiscriminately removed, square measure treated as missing values. The particular and therefore the foretold values obtained square measure compared by victimization Mean Absolute share Error (MAPE). we tend to observe that, the MAPE price is reduced from stage one to stage two, indicating the hybrid approach resulted in higher imputation compared to stage one alone

[6]Ensembles of learning machines represent one among the most current directions in machine learning analysis, and are applied to a large vary of real issues.[10] Despite of the absence of associate degree united theory on ensembles, there square measure several theoretical reasons for combining multiple learners, associate degreed an empirical proof of the electiveness of this approach.[8] During this paper we tend to gift a short summary of ensemble strategies, explaining the most reasons why they're able to outmatch any single classifier at intervals the ensemble, and proposing taxonomy supported the most ways that base classifiers may be generated or combined along.

the Analytic Hierarchy method is associate degree introduce particle to the Analytic Hierarchy method[2][1] A multi criteria deciding approach within which factors square measure organized during a class-conscious structure. The principles and therefore the philosophy of the idea square measure summarized giving general background informant particle of the sort of menstruation utilized, its properties and applications.

One of the crucial issues in {decision making |deciding |higher cognitive method} process is to assess the relative importance or weights of various attributes. [3]This paper presents associate degree objective methodology to work out relative criteria weights that relies on co relational analysis particularly principal elements analysis. Consistent with the methodology, delineate during this paper, there's a clear stage to work out weight coefficients from Kaiser-Meyer-Olkin live and from square issue loadings.

While gauging the performances of in operation entities victimization inexact data on the input and output importance weights, associate degree entity is taken into account Farrell economical as long because it outperforms its peers for a minimum of one possible combination of the weights for inputs and outputs. [4]This paper argues that Farrell potency computations square measure supported associate degree optimistic perspective and a Farrell-efficient entity might perform rather poorly once weights similar to realistic issues square measure allotted to inputs and outputs.[7] Associate degree entity is outlined as sturdy economical if its relative potency score reaches one altogether possible combos of the input and output weights. A applied mathematics based mostly approach is projected to perform what's spoken as sturdy potency analysis to spot sturdy economical entities. In distinction to Farrell potency analysis, sturdy potency associate degree analysis involves the computation of all-time low potency score which will be allotted to an entity relative to the best score among all the entities wherever the same combination of weights for inputs and outputs is applied. the assembly risk set underlying the projected approach is additionally outlined and understood. Associate degree experimental study illustrates that compared with Farrell potency analysis sturdy potency analysis has scammer discrimination capability and therefore the entity it identifies as economical has superior average performance.

[9]Business intelligence (BI) is that the method of gathering enough of the proper data within the right manner at the proper time, and delivering the proper results to the proper individuals for decision-making functions in order that it will still yield real business advantages, or has a positive impact on business strategy, tactics, and operations within the enterprises.[5] This paper was supposed as a brief introduction to the study of business intelligence in enterprise computing atmosphere. Additionally, the conclusions illustrate the challenges to broad and deep readying of business intelligence systems, and supply the proposals of constructing business intelligence more practical.

3. PLANNED SYSTEM

The formulation of the decision-making method of advanced systems has been mentioned and therefore the focus is directed on the way to assess relative importance once multiple criterion square measures is concerned. The relative importance of one type criterion over another is bestowed by weights. It's been found that the determination of weights involves associate analysis method, that shouldn't be merely treated as associate improvement. a scientific methodology has been planned to see the weights for multiple objectives befittingly. The discussion on consistency at the tip reveals that the weights don't seem to be essentially consistent once variable weights square measure applied and synthesized within the method of determination.

4. MULTI-OBJECTIVE DECISION MAKING

Multi-Objective Decision Making is a technique for ensuring the integrity of data in outsourcing storage service. It allows a client that has stored data at an un-trusted server to verify that the server possesses the original data without regaining it. The model gives out probabilistic evidences of possession by sampling random sets of blocks from the server, which radically cuts I/O expenses. The customer preserves a continuous amount of metadata to check the proof. The test/reaction protocol conveys a small, continuous amount of data, which minimizes network communication. Thus, the model for remote data checking supports large data sets in widely-distributed storage systems. We are quite aware about square measure of single objective call issues from our previous study of applied mathematics and alternative improvement fields like inventory management, and project management. However, several call things incorporate deciding once over one objective has to be thought of (called multi-objective call making). Such associate approach is typical to investment issues wherever business banks have to be compelled to balance come and risk.

5. MACHINE LEARNING

Machine learning deals with the problem of the way to build pc programs that improve their performance at some tasks through expertise. Machine learning algorithms have evidenced to be of nice sensible worth in an exceedingly sort of application domains. Not amazingly, the sphere of software system engineering seems to be a fertile ground wherever several software system development and maintenance tasks can be

developed as learning issues and approached in terms of learning algorithms. This book deals with the topic of machine learning applications in software system engineering. It provides a summary of machine learning, summarizes the state-of-the-practice, provides a classification of the present work, and offers some application pointers.

6. MODULAR NEURAL NETWORKS (MNN)

Modular neural networks, as combined structures, have conjointly a biological background: Natural neural systems square measure composed of a hierarchy of networks designed of parts specialized for various tasks. In general, combined networks square measure a lot of powerful than flat unstructured ones.

In this manner we tend to attain the conception of standard neural networks. many general problems have crystal rectifier to the event of standard systems.

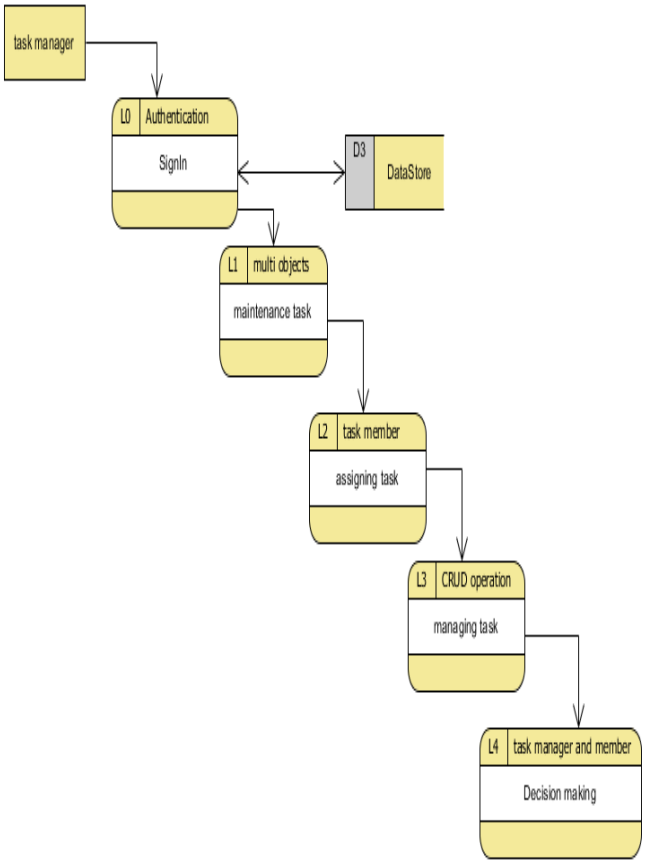


Figure.1 data flow diagram

•□Incorporating knowledge: Complete modules square measure associate degree extension of the approach mentioned in Sect. 10.3.5 of learning with hints.

•□Data merging and prevision averaging: Committees of networks may be thought-about as composite systems fabricated from similar parts.

•□Combination of techniques: quite one methodology or category of network may be used as building block.

•□Learning totally different tasks simultaneously: Trained modules are also shared among systems designed for various tasks.

•□Robustness and instrumentality: The combined network will grow step by step and may be created fault-tolerant.

7. CONCLUSION

The formulation of the decision-making method of complicated systems has been mentioned and therefore the focus is directed on the way to valuate relative importance once multiple criterion square measures concerned. The relative importance of 1 style criterion over another is conferred by weights. it's been found that the determination of weights involves associate degree analysis method, that shouldn't be merely treated as associate degree optimization. a scientific methodology has been projected to work out the weights for multiple objectives befittingly. The discussion on consistency at the top reveals that the weights don't seem to be essentially consistent once variable weights square measure applied and synthesized within the method of determination. The analysis of weights for multi-objective deciding could be a crucial issue in several applications like MODM and MFML. solely the load determination and consistency square measure centered within the conferred work

8. ACKNOWLEDGMENTS

The author would like to thank the Vice Chancellor, Dean-Engineering, Director, Secretary, Correspondent, HOD of Computer Science & Engineering, Dr. K.P. Kaliyamurthie, Bharath University,Chennai for their motivation and constant encouragement. The author would like to specially thank **Dr. A. Kumaravel, Dean, School of Computing,** Bharath University for his guidance and for critical review of this manuscript and for his valuable input and fruitful discussions in completing the work and the Faculty Members of Department of Computer Science &Engineering. Also, he takes privilege in extending gratitude to his parents and family members who rendered their support throughout this Research work.

9. REFERENCES

[1] G. Valentini and F. Masulli, "Ensembles of learning machines," in *Proc. WIRN VIETRI*, M. Marinaro and R. Tagliaferri, Eds., 2002.

[2] C. Wang, "Advances in information integration infrastructures supporting multidisciplinary design optimisation," *Enterprise Inf. Syst.*, Aug. 2012

[3] Y. M. Wang and J. K. Zhang, "A method based on standard and mean deviations for determining the weight coefficients of multiple attributes and its applications," *Appl. Stat. Manage.*, 2003

[4] P. Wang, S. Feng, and R. Zhang, "Method integration and variable weight synthesis in solving hybrid objective systems," Sep. 2008.

[5] X. Wang and X. Xu, "DIMP: An interoperable solution for software integration and product data exchange," *Enterprise Inf. Syst.*, Aug. 2012

[6] Q. Liu and C. Wang, "Multi-terminal pipe routing by Steiner minimal tree and particle swarm optimization," Aug. 2012

[7] W. Q. Liu, "The ordinary variable weight principle and multiobjective decision making," Mar. 2000.

[8] X. B. Lam, Y. S. Kim, A. D. Hoang, and C. W. Park, "Coupled aerostructural design optimization using the Kriging model and integrated multiobjective optimization algorithm," *J. Optim.*, Sep. 2009.

[9] L. Li and J. Liu, "An efficient and flexible web-services-based multidisciplinary design optimization framework for complex engineering systems," Aug. 2012.

[10] Y. J. Pang and K. D. Liu, "The consistency check in the analytic hierarchy process is not necessary condition of sequencing," *J. Hebei Inst. Arch. Sci. Technol*, 2002.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 291 - 295, 2015, ISSN:- 2319–8656

Android Application to Predict and Suggest Measures for Diabetes Using DM Techniques

V.Krishna Priya
Rajalakshmi Engineering College
8/14 F Kumar Brindavan Flats
43rd street nanganallur, ch-61
Chennai, India

A.Monika
Rajalakshmi Engineering College
2, Flowers road 4th lane,
Pursaiwakkam,ch-84
Chennai, India

P.Kavitha
Rajalakshmi Engineering College
819, Rajiv Gandhi St.,
Nazarathpet, Poonamalle.
Chennai, India

Abstract: Data Mining is an analytic process designed to explore data in search of consistent patterns and systematic relationships between variables, and then to validate the results by applying the patterns found to a new subset of data. Data mining is often described as the process of discovering patterns, correlations, trends or relationships by searching through a large amount of data stored in repositories, databases, and data warehouses. Diabetes, often referred to by doctors as diabetes mellitus, describes a group of metabolic diseases in which the person has high blood [3] glucose (blood sugar), either because insulin production is insufficient, or because the body's cells do not respond properly to insulin, or both. This project helps in identifying whether a person has diabetes or not, if predicted diabetic[4] the project suggest measures for maintaining normal health and if not diabetic it predicts the risk of getting diabetic. In this project Classification algorithm was used to classify the Pima Indian diabetes dataset. Results have been obtained using Android Application.

Keywords: Android Application, Diabetes, Data Mining

1. INTRODUCTION

Data mining, the extraction of hidden predictive information from large databases[1,12], is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Diabetes has become a most common disease in today's world. So for every individual it is important to take a precautionary measure to check if the person has any chances of getting diabetes. For this purpose we use data mining techniques to predict if a person is diabetic or not. It is attractive as the results are obtained through an android application installed in mobile device. The main reason for accuracy of results is that only most significant attributes causing diabetes are considered for analysis

Data mining tools [8] predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions [7]. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Diabetes (diabetes mellitus)[24] is classed as a metabolism disorder. Metabolism refers to the way our bodies use digested food for energy and growth.. Most of what we eat is broken down into glucose. Glucose is a form of sugar in the blood - it is the principal source of fuel for our bodies.

A person with diabetes has a condition in which the quantity of glucose in the blood [19] is too elevated (hyperglycemia). This is because the body either does not produce enough insulin [5], produces no insulin, or has cells that do not respond properly to the insulin the pancreas produces. This results in too much glucose building up in the blood. This excess blood glucose eventually passes out of the body in urine [19]. So, even though the blood has plenty of glucose, the cells are not getting it for their essential energy and growth requirements

2. DATASET

Dataset is composed of 768 instances. Each patient is characterized in data set by 8 attributes. All attributes are numerical values. This attributes are: Diastolic blood pressure, plasma glucose concentration a 2 hours in an oral glucose tolerance test, Diastolic blood pressure (mm Hg), triceps skin fold thickness (mm), 2-Hour serum insulin (mu U/ml), body mass index (weight in kg/(height in m)^2), diabetes pedigree function ,age (years), Class variable (0 or 1)[21].

3. CLASSIFICATION ALGORITHM

Classification [22] consists of predicting a certain outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute. The algorithm [16] tries to discover relationships between the attributes that would make it possible to predict the outcome. Decision tree [3] builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes [6]. A decision node has two or more branches. Leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called **root node**. Decision trees [3] can handle both categorical and numerical data.

3.1 C4.5 algorithm

The C4.5[18] algorithm constructs the decision tree with a divide and conquer strategy.
In C4.5algorithm, each node in the tree is associated with a set of cases. And also each cases are assigned with weights to take into account the unknown attributes values [17].

At the beginning, only the root is present, associated with the whole training set T S and with all case weights equal to 1:0. At each node the following divide and conquer method, the algorithm is executed, trying to find the locally best choice, with no backtracking allowed.

Let T be the set of cases associated at the node. The weighted frequency f req(Ci ; T) is computed (step (1)) of cases in T whose class is Ci , for i 2 [1; NC lass]. If all cases (step (2)) in T belong to a same class Cj (or the number of cases in T is less than a certain value) then the node is a leaf, with associated class Cj (resp., the most frequent class). The classification [22] error of the leaf is the weighted sum of the cases in T whose class is not Cj (resp., the most frequent class). If T contains cases belonging to two or more classes (step (3)), then the information gain of each attribute is calculated. For discrete attributes, the information gain is relative to the splitting of cases in T into sets with distinct attribute values. For continuous attributes, the information gain is relative to the splitting of T into two subsets, namely cases with attribute value not greater than and cases with attribute value greater than a certain local threshold, that is determined during information gain calculation [3].

The attribute [17] with the highest information gain (step (4)) is selected for the test at the node. Moreover, in case a continuous attribute is selected, the threshold is computed (step (5)) as the greatest value of the whole training set that is below the local threshold.

A decision node has s children if T1;Ts are the sets of the splitting produced by the test on the selected attribute (step (6)). Obviously, s = 2 when the selected attribute is continuous, and s = h for discrete attributes with h known values.

For i = [1; s], if Ti is empty, (step (7)) the child node is directly set to be a leaf, with associated class the most frequent class at the parent node and classification[22] error 0.

If Ti is not empty, the divide and conquer approach consists of recursively applying the same operations (step (8)) on the set consisting of Ti plus those cases in T with unknown value of the selected attribute. Note that cases with unknown value of the selected attribute are replicated in each child with their weights proportional to the proportion of cases in Ti over cases in T with known value of the selected attribute.

Finally, the classification error (step (9)) of the node is calculated as the sum of the errors of the child nodes. If the result is greater than the error of classifying all cases in T as belonging to the most frequent class in T , then the node is set to be a leaf, and all sub-trees are removed[3].

4. USAGE OF SIGNIFICANT ATTRIBUTES

The attributes used for calculation is the significant attributes done using Attribute Selection algorithm of WEKA[9] tool. The most significant attributes are plasma, body mass index, diabetes pedigree function, insulin level. These are the most significant attributes for the prediction of diabetes status of a person. The class attribute of the dataset specifies class 0 i.e not diabetic and class 1 i.e diabetic.

- **Not All Attributes Are Equal**

Whether you select and gather sample data yourself or whether it is provided to you by domain experts, the selection of attributes is critically important. It is important because it can mean the difference between successfully and meaningfully modeling the problem and not.

- **Misleading**

Including redundant attributes can be misleading to modeling algorithms. Instance-based methods such as k-nearest neighbor use small neighborhoods in the attribute space to determine classification and regression predictions. These predictions can be greatly skewed by redundant attributes.

- **Overfitting**

Keeping irrelevant attributes in your dataset can result in overfitting. Decision tree algorithms like C4.5 seek to make optimal spits in attribute values. Those attributes that are more correlated with the prediction are split on first. Deeper in the tree less relevant and irrelevant attributes are used to make prediction decisions that may only be beneficial by chance in the training dataset. This overfitting of the training data can negatively affect the modeling power of the method and cripple the predictive accuracy.
It is important to remove redundant and irrelevant attributes from your dataset before evaluating algorithms. This task should be tackled in the Prepare Data step of the applied machine learning process.

- **Feature Selection**

Feature Selection or attribute selection is a process by which you automatically search for the best subset of attributes in your dataset. The notion of "best" is relative to the problem you are trying to solve, but typically means highest accuracy.
A useful way to think about the problem of selecting attributes is a state-space search. The search space is discrete and consists of all possible combinations of attributes you could choose from the dataset. The objective is to navigate through the search space and locate the best or a good enough combination that improves performance over selecting all attributes.
Three key benefits of performing feature selection on your data are:
Reduces Overfitting: Less redundant data means less opportunity to make decisions based on noise.
Improves Accuracy: Less misleading data means modeling accuracy improves.
Reduces Training Time: Less data means that algorithms train faster

5. SYSTEM ARCHITECTURE

The system architecture (see Figure 1) describes the flow of the project work. The first step in the process is the collection of data needed for the work. Here the dataset used is Pima Indian diabetes dataset[21], which is collected in the first step. The next step in the process is preprocessing of the data[4]. Here we covert the raw data into understandable format. Now the preprocessed data is classified into a decision tree[3] to predict the status of a person whether diabetic or not using the algorithm(C4.5)[25]. The user enters the details to know his results for the test into an android app[22] installed in his

mobile device. The attributes entered by the user is compared with the decision tree and the results are generated.

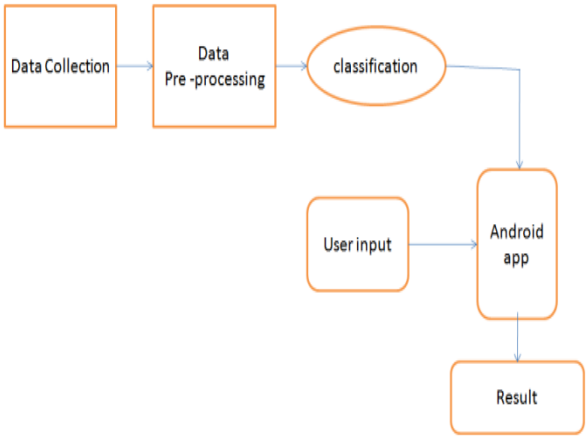


Figure 1. System Architecture

6. SCREENSHOTS

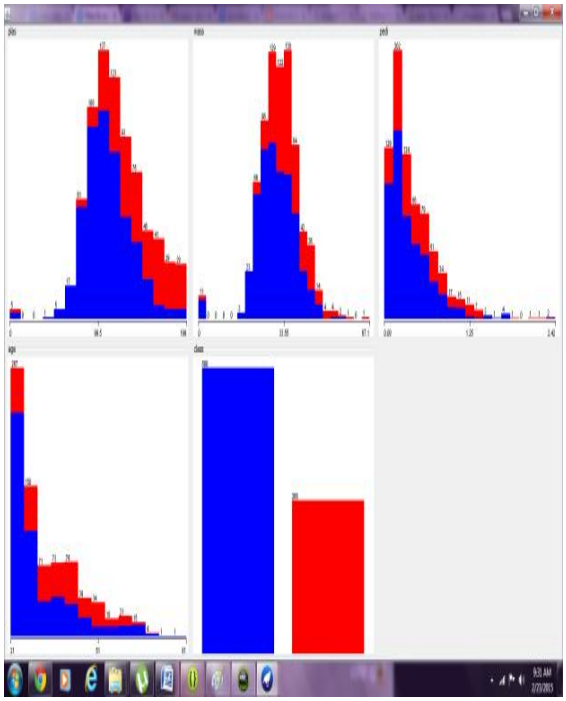


Figure 2. Pre-Processing

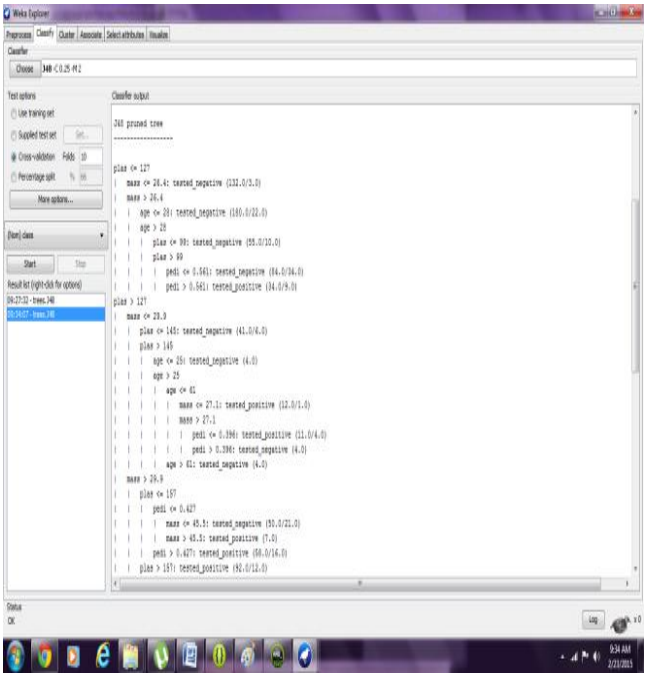


Figure 3. Decision tree

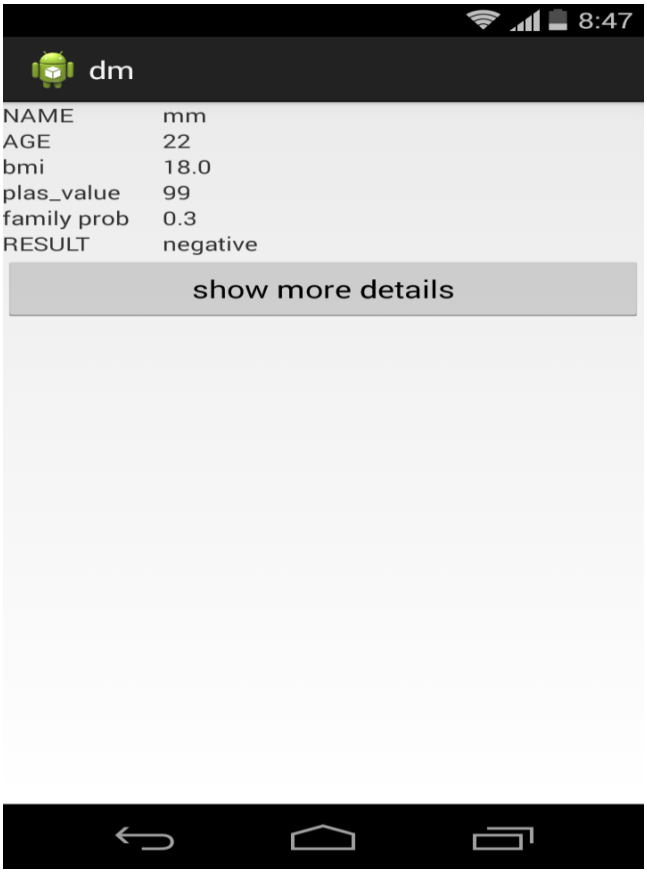


Figure 4. Result after entering details in Android application

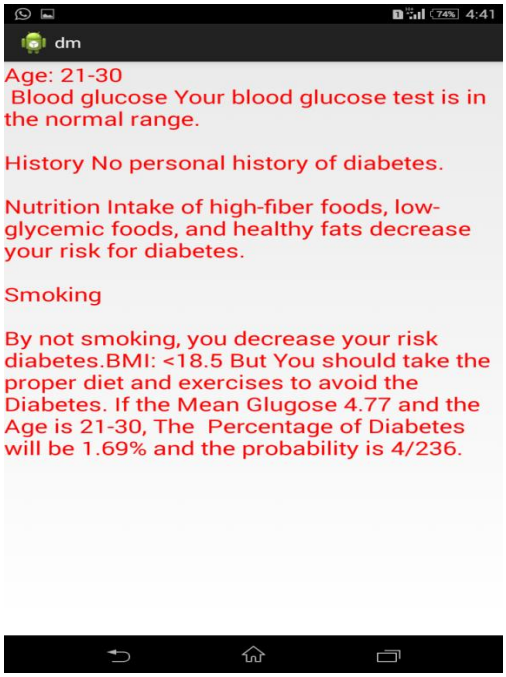


Figure 5. Suggestive measures for user

7. CONCLUSION

The discovery of knowledge[9] from datasets is important in order to make effective diagnosis. The aim of data mining is to extract information stored in dataset and generate clear and understandable patterns. This study aims at the discovery of a decision tree model for the prediction of diabetes. Pre-processing is used to improve the quality of data. While preprocessing[6], the significant attributes of the dataset are considered for prediction of diabetes. This is an important factor for consideration. The decision tree algorithm[14] used for classification also produces maximum accuracy when compared to other algorithms of classification. Finally the results of the system are obtained in an android application which is very useful for the present generation.

8. FUTURE SCOPE

In future this system can designed for any prediction of any other disease such as cancer, thyroid, lung diseases etc., if these an android application of such disease prediction would be of great use in the near future. Another future enhancement would be to reduce the no of attributes considered for the prediction purpose. Considering less no of attributes and produce more accurate results is needed as an enhancement for the existing system.

9. ACKNOWLEDGMENT

Sincere thanks to our guide Mrs.P.Revathy who contributed to the development of this paper

10. REFERENCES

[1] P. Yasodha, M. Kannan, "Analysis of a Population of Diabetic Patient Databases in Weka Tool", International Journal of Scientific & Engineering Research Volume 2, Issue 5, May-2011

[2] WEKA, by university of Waikato, http://www.cs.waikato.ac.nz/ml/weka/

[3] T. Mitchell, "Decision Tree Learning", in T.Mitchell, Machine Learning (1997) the McGraw- Hill Companies, Inc., pp.

[4] Han, J., Kamber, M.: Data Mining; Concepts and Techniques, Morgan Kaufmann Publishers (2000).

[5] Gloria L.A. Beckles and Patricia E. Thompson-Reidy the authors of" Diabetes and Women's Health Across the Life Stages".

[6] Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining Concepts and Techniques" Third edition .

[7] Folorunso O and Ogunde A. O (2004), "Data Mining as a Technique for Knowledge

[8] Management in Business Process Redesign" The Electronic Journal of Knowledge Management Volume 2 Issue 1, pp 33-44

[9] P.Yashoda, M.Kanan, Analysis of a population of diabetic patients databases in WEKA tool, IJSER, vol2, issue5, may 2011.

[10] Mukesh kumari, Dr. Rajan Vohra ,Anshul arora Prediction of Diabetes Using Bayesian Network (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5174-5178

[11] M. Khajehei, F. Etemady, "Data Mining and Medical Research Studies," cimsim, pp.119-122, 2010 Second International Conference on Computational Intelligence, Modelling and Simulation, 2010

[12] Kaur H, Wasan SK," Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science,2(2):194-200,2006

[13] Analysis of a Population of Diabetic Patients Databases with Classifiers using c4.5 Algorithm" World Academy of Science, Engineering and Technology International Journal of Medical, Pharmaceutical Science and Engineering Vol: 7 No: 8, 2013.

[14] Margaret H. Dunham,-"Data Mining Techniques and tice hall publishers

[15] P. Radha , Dr. B. Srinivasan Predicting Diabetes by cosequencing the various Data Mining Classification Techniques IJISET - InternationalJournal of Innovative Science

[16] E.Knorr.E and R.Ng, "Algorithms forming distance - based outliers in large datasets", in proceedings of 1998 International Conference on Very Large Data Bases (VldB'98), pp. 392-403 New York, 1998.

[17] E.Jiawei Hen and Micheline Kamber "DataMining Concepts and Techniques", *CA:Elsevier Inc,SanFranciso*, 2006

[18] U.M.Piatetsky-Shapiro and G.Smyth "From DataMining to Knowledge Discovery : An Overview",1996, pp.1 -36

[19] S.C.Liao & M.Embrenchts, "Data Mining techniques applied to medical information", *Med.Inform*, 2000, pp.81 102.

[20] L.Breiman, J.Friedman, J.Olsen C.Stone, "Classification and Re-gression Trees", *Chapman & Hal, 1984, 122-134. Engineering & Technology, Vol. 1 Issue6, August 2014.*

[21] *https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes UCI MACHINE LEARNING REPOSITORY*

[22] Szakacs-Simon, P. Dept. of Autom., "Transilvania" Univ., Brasov, Romania Moraru, S.A. ; Perniu, L.Android application developed to extend health monitoring device range and real-time patient tracking International Journal of Advanced Research in Computer Science and Software Engineering
[23] Rohanizadeh.s "A proposed data mining methodogy application to industrial procedures"
[24] en.wikipedia.org/wiki/Diabetes_mellitus

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 296 - 299, 2015, ISSN:- 2319–8656

Identity Recognition Using Edge Suppression Method

R.Ram kumar
Bharath University
Chennai, India

V.Sanjeevi
Bharath University
Chennai, India

K.Sivaraman
Bharath University
Chennai, India

Abstract: There are multiple crimes happening even in the presence of cctv cameras still we couldn't able to find the identity due to shadows in the image. In this work, we have used edge suppression method and affine transformation to recognize identity even under severe shadow. We also used gradient field to calculate the position of light sources in the place. For classification of identity from the databases we use k-nearest neighbor rule. Additionally we also were using this principal component analysis for feature extraction. Classification can be done in real time environments by accessing authorized databases or also by standard databases.
Keywords: K-NN Classifier, Tensors, flash image, shadow removing, Feature extraction

1. INTRODUCTION

Face recognition is one of the applications for automatically identifying and recognizing the face of an human by an computer. This kind of image processing applications are used in many ways, in this work face recognition system is used not only recognizing face also taken upto his chest which increases multiple reference points such as neck, shoulder and more. The another important issue is shadow presence in the image due to this there is a chance of error occurrence in finding the correct person. So, to reduce the impact of shadow on the image diagonally projecting sensors[2] are used this suppresses the edges of the image. Here we also were accounting the chromaticity[8] of the environment lighting. Ramesh Raskar[1] had mentioned that the chromaticity of the environment lightning will be approximately as same as the chromaticity of the diffused light. And it is found that for different magnitude in the chromaticity of differently exposed regions has different rgb, hue and hsv color values[5].We proposing that along with face recognition recognizing of other identities also important because the identity database like aadhar card are capturing images up to chest level not only up to face. So it will be useful when using it with real databases. As compared with thresholding we clear the textures in the scene by approaching principled way. Here we are not accounting any assumptions on the environment lightning, mapping of exposure or reflection. The diagonally projecting tensors helps us to remove those shadows by using an highly exposed image as an reference called as flash image.

2. EXISTING METHODOLOGY:

All images are a combination of primary colors like red, blue, green. All the image processing process uses RGB images which has more number of color components and it makes the processing slower. In recovering the illumination map, we make the usual assumption that the scene texture edges do not coincide with the illumination edges. But the foreground layer, edges of the foreground object which exactly align with the background edges cannot be recovered. And the existing systems has face recognition only, which may fail in case of twins and people having similar faces and also in usage of mask.

.

3. PROPOSED METHODOLOGY:

To predict the presence of light source the two pictures are converted into YUV format which reduces the color components for processing it into a gradient pictures with x and y-axis components. That helps in finding the position of light sources which in turn helps in removing shadows. Then those diagonally projecting tensors are applied on the pictures to analyze the vector and scalar values of two pictures(real image and flash image) then affine transformation uses those values to correct the pictures like rotating and aligning. Thus a shadow free image is obtained as a result. This resulted picture is used to search in databases. Here we are proposing a id recognition instead of regular face recognition which uses the principal component analysis to feature extract the face, neck, shoulders from the images. By using the calculated eigen vectors and values KNN classifier produce the result by comparison.

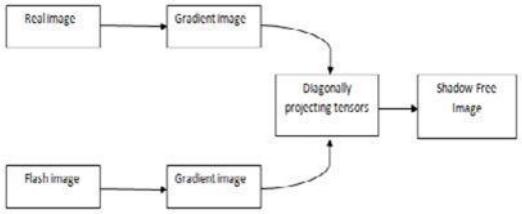


Fig.1 Block diagram

4. GRADIENT COMPUTATIONS:

This computations are existing mathematical functions only. This is used to obtain the gradient value in the image environment. The gradient magnitude and direction are encoded by computing. By [9]adding partial derivatives of X and Y directions gradient vector is formed.

$$\Delta I = (\partial I/\partial x, \partial I/\partial y) \quad \ldots(1)$$

we could write this continuous function, I(x, y) as:

$(\partial I(x,y)/\partial x) = \lim \Delta x \to 0\ (I(x+\Delta x,y)-I(x,y))/\Delta x$ …(2)

While if it is an discrete case, only one pixel intervals can be taken. Hence we can take the difference between pixel before or one pixel after the I(x,y).By the usage of correlation we can define the pixels after and before I(x,y) symmetrically, and compute:

$(\partial I(x,y)/\partial x) = (I(x+1,y) - I(x-1,y))/2$ ...(3)

5. ARCHITECTURE DIAGRAM:

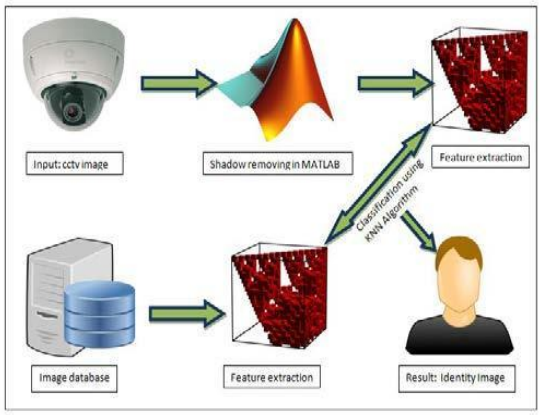


Fig.2 Architecture Diagram

6. AFFINE TRANSFORMATION ON GRADIENT:

Let intensity image and denote each pixels gradient vector at I. The Gσ smoothed structure tensor is defined as

$$\mathbf{G}_\sigma = (\nabla I \nabla I^T) * K_\sigma = \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix} * K_\sigma$$

where $*$ denotes convolution and Kσ is a normalized 2D Gaussian kernel of variance σ. The matrix Gσ can be de-composed as

$$\mathbf{G}_\sigma = V \Sigma V^T = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{bmatrix},$$

here the eigen-vectors v1,v2 are corresponding to the Eigen-values λ1, λ2 respectively and λ2 ≤ λ1. The eigen-values and eigen-vectors of Gσ give information about the local intensity structures in the image[8]. For homogeneous regions, λ1 = λ2= 0. If λ2= 0 and λ1 > 0, it signifies the presence of an intensity edge. The eigen-vector v1 (corresponding to the higher eigen-value λ1) corresponds to the direction of the edge.

7. TENSORS:

$$D^{self} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{bmatrix}$$

$$D^{self}\mathbf{v}_1 = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{bmatrix} \mathbf{v}_1$$

$$= \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Here we have used tensors to find out the difference between the scalars and vectors of real image and flash image. These tensors are idely used mathematical component to derivethe scalar and vectors, actually it'san generalization of scalars and vectors; a zero rank tensor is a scalar,and a first ranktensor is a vector and a tensor also include linear map, dotproduct and cross product.

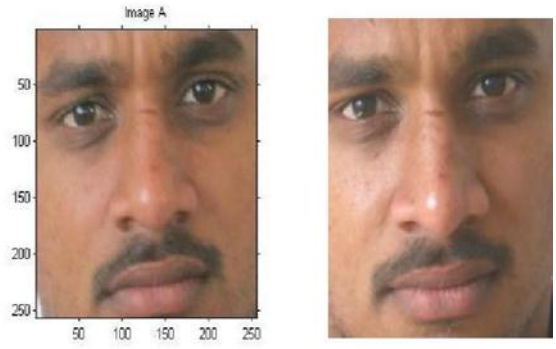


Fig.3 (i) Obtained image (ii) Shadow free image

8. SELF PROJECTION TENSORS:

We removed edges from a single image by using this already proven diagonally projecting tensors from the image itself. The idea is to project the image gradient vector onto its own orthogonal direction and hence the name self-projection tensors. This analysis will lead us to our main idea of cross projection tensors is to estimate these tensors from a second image and apply them to suppress edges in the given image. As Amit Agarwal [7] proposed the technique of gradient projection to remove artifacts from flash image using a non-flash real image. They project the flash image gradient onto the direction of the ambient image gradient to remove spurious edges from flash image due to glass reflections. They use the idea that the direction of the image gradient remains stable under illumination changes. We first show that taking a projection can also be defined by an affine transformation of the gradient field. The Eigen-vector v1 of the structure tensor matrix G correspond to the Direction of the edge. Suppose by an affine transformation of the gradient field. The Eigen Vector v1 of the structure tensor matrix G correspond to the direction of the edge. Suppose we define the self-projection tensor D^{self} as $u_1=v_1$ u $_2=v_2$,$\mu_1=0$ $\mu_2=1$. It is easy to see that an affine transformation of the image gradient using D^{self} will remove the local edge

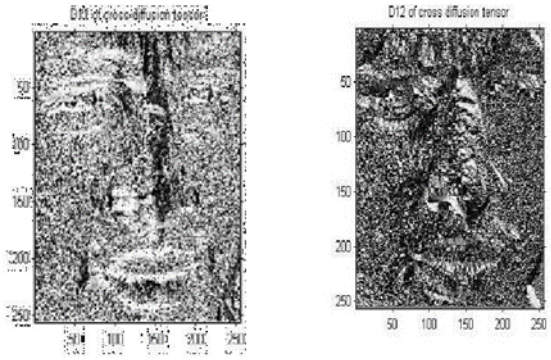


Fig.4 YUV Image from RGB image

9.DIAGONALLY PROJECTING TENSOR:

To remove the scene texture edges from an image by transforming its gradient field using diagonally projecting tensors obtained from a second image of the same scene. The final image is obtained by a 2D integration. If A is also homogeneous (λA1 = 0), set μ 1 =μ2 = 0. These results in If A is also homogeneous (λA1
= 0), set μ1 =μ2 = 0. These results in

$$D(x,y) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

If A is not homogeneous (λA1 > 0), set μ1 = μ2 = 1

$$D(x,y) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

This result in and edges which are in A but not in B can be retained. Else, if there is an edge in B (λB1 > 0),
remove that edge by setting μ1 = 0, μ2 =1.

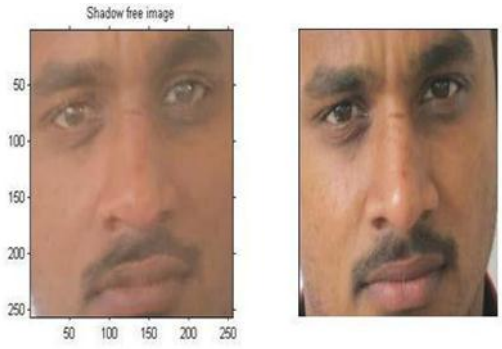


Fig.5(i) shadow free image (ii)illuminated image

10.FEATURE EXTRACTION:

Feature extraction is one of the image processing processes which can be used to extract the information from the image. This information must be valuable to the later step of identifying the subject with an acceptable error rate. This process is proved to be efficient in computing time and usage of memory. Again here we extracting more details rather than the usual information extraction of faces. The optimized output is obtained for the classification.

11. PRINCIPLE COMPONENT ANALYSIS:

Again we are using the proven principle component analysis which is widely used in all the sectors to analysis the form from computer image processing to neural analysis due to its extraction of confusing datasets using non-parametric method which is an simpler one. And it is well known as most reliable results from applied linear algebra. The Reason for using PCA in this identity recognition is to express the 2-D facial image into the large 1-D vector of pixels in the compact principal components of the feature space. Sometimes it also helps us in revealing hidden, simplified structure by reducing the complex dataset into the lower dimensions.

12. CLASSIFICATION:

Classification is one of the data mining process of analyzing and finding the similar things to group under some keyword. Here we done the classification of the test images with the images in the present database. For this, we go for the reliable KNN classifier**.**

13. K-NN CLASSIFIER:

While looking for the algorithm to search the match in database K-NN classifier seems to be the proven algorithm especially in searching images, recognizing patterns and data mining. K-NN algorithm uses the similarity measures to classifies the matches in a database. A case is arranged by a lion's share vote of its neighbors, with the case being allocated to the class most normal amongst its K closest neighbors measured by a separation capacity. On the off chance that K = 1, then the case is essentially allotted to the class of its closest

14. RESULTS:

We cleared shadows from the Real-environment (no-flash) image A by using the image with flash. From the real image, flash image was created using the separate photo editor tool like lumia creative∇ studio. Using F and trans-form the gradient field A the cross projection tensor Df. the shadow removed image has less color artifacts and less illuminated map when compared with the previous highly textured one. Always there will be some difference in color tones due to the environment lightning. Our edge suppression method requires no color calibration or artifacts and when compared to the result using gradient projection. while converting the image into flash image there will be some variations in lighting and this doesn't provide any illumination mapping. The illumination map obtained by our approach better represents the diffuse ambient illumination. Even the white balance in real and flash images is different but Our resulted image has no color artifacts. We can proceed to classification process after getting the shadow free image. Image will be given as a query input and the matches are searched in the database. The output is shown in figure 6.

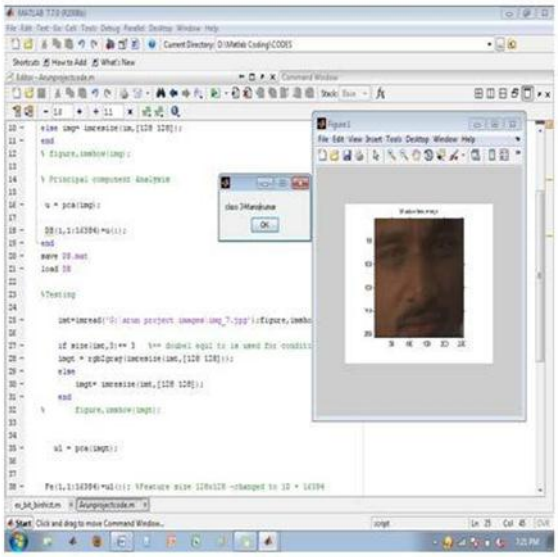


Fig.6 Output

15. CONCLUSION:

To remove shadow in a image, a tensor-based identity recognition method is proposed the illuminations is effectively reduced by edge suppression method. We presented an approach for edge-suppressing operations on an image. The light sources are found by applying the gradient field effect. In the recognition phase, Principal component analysis is used for feature extraction. The K-nearest-neighbor rule is applied for classification. Experiments are carried out upon standard databases and if access available can be done in real-time too. The results reveal that the proposed method achieves satisfactory recognition rates under varying illumination conditions. We hope that it can be able to use in verification and identity authentication processes etc because of its easy handling.

16. REFERENCE:

[1] Ramesh Raskar, S.Nayar and Y. Li. *Removing photography artifacts using gradient projection and flashexposure sampling.* ACM Trans. Graph., 24(3):828–835, 2005.

[2] Face Recognition under Severe Shadow Effects Using Gradient Field Transformation Parisa Beham M *, Bala Bhattachariar J.U, ISSN 2250-3153. International Journal of Multimedia and Ubiquitous Engineering

[3] ClopiNet, 955 Creston Rd., Berkeley, CA 94708, USA. isabelle@clopinet.com 2 IBM Research GmbH, Zurich Research Laboratory, Saumerstrasse 4, CH-8803 Ruschlikon, Switzerland. ¨ ael@zurich.ibm.com

[4]H. Barrow and J. Tenenbaum. Recovering intrinsic scene characteristics from images. In Computer Vision Systems, pages 3–26, 1978

[5] H.Chen, P.Belhumeur, and D. Jacobs. In search of illumination invariants. In Proc. Conf. Computer Vision and Pattern Recognition, pages 254–261, 2000.

[6] A.Elgammal, D.Harwood, and L. Davis. Non-parametric model for background subtraction. In Proc. European Conf. Computer Vision, pages 751–767, 2000.

[7] Edge Suppression by Gradient Field Transformation using Cross-Projection Tensors by Amit Agrawal, Ramesh Raskar and Rama Chellappa, 2008.

[8] Commentary paper on shadow removal in indoor scene by Sofke.M, 2008

[9] G.Aubert and P. Kornprobst. Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations, volume 147 of Applied Mathematical Sciences. SpringerVerlag, 2002

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 300 - 307, 2015, ISSN:- 2319–8656

Detecting And Resolving Privacy Conflicts in Online Social Networks using AC2P Protocol

J.S.Harilakshmanraj
Sri Ramakrishna Engineering College, Coimbatore
Tamilnadu, India

N.Rajkumar
Sri Ramakrishna Engineering College, Coimbatore
Tamilnadu, India

Abstract: Online Social Network (OSN) sites act as a medium to spread their own views, activities and their thoughts to some camaraderie. Contents of this network are spread over web, so it was hard to determine by a human decision. Currently, they do not provide any mechanism to ensure privacy concerns towards data associated with each user. Due to this problem, number of users lacks from their ownership control. In this paper, we proposed **AC2P** (Activity Control-Access Control Protocol) for information control on the web. Alternatively, Tag Refinement strategy determines illegal tagging over images and send notification about particular image spread within different communities/groups. These techniques reduce risk of information flow and avoid unwanted tagging toward images.

Keywords: Decision based access control, Tag refinement, Online social Networking.

1. INTRODUCTION

Online Social Networks act as medium to share both personal, public information and helps to formulate network using friends, colleagues, family and even with unknown persons. It has experienced tremendous growth in recent years.Facebook is one of most frequently used social networking site, which has more than 800 million active users and over 40 billion pieces of contents like web links, news, blog posts, photos are being shared each month[2].To protect user data access control mechanism has become much needed one,[4],[5].

At present, OSN provide *user- wall* to every user, where user and their friends can post both views and content using those walls. Subsequently, users upload both content as well as *tag* other users, who appear in that content. Each tag acts an explicit reference which links to a user's space. To protect user data, OSNs require user system and policy administrator for regulating data in social network.

A simple access control mechanism allows users to govern access to information contained in their spaces, but has no control over the data, which has presented outside their spaces. For example, if a user posts a comment in a friend's space, she/he cannot specify, who should view the comment.

In another scenario,when user upload photo and tags friends who appear in the photo, he/she cannot state any privacy norms about the photo. In this paper, we propose a solution to sophisticate collaborative management of shared data in OSNs.Based on these sharing patterns, AC2P protocol is used to capture the core features of user authorization requirements that have not been accommodated, so far, By existing access control system for OSNs.(e.g.[6],[8],[9],[10]and[11]).

Accessing the implications of access control mechanisms traditionally rely on the security analysis techniques, (e.g. Operating system, [7], Trust management,[12] and role-based access control,[3] [13]).

2. BACKGROUND

At present, SNS (Social Networking Sites) allow merchants and third parties to take advantage of user information without their agreement. Some important privacy issues in SNSs are: [1],

- The privacy tool is very hard to learn and to use them, due to which people feed up and they end up doing nothing.

- User can directly control their profile information, but cannot control what others reveal about them.
- Privacy tools (or) options are desired to provide, for choosing "*Friends*", *"Friends of Friends*" (*FOF*) (or) "Everyone", but it is not yet simplified.[14].
- With the third party integration, it becomes more risky, that your information is being shared among various stakeholders,[2].

(i) Disclosing the user's identity

At present, SNSs motivates users to share profile images. So, there is a risk that propagates with technologies like Content based Image Retrieval (CBIR) by analyzing the specification of an image, which reveal details of place from where the image was taken. Most SNS users are able to share any images(or)videos regardless of who is in that specific content.So,there is a high risk of publishing user identity and location even sometimes without user knowledge.

(ii) Cyber Crime-Related Field

Some rule defines that, however vulnerable to cyber criminals who pretend themselves as a friend using fake names and gain access to all information shared by naive users.

Cutillo et al [16] state some SNS's should fulfill the following privacy requirements.

Basic Privacy Requirements: [16]

a) End-To-End Confidentiality- All communications are needed to be confidential and only the sender and receiver should have control of access to the data.

b) Privacy- Personal information of a user should not publish to any other users apart from these explicitly mentioned by the user.

c) Access Control- User should be able to manage and control over their profiles as well as attributes of their profiles.

d) Authentication- For satisfying the previous requirement of a receiver's message should be able to authenticate the sender of the message.

e) Data Integrity- For each swapped message, whether it is acknowledged or a request, original authentication and also modification detection are needed to be performed.

f) Availability- All Public data has to be accessible and all messages should be delivered in time.

3. PROBLEM DEFINITION

Major computer security aspects are: Confidentiality, Integrity and Availability. At many Social Networking Sites, have limited security protection. A developer concentrates to enhance communication between users; therefore no security threats are being identified, so far. This work identifies the threats in OSNs and finds a solution for both content sharing as well as for image tagging activities.

4. ACTIVITY CONTROL MECHANISM FOR OSNs

In this section we formalize a AC2P Protocol for OSNs (Section 4.a) as well as Decision Scheme (Section 4.b) and Decision evaluation mechanism (Section 4.c) for the specification and enforcement of privacy policies toward OSNs.

4.1 AC2P PROTOCOL MODEL

To determine the consequence of information sharing, users require good understanding towards visibility of information that to be probed. However, privacy controls in OSNs are complicated and unintuitive.

This protocol consists of three major components namely:

(i) SNS Server.
(ii) Evaluation Schema.
(iii) Host based Web Server and
(iv) Decision-Based System.

Using above components, privacy has been preserved in OSNs.

SNS Server: It Gets Request/Response from OSNs user. Appropriate request messages are transmitted to application servers and it gets a notification from the application server (Alert-MSG), based on the user decision, particular OSN content will be allowed for other user's visibility.

Evaluation Schema: It verifies the content and check whether it is unique (or) not. It acts as *Plagiarism checker* to validate the uniqueness.

Host based Web server: It acts as authentication measurement system. If user blocks visibility of content to particular users. Then level of privileges is being measured for each and every user. Measurements are done in forms of

(i) High-Risk
(ii) Intermediate-Risk
(iii) No-Risk.

Decision-based System: An OSNs user has direct control over to set privacy. This component offers two ways of decision making towards the content:

- Allow the content (Others Visibility)
- Refuse the content(DENY)

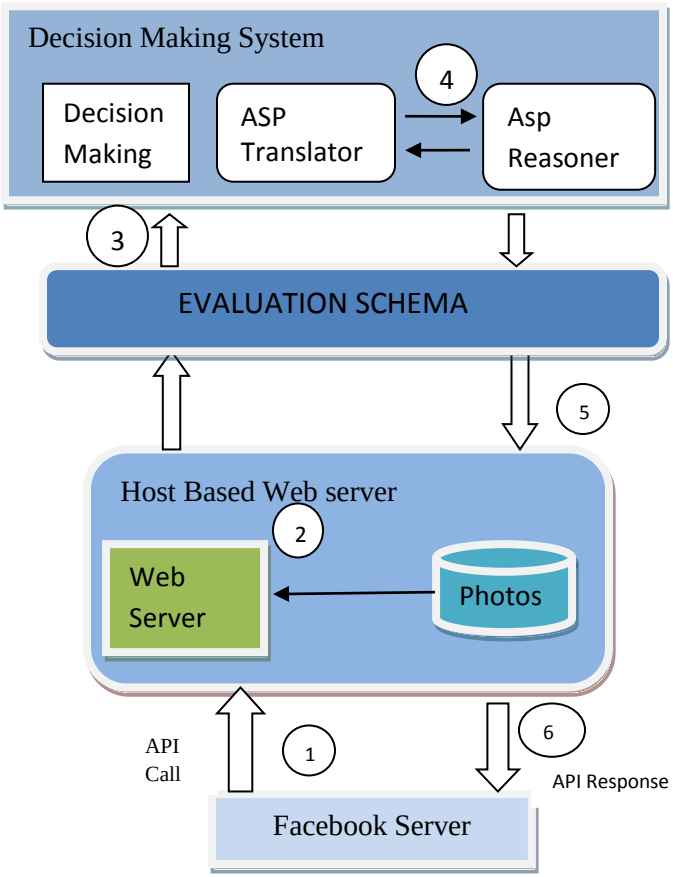


Fig.4.1 Framework of AC2P Protocol.

In fig 4.1 describe about components and strategies involved in AC2P Protocol mechanism, which used to protect OSNs contents from unauthorized users. It follows certain procedures, is being stated below:

1. *API call procedure*- it intimates host server whether to share the content or not.
2. *An Access request*- Notification is sent to the Evaluation mechanism (Section 4.3).
3. Examining the uniqueness of OSNs content is done.
4. After evaluation, *Final decision* is taken.
5. The Final decision will be either to "*Allow*" or "*Deny*" the OSNs content.
6. Using, the *API Response* final decision is notified to FB server.

4.2 EVALUATION SCHEMA

This schema used to predict the uniqueness of content in OSNs. For Example, if Bob post some content in public-view, john see that content as well as trim some information and post to his wall. In above scenario, some modification is been done towards the content. To predict that activity in OSNs, this schema used to predict the uniqueness value and send notification to particular user about their content was been access by other user, whether content should allowed or not. Alert notification is been generated.

4.3 *PROTOCOL -POLICY SPECIFICATION*

In Fig 4.2.,a disseminator used to share other profile information to others. So this kind of access specific schemas is being used widely [17][18].By using this kind of access privacy setting and access control norms will not be suitable for a privacy protection scenario. Some modification needs to be done.By using single controller, the resource-owner, to specify access control policies. A policy evaluation scheme is used to evaluate the DV (Decision-Value).A DV value state two possibilities either "Allow" or "Deny". This decision is taken based on some constraints.

$$\text{Decision} = \begin{cases} \text{Permit} & \text{if } DV\text{ag} > \text{Sc}. \\ \text{Deny} & \text{if } DV\text{ag} \leq \text{Sc} \end{cases} \quad \ldots..(1)$$

If the Sc is high, there is a chance of *Deny* access, take cares of high sensitive data.otherwise, the final verdict is most likely to *Allow* the data access.

(i) Owner-overrides: The owner decision is the final decision, It has highest priority. Based on the weight age of decision making scheme, we set $w_{ow}=1$, $w_{cb}=0$,then

$$\text{Decision} = \begin{cases} \text{Permit, if } DV\text{ag}=1 \\ \text{Deny, if } DV\text{ag}=0 \end{cases} \quad \ldots\ldots.(2)$$

(ii) Majority-Permit: The sending request is greater than the number of controller to deny, the final decision will be

$$\begin{cases} \text{Permit, if } DV\text{ag} \geq ½ \end{cases}$$

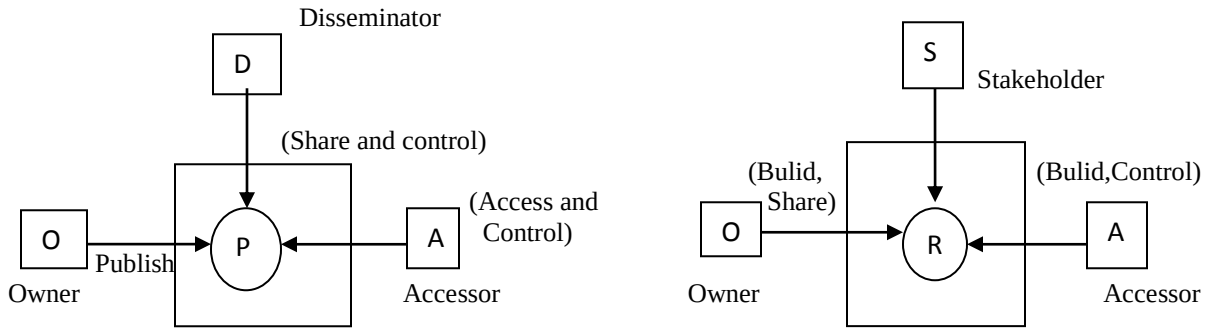
Decision= Deny, if $DV_{ag} < ½$ ….(3)

By using the above strategies, owner (U*i*) Will take the final decision to *Allow* (or) *Deny* the data object.

4.4 *DECISION EVALUATION MECHANISM*

To make an authorized decision from the user(U*i*) of content policy evaluation schema used for decision making purpose.It makes privacy setting based on certain norms desired by an owner(U*i*), In fig.4.3 illustrate overall scenario of access mechanism and its functionalities. Decision aggregate is being generalized and not been specialized under some constraints, but determines whether to *refuse* (or) to *allow* the content.

The probability flow model (*PFM*) used to predict,whether the user(U*j*) will get permission for content from the owner(U*i*).It is important to notify that OSNs community is defined for each and every user separately.Some sub-community use to define particular contexts.It is also possible to develop aggregations of OSN community in Social Neighborhood(SN) to form own communities. Using this model, some evaluation is being done. We used to collect some benchmark datasets from Facebook.Using, those datasets some evaluation are been done.



a) A Disseminator shares other profile

b) A user shares his/her relationships

FIG 4.2. Profile and Personal information Sharing

schema is being processed.Based on the hop value and community value(C*i*),the distribution size is being calculated.

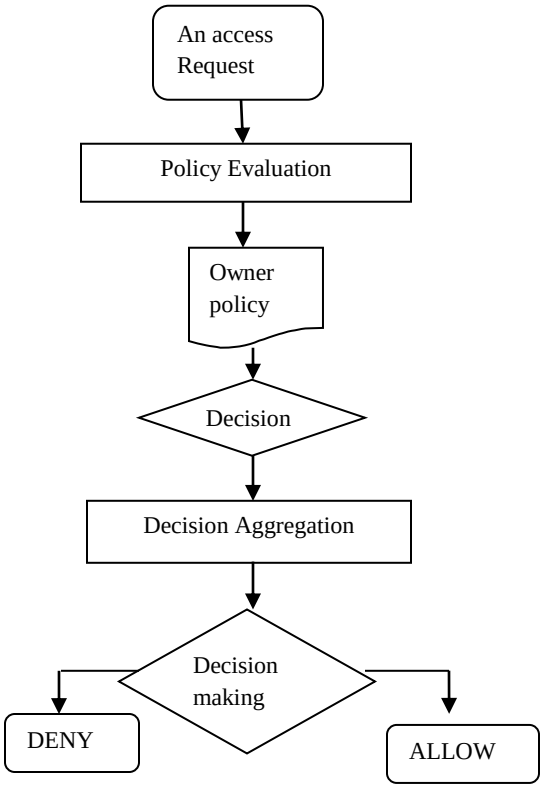


FIG.4.3 Schematic structure of Decision Evaluation Mechanism

4.4.1 A Voting Schema for Decision-Making

Various types of voting schemas are used for decision making [19].We proposes a voting scheme to achieve an effective user based conflict resolution in OSNs.Our voting schema consists two voting mechanisms.

(i) **Decision based voting.**

(ii) **Sensible voting.**

Decision based voting: A decision value (*DV*) is enhanced from the policy evaluation is defined as follows, where Evaluation *(P*) returns the decision of a Policy (*p).*

$$\text{Decision} = \begin{cases} 0 & \text{if Evaluation (p) =Deny} \\ 1 & \text{if Evaluation (p) =Permit} \end{cases} \quad \ldots (4)$$

Sensible voting: Each user assigns an SL to the shared data item to reflect her/his privacy concern. A sensitivity score (Sc) (in the range from 0.00 to 1.00) for the data item can be calculated based on the following equation:

$$S_c = (SL_{ow} + SL_{cb} + \sum_{i \in SS} SL^{i}_{st}) \times \frac{1}{m} \ldots\ldots\ldots (5)$$

5. Tag Refinement Strategy

If user were tagged in particular photo, user can ensure privacy control to the particular photograph. Customized access permission is used to control and avoid undesired tagging towards photograph. In fig 5.a., states the complete scenario and overall behavior of Access customization towards tagging image.

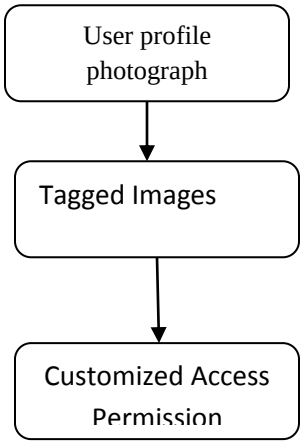
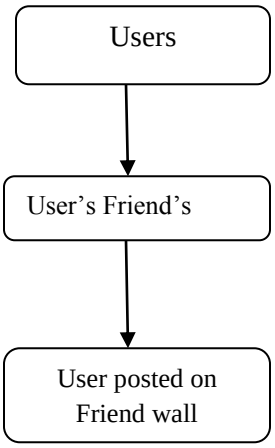


Fig 5.1 Access Customization over Tagged Images

5.2 Visibility Control Policy

If users post any message on a public wall, then the visibility of that post is governed as per the privacy policies of the user on whose wall we posted. However, In some situations user who is posting may want to control who among our common friends can view that post. This policy enables us to allow/disallow users among our common friends to view our post. It has been enabled by Access control schema.



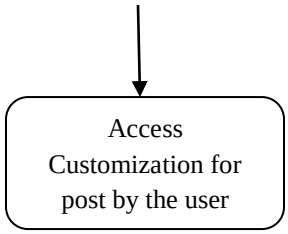


Fig 5.2 Visibility Control Policy

6. Evaluation Of Access Control Mechanism

A VisibilityControlList (VCL) is being used for the evaluation process. It is an enhancement of ACL (Access Control List) which used to state the privacy-level of each user and it state the privileges assigned to each individual user.

Blog_ID	Allow	Deny
1	✓	×
2	×	×
3	✓	✓
4	×	✓

Table 6 Visibility Access List

Based on VCL access permissions-whether to "Block "or "Deny " the user for appropriate content accessing over OSN.In fig 5.a Evaluation result indicates the level of privacy towards contents that was being preserved by AC2P Protocol .

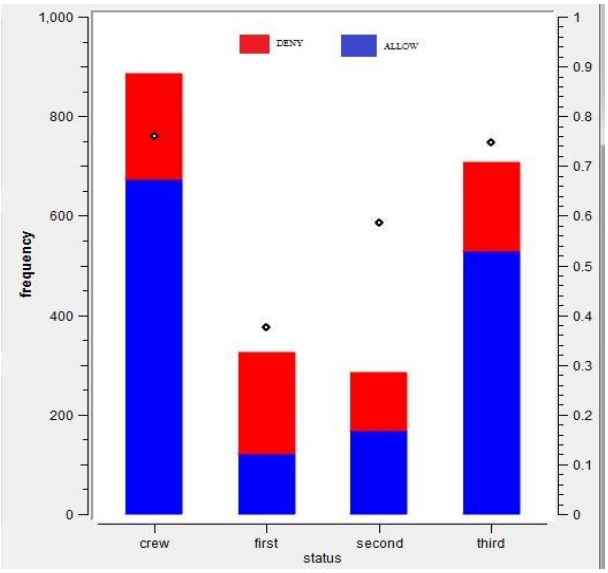


Fig 6.1 Analyzing of DCAM model.

In Fig 5.a, overall performance of a AC2P Protocol is evaluated based using (DV) value. This value is used for the prediction over OSN content.

6.1 Performance Evaluation

Based on the contents, which was shared between users, are determined on the basis of *outflow* and *inflow* strategy. It is determined using some constraint equ(6).

$$Ci = \begin{cases} \text{Outflow/Inflow} & \text{Outflow} < \text{Inflow} \\ 1 & \text{Outflow} > \text{Inflow} \end{cases} \quad \ldots\ldots\ldots(6)$$

Outflow = the number of interactions, user U*i* has with her friend.

Inflow = the number of interactions U*i's* friends

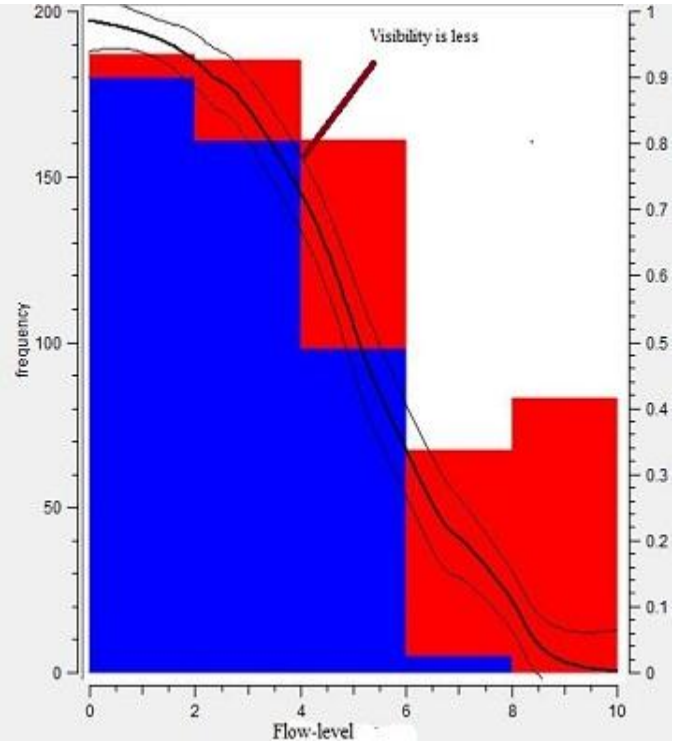


Fig 6.2 Probing the risk of leakage over OSNs contents

Finally, AC2P Protocol reduces risk of information leakage and assures ownership policy over user (U*i*).Above fig 6.b state the privacy-level managed using DCAM model.

7. CONCLUSION

In this paper, we has been proposed a better solution for both illegal content accessing of shared data and unwanted image tagging on OSNs. AC2P protocol was developed with Decision Scheme and Decision evaluation mechanism. In addition, we have introduced an approach for representing and reasoning about our proposed model.Tag Refinement is being proposed to avoid unwanted tagging and it preserve user and ensure ownership.So,user can make the decision to allow/disallow users among our common friends to view our post.

8. FUTURE WORK

In future work, we are planning to determine the comprehensive privacy conflict resolution approach[21],[22]and to probe the services of collaborative management of shared data in OSN's.We would study inference-based techniques [20] for automatically configure privacy preferences in the AC2P Protocol. Besides, we plan to systematically integrate the notion of trust and reputation into Decision making model and investigate a comprehensive solution to cope with collusion attacks for providing a robust Decision making service in OSNs.

REFERENCES

[1] Mark Hachman (April 23,2012). "Facebook Now Totals 1.20 billion Users, Profits Slip". PCMag.com. Retrieved September 24, 2013.

[2] Facebook Statistics, http://www.facebook.com/ Press/ info .php?statistics, 2013.

[3] G. Ahn and H. Hu, "Towards Realizing a Formal RBAC model in Real Systems," Proc. 12th ACM Symp. Access Control Models and Technologies, pp. 215-224, 2007.

[4] Facebook Privacy Policy, http://www.facebook.com/policy.php/, 2013.

[5] Google+ Privacy Policy, http://http://www.google.com/intl/en/+/policy/, 2013.

[6] B. Carminati, E. Ferrari, and A. Perego, "Rule-Based Access Control for Social Networks," Proc. Int'l Conf. On the Move to Meaningful Internet Systems, pp. 1734-1744, 2006.

[7] M. Harrison, W. Ruzzo, and J. Ullman, "Protection in Operating Systems," Comm. ACM, vol. 19, no. 8, pp. 461-471, 1976.

[8] B. Carminati, E. Ferrari, and A. Perego, "Enforcing Access Control in Web-Based Social Networks," ACM Trans. Information and System Security, vol. 13, no. 1, pp. 1-38, 2009.

[9] P. Fong, "Relationship-Based Access Control: Protection Model and Policy Language," Proc. First ACM Conf. Data and Application Security and Privacy, pp. 191-202, 2011.

[10] P. Fong, M. Anwar, and Z. Zhao, "A Privacy Preservation Model for Facebook-Style Social Network Systems," Proc. 14th European Conf. Research in Computer Security, pp. 303-320, 2009.

[11] S. Kruk, S. Grzonkowski, A. Gzella, T. Woroniecki, and H. Choi "D-FOAF: Distributed Identity Management with Access Rights Delegation," Proc. Asian Semantic Web Conf. (ASWC), pp. 140-154,2006.

[12] N. Li, J. Mitchell, and W. Winsborough, "Beyond Proof-of- Compliance: Security Analysis in Trust Management," J. ACM,vol. 52, no. 3, pp. 474-514, 2005.

[13] H. Hu and G. Ahn, "Enabling Verification and Conformance Testing for Access Control Model," Proc. 13th ACM Symp. Access Control Models and Technologies, pp. 195-204, 2008.

[14] Aimeur, E.; gambus,S.; Ai Ho; , "UPP: User Privacy Policy for Social Networking Sites,"Internet and Web Applications and Services,2009. ICIW '09.Fourth International Conference on vol., no., pp.267-272,24-28 May 2009.

[15] A. Ho, A4. Maiga, and E. Aimeur, "Privacy protection issues in social networking sites,"IEEE/Acs International Conference on Computer Systems and Applications 2009 (AICCSA 2009),PP.271-278,Country,2009

[16] SeyedHossein Mohtasebi and Ali Dehghantanha," A Mitigation Approach to the Malwares Threats of Social Network Services," Muktimedia Information Networking and Security,2009. MINES'09. International Conference on, vol.1,no.,pp.448-459,2011

[17] B. Carminati, E. Ferrari, and A. Perego, "Rule-Based Access Control for Social Networks," Proc. Int'l Conf. On the Move to Meaningful Internet Systems, pp. 1734-1744, 2006.

[18] P. Fong, "Relationship-Based Access Control: Protection Model and Policy Language," Proc. First ACM Conf. Data and Application Security and Privacy, pp. 191-202, 2011.

[19] L. Lam and C.Y. Suen, "Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance," IEEE Trans. Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 27, no. 5, pp. 553-568, Sept. 1997.

[20] A. Squicciarini, S. Sundareswaran, D. Lin, and J. Wede, "A3p: Adaptive Policy Prediction for Shared Images over Popular Content Sharing Sites," Proc. 22nd ACM Conf. Hypertext and Hypermedia, pp. 261-270, 2011.

[21] H. Hu, G. Ahn, and K. Kulkarni, "Anomaly Discovery and Resolution in Web Access Control Policies," Proc. 16th ACM Symp.Access Control Models and Technologies, pp. 165-174, 2011.

[22] H. Hu, G.-J. Ahn, and K. Kulkarni, "Detecting and Resolving Firewall Policy Anomalies," IEEE Trans. Dependable and Secure Computing, vol. 9, no. 3, pp. 318-331, May 2012.

Mr.J.S.HARILAKSHMANRAJ, Student at Department of Software Engineering at Sri Ramakrishna engineering college, Tamilnadu.He received Bachelor Degree from Kumaraguru college of Technology, Tamilnadu.His research interest focus on Social Network Analysis, Data Mining and Web-mining.

Dr.N.RAJKUMAR,Head of department of M.E. Software Engineering, at Sri Ramakrishna Engineering College, Tamilnadu. He received Ph.D Degree at Bharathiyar University during 2005.His specialization was Datamining, webmining.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 308 - 311, 2015, ISSN:- 2319–8656

A Detailed Study on Prevention of SQLI attacks for Web Security

Navjot Verma
Center of IT and Management
Punjabi University Regional
Mohali India

Amardeep Kaur
Center of IT and Management
Punjabi University Regional
Mohali India

Abstract: *SQL injection is the major susceptible attack in today's era of web application which attacks the database to gain unauthorized and illicit access. It works as an intermediate between web application and database. Most of the time, well-known people fire the SQL injection, who is previously working in the organisation on the present database. Today organisation has major concern is to stop SQL injection because it is the major vulnerable attack in the database. SQLI attacks target databases that are reachable through web front. SQLI prevention technique efficiently blocked all of the attacks without generating any false positive. In this paper we present different techniques and tools which can prevent various attacks*

Keywords: SQL Injection, SQL injection Prevention, web application, database, vulnerable.

1. INTRODUCTION

Web applications are being in a much wider area these days, online shopping, online banking and social networking is some of the key users of these [21]. All these users have the utmost priority for their privacy and security and these are the most vulnerable while being online. To secure these applications two phases are implied. These phases are:

1.1 Data base layer: The *database layer* provides an object vision of database information by applying schema semantics to database records, so isolating the upper layers of the directory service from the underlying database system. The database layer is an inner boundary that is not exposed to users. No database admission calls are made directly to the Extensible Storage Engine; as an alternative, all database right to use is routed through the database layer [23].

1.2 Application layer: refers to techniques of shielding Web applications at the application layer (last layer of the seven-layer OSI model) from nasty attacks that may picture private information. Protection is applied to the application layer especially to protect against illegal access and attacks.

Advantages of Web Security:

1. Internet sites are well-liked targets for crackers, and even without mean forces security holes can permit accidents happen.
2. A secure network is a good network.
3. This doesn't make it easy to take part in a web site. Scripts may require access to sensitive information, or at least information you don't want in the public domain before you are prepared.

SQL injection is one of the most serious threats to the data security of all web applications. SQL injection attack allows attackers to gain control of the original query, illegal access to the database and extract or transform the database [1]. The main cause of SQL injection vulnerabilities is: attackers use the input support to attack strings that contains special database commands. An SQLIA occurs when an attacker change the SQL control by inserting new keywords [2]. A successful SQLI attack hinder privacy integrity and availability of information in the database. In most of cases, SQL Injection is used to initiate the denial of service attack on web applications. The strictness of the attacks depends on the role or account on which the SQL statement is executed.
An attacker needs to know loop holes in the application before launch an attack. Attackers use: input format, timing, performance and error message to decide the type of attack suitable for an application. Database is the heart of many web applications, basis for which database more and more coming under great number of attacks. SQLIAs occur when data provided by the user is incorporated directly in the query and is not appropriately validated.

1.3 Vulnerability

Table 1 present the most common security vulnerabilities found in web programming languages [21].

Table 1: Types of Vulnerabilities

Vulnerability Types	Description
Type I	No clear distinction between data types received as input in the programming language used for the web application development

Type II	Delay of operation analysis till the runtime phase where the current variables are considered rather than the source code expressions.
Type III	The validation of the user input is not definite. Attacker taking advantages of insufficient input validation can utilize mean code to conduct attacks.
Type IV	Puny concern of type specification in the design. A number can be used as a string or vice versa.

2. TYPES OF SQLI ATTACK

The SQL injection attacks can be performed using a variety of techniques. Some of them are specified as follows:

First Order Attack: Attackers aim the database with strings attached to an input field and receives the answer immediately. Such attacks which exploit the lack of validation in the input field parameter are known as first order attacks [21].

Second Order Attack: An attacker attacks the database with inserting mean queries in a table but implement these queries from other actions [21].

Tautology Attack: Conditional operators are used by the attackers in the SQL queries such that the query always evaluates to TRUE [1,2,6,10].
For example, SELECT * FROM employee WHERE name = '' OR '1'='1';

Logically Incorrect Queries: An illegal query used by the attacker to glance at the whole database [1,2,6,10].
For example, "SELECT * FROM employee WHERE id =" + name + ";"

Piggy-backed Query: In this attack, attacker tries to add on supplementary queries but terminates the first query by inserting ";" [1,2,7].
For example, SELECT * FROM employee WHERE id=1; DROP TABLE employee;

Inference: The main goal of the inference based attack is to change the activities of a database or application. There are two well-known attack techniques that are based on inference: blind injection and timing attacks

Timing attack: In these types of attack an attacker observe the database delays in database response and gather the information. WAITFOR, IF, ELSE, BENCHMARK [1,2] cause delay in database response.
For example, SELECT * FROM employee WHERE id=1-SLEEP(15);

Blind injection: In this situation an attacker performs queries that have a Boolean result [1].
For example: SELECT * FROM employee WHERE id = '1008' AND 1=1;

Alternate Encoding: Attacker modifies the injection query by using alternate encoding such as hexadecimal, ASCII and Unicode [1,2] .
For example: SELECT * FROM employee WHERE id=unhex('05'); .

Union Query: An attacker makes use of vulnerable parameters and attach injected query to the safe query by the word UNION and get data about other tables from the application.
For example: Select * from company where name="" union select * from employee –,,and Password="anypwd"

Stored Procedure [10]: A stored procedure is a cluster of Transact-SQL statements compiled into a single execution plan. As stored procedure could be coded by programmer, attacker can execute these built in procedures**.**

3. RELATED WORK

There are many ways to prevent SQL injection attacks [1,2,7,14,15]. The various types of existing techniques for preventing SQL injection are as follows:

Negative Tainting, [1] Preventing SQL injection attack using negative tainting provide uniqueness by using linked list. This approach works on the untrusted strings and provides good response time for large database programs. This approach consists of (1) Identifying hot spot from the application (2) To find out SQL injection attack using negative tainting. (3) Inserting newly identified SQL injection attacks to get better accuracy.

Positive Tainting, [2] Positive tainting focuses on the recognition and marking of trusted strings. It uses the concept of syntax sensitive estimation. This system works in following manner- (1) Identifying trusted data source. (2) Allowing only data from such sources to suit a SQL keyword or operator in query strings. Trusted data strings can be more readily known. WASP (Web Application SQL injection Preventer) tool have implemented this approach. This approach is defined at the application level and it requires no alteration of the runtime (JVM) system, and it imposes low execution overhead. Positive tainting used to check SQLIA at the runtime. WASP tool works fruitfully but it blocked over 12000 attacks without generating false positives.

Input Filter Technique, [4] An SQL injection attack is interpreted differently on different databases. This technique provides the general solution to solve this problem. Depending on number of space, double dash and single quote the count value of the input value is increased by 1 because default count of the query is 1. Then the fixed count value and dynamically generated count compared to check SQL injection attack. But this technique has limitation that it works on single quote, double dash and space only.

Query Transformation and Hashing, [7] This technique uses a lightweight method to prevent SQL injection and works in two ways. Fist is to convert the query into structural form than parameterized form. Second apply an appropriate hash function to create unique hash key for each transformed query by using suitable hash function. Only the hash keys are stored instead of transformed query. A primary index can be created for fast and proficient searching. As this approach is proficient but it does not prevent second order SQL

injection attacks and this approach can neither be applied to prevent XSS attacks.

SQL-ID, [14]

Kemalis and Tzouramanis have suggested novel specification-based methodology for the detection of exploitations of SQL injection vulnerabilities in "Specification based approach on SQL Injection detection" [3]. A Java-based application monitors by this system and identify SQL injection attacks in real time.

Dynamic Candidate Evaluations Approach, [15]

Bisht et al. propose CANDID. It is a Dynamic Candidate Evaluations method for automatic prevention of SQLInjection attacks. This structure dynamically extracts the query structures from every SQL query location which are intended by the developer (programmer). Hence, it solves the issue of manually modifying the application to create the prepared statements.

AMNESIA,[16] - AMNESIA approach for tracing SQL input flow and generating attack input, JCrasher for generating test cases, and SQLInjectionGen for identifying hotspots. The experiment was conducted on two Web applications running on MySQL1 1 v5.0.21. Based on three attempts on the two databases, SQLInjectionGen was found to give only two false negatives in one attempt. This framework is efficient considering the fact that it emphasizes on attack input precision. Besides that, the attack input is properly matched with method arguments. The single disadvantage of this approach is that it involves a number of steps using different tools.

SQLrand,[17] SQLrand approach is proposed by Boyd and Keromytis. To implement this approach, they use a proof of concept proxy server in between the Web server (client) and SQL server; they de-randomized queries received from the client and sent the request to the server. This de-randomization framework has 2 main advantages: portability and security. The proposed scheme has a good performance: 6.5 ms is the maximum latency overhead imposed on every query.

SQLIA Prevention Using Stored Procedures,[18] Stored procedures are subroutines in the database which the applications can make call to . The prevention in these stored procedures is implemented by a combination of static analysis and runtime analysis. The stagnant analysis used for commands identification is achieved through stored procedure parser and the runtime analysis by using a SQLChecker for input identification. This paper has proposed a combination of static analysis and runtime monitoring to fortify the security of potential vulnerabilities.

Adaptive algorithm,[19]

This method consists of the top features of parse tree validation technique and code conversion method. This technique parse the user input and verify whether its prone, if there is any chance of vulnerability present then code conversion will be applied over the input. This paper had also surveyed various SQL injection methods and techniques against SQL injection. It has also presented the algorithm to apply for the vulnerable code.

4. COMPARISON

The main goal for future improvement is to improve the efficiency of the technique by reducing false positive. Table 2, [1] shows a chart of the schemes and their prevention capabilities against various SQL injections attacks and précis the results of this comparison.

Table 2: comparison of various prevention schemes and various attacks

Schemes	Tautology	Logically Incorrect Queries	Union Query	Stored Procedure	Piggy Backed Queries	Inference Attack
AMNESIA	YES	YES	YES	NO	YES	YES
SQLrand	YES	NO	NO	NO	YES	YES
CANDID	YES	NO	NO	NO	NO	NO
SQLGuard	YES	NO	NO	NO	NO	NO
SQLIPA	YES	YES	YES	NO	YES	YES
Negative Tainting	YES	YES	YES	NO	YES	YES
Positive Tainting	YES	YES	YES	YES	YES	YES

5. CONCLUSION

With above data in place it can be reviewed that database security is major issue in today life. This paper presents a survey report on the SQL injection attacks and how attacks are implemented on the database using SQL queries. We first identified the various types of SQL injection attacks and then we examine various prevention techniques. Finally a comparative analysis of various types of prevention techniques of SQL injection is presented.

6. REFERENCES

[1] A. S. Gadgikar, "Preventing SQL injection attacks using negative tainting approach," in IEEE International Conference on Computational Intelligence and Computing Research, 2013, pp. 1–5.

[2] W. G. J. Halfond, A. Orso, and P. Manolios, "Using positive tainting and syntax-aware evaluation to counter SQL injection attacks," in Proceedings of the 14th ACM SIGSOFT International Symposium On Foundations of Software Engineering - SIGSOFT '06/FSE-14, 2006, pp. 175–185.

[3] S. Roy, A. K. Singh, and A. S. Sairam, "Detecting and Defeating SQL Injection Attacks," International Journal of Information and Electronics Engineering., vol. 1, no. 1, pp. 38–46, 2011.

[4] S. Bangre and A. Jaiswal, "SQL Injection Detection and Prevention Using Input Filter Technique," International Journal of Recent Technology and Engineering (2012), vol. 1, no. 2, pp. 145–150, 2012.

[5] E. Bertino, A. Kamra, and J. P. Early, "Profiling database applications to detect SQL injection attacks," in Conference Proceedings of the IEEE International Performance, Computing, and Communications Conference, 2007, pp. 449–458.

[6] A. Sadeghian, M. Zamani, and A. A. Manaf, "A Taxonomy of SQL Injection Detection and Prevention Techniques," in 2013 International Conference on Informatics and Creative Multimedia, 2013, pp. 53–56.
[7] D. Kar and P. Suvasini, "Prevention of SQL Injection Attack Using Query Transformation and Hashing," in Proceedings of the 2013 3rd IEEE International Advance Computing Conference, IACC 2013, 2013, pp. 1317–1323.
[8] R. Dharam and S. G. Shiva, "Runtime Monitors for Tautology based SQL Injection Attacks," IEEE Int. J. Cyber-Security Digit. Forensics, vol. 53, no. 6, pp. 253–258, 2012.
[9] P. Kumar and R. Pateriya, "A Survey on SQL injection attacks, detection and prevention techniques," in Computing Communication & Network Technologies, 2012, no. July, pp. 1–5..
[10] X. Fu, X. Lu, and B. Peltsverger, "A static analysis framework for detecting SQL injection vulnerabilities," in 31st Annual International Computer Software and Application Conference, 2007, no. Compsac, pp. 87–96.
[11] S. Thomas, L. Williams, and N. Carolina, "Using Automated Fix Generation to Secure SQL Statements [Short presentation paper]," 2007.
[12] K.-X. Zhang, C.-J. Lin, S.-J. Chen, Y. Hwang, H.-L. Huang, and F.-H. Hsu, "TransSQL: A Translation and Validation-Based Solution for SQL-injection Attacks," in 2011 First International Conference on Robot, Vision and Signal Processing, 2011, pp. 248–251.
[13] K. Kemalis and T. Tzouramanis, "SQL-IDS : A Specification-based Approach for SQL-Injection Detection," pp. 2153–2158, 2008.
[14] P. Bisht, "CANDID : Dynamic Candidate Evaluations for Automatic Prevention of SQL Injection Attacks," ACM Int. J. Comput. Sci., vol. V, no. 2, pp. 1–38, 2010.
[15] G. T. Buehrer, B. W. Weide, and P. A. G. Sivilotti, "Using Parse Tree Validation to Prevent SQL Injection Attacks," no. September, pp. 106–113, 2005.
[16] S. W. Boyd and A. D. Keromytis, "SQLrand : Preventing SQL Injection Attacks," IEEE Appl. Cryptogr. Netw. Secur.,2004,vol. 3089, pp. 292–302..
[17] W. G. J. Halfond and A. Orso, "Preventing SQL injection attacks using AMNESIA," in Proceeding of the 28th international conference on Software engineering - ICSE '06, 2006, p. 795.
[18] Z.Su and G. Wassermann, "The Essence of Command Injection attacks in Web Applications", 33rd ACMSIGPLAN, SIGACT Symposium on Principles of Programming Languages, Charleston, South Carolina, USA, 2006,pp. 372-382.
[19] Ashish John "SQL Injection Prevention by Adaptive Algorithm," IOSR Journal of Computer Engineering, 2015,Vol 17,pp. 19-24
[20] Diksha Upadhya "A Survey on SQL Injection - Vulnerabilities Attacks and Prevention Techniques".
[21] K. S. Chavda, "International Journal of Advance Engineering and Research," Sci. J. Impact Factor, vol. 1, no. 12, pp. 173–179, 2014.
[22] A. John, A. Agarwal, and M. Bhardwaj, "An adaptive algorithm to prevent SQL injection," *An Am. J. Netw. Commun.*, vol. 4, pp. 12–15, 2015.
[23] https://technet.microsoft.com/en-us/library/cc961800.aspx

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 312 - 317, 2015, ISSN:- 2319–8656

Virtual Organization of Grid – Pipeline Virtual Organization (PVO) Approach

J.Prema
Department of CSE
SRM University
Ramapuram
Chennai, India

Dojohn Loyd B.
Department of CSE
SRM University
Ramapuram
Chennai, India

Abstract: Grid computing is a modularized way of structuring network resources so as to share the information and resources, to perform heavily intense problems. Grid computing is a part of distributed processing which allows the distribution of the problem over multiple computational resources. The formation of grid decides the computation cost, the performance metrics of the operation to be carried out. The grid computing favours the formation of an adhoc structure formed at the time of request (it does not hold an infrastructure) that is termed as virtual organization. This paper presents dynamic source routing for modeling virtual organization.

Keywords: Grid computing, grid layers, virtual organization, MSVOF Mechanism, minimum path algorithm, pipelining VO formation

1. INTRODUCTION

In the modern world, scientific problem has grown beyond proportion; hence the computation could take time anywhere between hours and years. This is due to the presence of huge data set or the complexity of the computation. The best way to deal with this is the distribution of problem's data set over multiple computational units which led to the concept of GRID computing. Grid is a type of parallel and distributed system that enables sharing, selection and aggregation of geographically distributed autonomous resources dynamically based on their availability, capability and cost.

2. ARCHITECTURE OF GRID

Grid computing involves a five layered architecture providing protocols and services at different levels of layers.

2.1 Layers of Grid

2.1.1 Fabric Layer:

Fabric layer takes the bottom of the layered architecture that provides shareable resources such as network bandwidth, CPU, time, memory, etc. Operating System, queuing system and processing kernel also forms the part of this layer.

2.1.2 Connectivity Layer

The protocols related to the communication and authentication becomes the part of this layer.

2.1.3 Resource Layer

This layer specifies the protocols for operating with the shared resources. These protocols are concerned about the allocation and monitoring of resources.

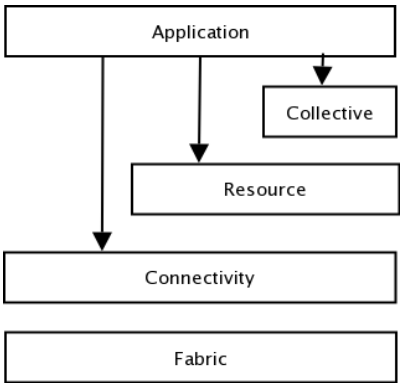


Figure : Architecture Of Grid

2.1.4 Collective Layer

This layer holds general purpose utilities like directory services, diagnostic services, data replication services, etc.

2.1.5 Application Layer

This layer is the placeholder where the user's applications are deployed.

2.2 Types of Grid

Grids are categorized into five types based on their utility.

2.2.1 Computational Grids:

These are the grids capable of providing secure access to computational resources for complex computation which would require high computing machines.

2.2.2 Utility Grids:

These are the grids that promote the sharing of CPU cycles, software and peripherals like sensors, etc.

2.2.3 Data Grids:

These are the grids that offers database storage related functionalities.

2.2.4 Collaborative Grids:

These are the grids that lead to formation of social web forums.

3. VIRTUAL ORGANIZATION IN GRID

A virtual organization is a collection of geographically distributed functionally and diverse entities that are linked to each other by means of electronic communication. VO are always formed suddenly in an adhoc fashion, they perform task assigned and then dismantle them immediately. The mechanism that decides the formation of virtual organization pattern contributes to the revenue and the cost of the computation.

Dynamic grid Virtual organization is collections of workspaces with the computing systems which gets tasks in a sequential manner and gets a separate method for each computing system and get that job completed and gives the output.

3.1 Necessity for VO Formation Algorithm:

Virtual Organization is formed with electronic communication. But when it is grouped together formation is more important to organize it in a sequential way. So that the task can be performed in a good manner and then the cost also will be decreased so that we are using formation algorithms for VO. We have used some previous algorithms like Merge and split VO formation (MSVOF) and Random Select Virtual organization Formation (RSVOF), Same Size Virtual organization formation (SSVOF), Grand Coalition VO formation (GVOF). In this paper we have introduced Pipelining Virtual Organization Formation (PVOF).

4. MSVOF MECHANISM

This mechanism is to find the minimum cost between the nodes and splitting it into some partitions and execute their program. This algorithm comes after some analysis on Random Select VOF and Grand Coalition VOF.

VO ={{Vo1},... {Vom}}
Map program P on each Si € VO
repeat
stop ←T rue
for all *Si, Sj € VO, i≠ j* ***do***
visited[Si][Sj] ← False
end for
{Merge process starts:}
repeat
flag ←T rue
Randomly select Si,Sj € VO for which
Visited [Si][Sj]=False; i ≠ j
visited[Si][Sj]←T rue
B&B-MIN-COST-ASSIGN(Si U Sj)
{Map program T on Si [Sj}
if *(Si U Sj) €{Si,Sj}* ***then***
Si ←Si U Sj {merge Si and Sj}
Sj←ϕ {Sj is removed from VO}
for all *Sk €VO, k ≠ i* ***do***
visited[Si][Sk] ←False
end for
end if
for all *Si,Sj€VO, i ≠ j* ***do***
if *not visited [Si][Sj]* ***then***
flag ←False
end if
end for
until *(|VO|=1) or (flag ←T rue)*
{Split process starts:}
for all *Si € VO where |Si| > 1* ***do***
for all *partitions {Sj,Sk} of Si,*
where Si=SjU Sk, Sj ∩ Sk=ϕ ***do***
B&B-MIN-COST-ASSIGN(Sj)
{Map program P on Sj}
B&B-MIN-COST-ASSIGN(Sk)

{Map program P on Sk}

if *{Sj,Sk}€Si* ***then***

Si ←Sj {that is VO=VO\Si}

VO=VO U Sk

stop ←False

Break (one split occurs; no need to check other splits)

end if

end for

end for

until *stop=T rue*

Find k=arg max Si € VO {v(Si)/\|(Si)|}

Map and execute program P on VO Sk

Algorithm-1 : MSVOF Algorithm

Symbols	Description
VO	Virtual Organization
P	Program
Si, Sj, Sk	Nodes

We can see this algorithm has 2 major parts. In the first part of this algorithm it finds the min cost of consecutive nodes and merging into VO through merge algorithm. In the second part the merged nodes are separated through split algorithm based on their cost. It is based on the nodes cost and not about the job. Also, the execution time is not measured accurately. So we are moving for another formation mechanism called PVOF (Pipelining Virtual Organization Formation).

Here we are proposing pipelining VO formation. This is used to split the job first, then process the job, then concatenate it for the desired output. To get the shortest distance and choosing the correct node we are using the min path algorithm. It used to connect all the nodes and the shortest distance between all the nodes.

4.1 Min Path Algorithm:

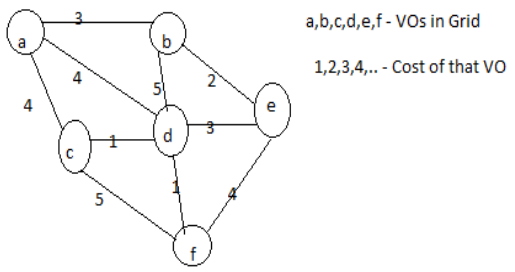


Fig. 1.1 GIven VO Path to find MIn Path

To find the minimum cost we are using the min path algorithm. From this all the modes of the VO will be connected and the resources will be shared among the Grids. The cost will be the distance between one node to the other. This min path algorithm to define the minimum cost of the traversing nodes as well as the time consumed is given as follows:

```
IntV = VO->V;
  structnode result[V];
  intx = 0;  // E An index variable, used for result[]
  intj = 0;  // A variable, used for sorted nodes
// array of nodes
  nsort(VO->node, VO->X, sizeof(VO->node[0]), class);
  // Allocate memory for creating V ssubsets
  structsubset *subsets =
    (structsubset*) malloc( V * sizeof(structsubset) );
  // Create V subsets with single elements
  for(intv = 0; v < V; ++v)
  {
    subsets[v].parent = v;
    subsets[v].rank = 0;
  }
  // Number of nodes to be taken is equal to V-1
  while(x< V - 1)
  {
    // Step 2: Pick the smallest node. And increment the
index
     // for next iteration
     structnodenext_node = VO->node[i++];
      intk = find(subsets, next_node.src);
     intl = find(subsets, next_node.dest);
      // If including this nodedoes't cause cycle, include it
     // in result and increment the index of result for next
node
     if(k != l)
     {
       result[x++] = next_node;
       Union(subsets, k, l);
     }
     // Else discard the next_node
  }
      for(i = 0; i< x; ++j)
```

```
    print("%d -- %d == %d\n", result[j].src, result[j].dest,
                                result[j].weight);
  return;
}
```

Algorithm 2 : Min Path Cost Algorithm

Symbols	Description
V	Virtual Organization
n	No. of Nodes
k	Index
node	VO node

By using this algorithm we are finding the minimum cost path for the given VO formation. With respect to the given cost, the minimum path has been detected. The VO following this will be the final solution for the given problem.

After detecting the minimum path we have to divide the original job into the sub tasks. Then the subtasks will be divided into some other sub tasks. In this Fig 2.1 the minimum path calculated VO has been given.

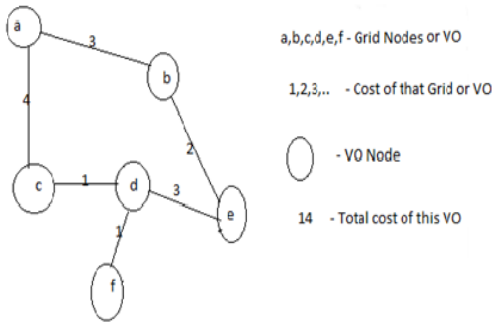


Fig 2.1 MIn path calculated VO

5. PIPELINING VIRTUAL ORGANIZATION FORMATION:

We are going for the pipeline algorithm which can be used to divide the task into sub tasks and then process it. It will process the subtask1; at the same time it will process the subtask2 in another VO. So it works with parallel time at distributed VO grids. It is given by the algorithm as follows:

G={{Vo1},{Vo2},{Vo3},…}
for all VOx,VOy €G where x≠y do
for all partitions {Vox,Voy} of G
where Vox=Vox U Voy ∩ Voz
Map program P for min cost on Voy
Map program P for min cost on Voz
If {Voy,Voz}>Vox then
Vox←Voyi.e G={G/Voy}
G=GUVoz
Stop ←false
Break (one split occurs;no need to check other splits)
end if
end for
end for
until stop = True
Find z=arg max Vo€G{Vox/Voz}
Map and execute program P on VOz

Algorithm 3 : Parallel VO Formation

Symbols	Description
G	Virtual Organization
P	Program
Vox, Voy	VO node

This algorithm is used to break the task into sub tasks and then doing the jobs in parallel time. In this algorithm G is having the virtual organization group which is divided into some other VO. Then the task will be provided into that VO. VO1 is the group having some grid formation. In that, the tasks will be shared among them and processed.

The minimum cost of VO will be calculated before executing this algorithm. Based on that cost, the task will be shared and executed. The advantage of this approach is to achieve the output by using parallel systems and memory usage efficiently.

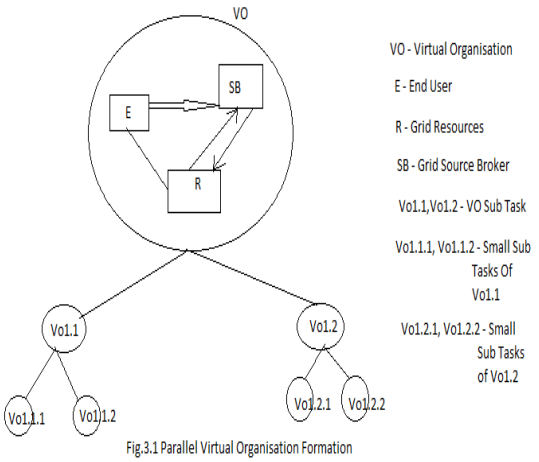


Fig.3.1 Parallel Virtual Organisation Formation

6. EXPERIMENTAL ANALYSIS:

In network stimulator (NS2), we have implemented PVOF and MSVOF with different number of nodes and cost. It is given from the output that PVOF is better than MSVOF based on their performance, since PVOF is performing many tasks at the same time. So it is better in performance. NS2 stimulates the MSVOF as well as PVOF with the exact monitoring values.

Based on the cost and number of nodes the PVOF will give output. If the number of nodes is increased, then the minimum cost will also be calculated and the cost also increased. Based on the experimental and performance analysis it shows that the PVOF is better than MSVOF. It has been calculated through our experiments.

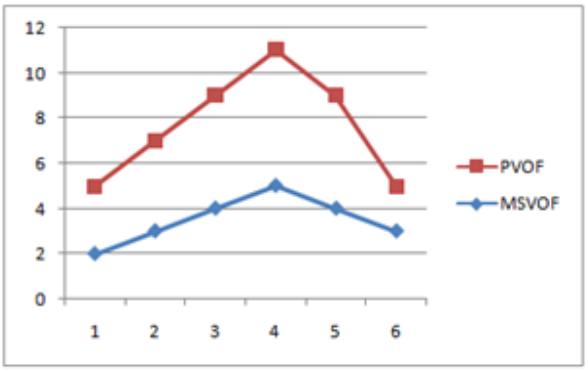


Fig 5.1 Experimental Analysis

The PVOF is based on time and not based on the number of nodes, because it splits the task with respect to time. MSVOF does not split the task. PVOF gives the output based on time. The proposed PVOF gives better performance than MSVOF based on the experimental results.

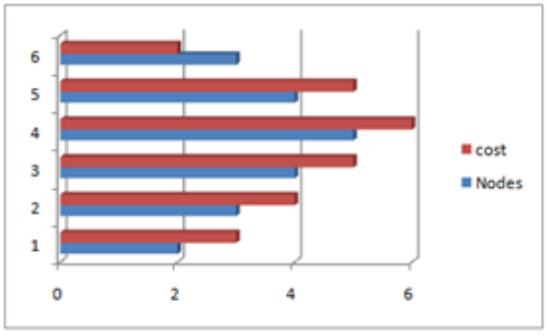


Fig 4.1 Perfomance Analysis Graph

7. RELATED WORKS:

Grid Information service architecture defines low level enquiry protocol and registration protocol to incorporate individual entities and discovery strategies.

These combined Grid protocols are constructing high level services that are having the capabilities of brokering, monitoring, fault detection and trouble shooting.

Through Monitoring and Discovery System (MDS-2) architecture, information services have been implemented in Globus tool kit which is widely used. This architecture has to develop flexible configuration tools for VO formation and should extend our security model. [1]

The Xtreem OS (Linux-based Operating System) is used to support VO Management. It is based on Linux operating system. It is providing core grid services to handle different collaboration models. It ensures the scalability and compatibility across all the applications. This is implemented in three flavors, i.e. Clusters, PCs and Mobiles. This will pose new challenges in cloud computing for trustworthy mechanism and services. This will take an effort to explore the architecture for Cloud Domains for managing credentials and ensuring security among users and resources. [2]

Security will play an essential part in exchanging data. SSL (Secure Socket Layer) is used to exchange data instead of message level security. When the message level security is not required we can use SSL. In message level security Web Services (WS) Secure Connection is used. But it is not suited for huge number of clients. [3]

When Globus toolkit is facing few security issues, it has been proposed with some improvements from the previous versions and evolution of open grid services architecture. It needs some more to extend experimental works in WS routing for compatibility, implementing the standard services and identity mapping. [4]

The Virtual Grid is similar to physical grid with the workspaces of their applications. It should have coarse grained control over workspace images via attestation. Thus virtual organization works flexibly and independently in a specific way and provides the workspace to be deployed from different communities and platforms. It reduces the administrative cost of maintaining grid resources. Much research is going for virtual playground which consists of group of virtual organization and also ensuring the security in the virtual playground. [5]

The Grid Trust Security framework is used for trust and security in the Next Generation grid architecture. The Grid Trust approach has 3 layers. Grid Application layer, Grid Service middleware layer, grid foundation layer. This architecture works with the last 2 layers. It composes of tools and services authorizing credentials and ensuring that the service task has been completed. It uses a secure resource broker and a reputation service in the middleware and in the foundation layer. Also, it has VO level policies and computational level policies. [6]

8. CONCLUSION:

We have already started using Grid Programming, so it is important to know what kind of results we are going to give to the real world. Here we have used PVOF for organization of the Virtual Grids. In future, exploring the dynamic representation of grid and examining the process flow will be an added advantage. It will reduce the cost management and the execution time.

9. ACKNOWLEDGMENTS:

I hereby thank Mr.DOJOHN LOYD B. for his valuable guidance for preparation of this paper. I am also grateful to our HOD, Mr.J.Jagadeesan and other staff members of Computer Science Department, SRM University, Ramapuram Campus, for their support and guidance.

10. REFERENCES:

[1] Karl Czajkowski, Steven Fitzgerald, Ian Foster, Carl Kesselman, "*Grid Information Services for Distributed Resource Sharing"*

[2] Massimo Coppola, YvonJégou, Brian Matthew, Christine Morin, Luis Pablo Prieto, Óscar David Sánchez, Erica Y. Yang, Haiyan Yu, "*Virtual Organization Support within a Grid-Wide Operating System"*

[3] Von Welch, Frank Siebenlist, Ian Foster, John Bresnahan, Karl Czajkowski, JarekGawor, Carl Kesselman, Sam Meder, Laura Pearlman, Steven Tuecke, "*Security for grid Sevices"*

[4] Satoshi Shirasuna, AleksanderSlominski, Liang Fang, Dennis Gannon, "*Performance Comparison of Security Mechanisms for Grid Services".*

[5] K. Keahey et al., "*Virtual Workspaces: Achieving Quality of Service and Quality of Life in the Grid*," Scienti!c Programming J., special issue on dynamic grids and worldwide computing, vol. 13, no. 4, 2005, pp. 265–275

[6] Philippe Massonet, "*Trust and security for next generation Grids"*

[7] Lena Mashayekhy, Student Member, IEEE, and Daniel Grosu, Senior Member, IEEE. "*A Merge-and-Split Mechanism for Dynamic Virtual Organization Formation in Grids",* VOL. 25, NO. 3, MARCH 2014

[8] Krsul, I., A. Ganguly, J. Zhang, J. Fortes, and R. Figueiredo. "*VMPlants: Providing and Managing Virtual Machine Execution Environments for Grid Computing. in SC04".* 2004. Pittsburgh, PA

[9] Keahey, K., I. Foster, T. Freeman, X. Zhang, and D. Galron, "*Virtual Workspaces in the Grid."* ANL/MCS-P1231-0205, 2005

[10] Foster, I., C. Kesselman, and S. Tuecke, "*The Anatomy of the Grid: Enabling Scalable Virtual Organizations."*International Journal of Supercomputer Applications, 2001. 15(3): p. 200-222

[11] Chase, J., L. Grit, D. Irwin, J. Moore, and S. Sprenkle, "*Dynamic Virtual Clusters in a Grid Site Manager."* accepted to the 12th International Symposium on High Performance Distributed Computing (HPDC-12), 2003.

[12] Foster, I. and Kesselman, C. "*Globus: A Toolkit-Based Grid Architecture".* Foster, I. and Kesselman, C. eds. The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann, 1999, 259-278.

[13] Foster, I., Kesselman, C., Tsudik, G. and Tuecke, S. "*A Security Architecture for Computational Grids".* ACM Conference on Computers and Security, 1998, 83-91

[14] Nakada, H., Sato, M. and Sekiguchi, S. "*Design and Implementations of Ninf: towards a Global Computing Infrastructure. Future Generation Computing Systems,"* 1999.

[15] Wolski, R. Forecasting "*Network Performance to Support Dynamic Scheduling Using the Network Weather Service."* In Proc. 6th IEEE Symp. on High Performance Distributed Computing, Portland, Oregon, 1997.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 318 - 321, 2015, ISSN:- 2319–8656

Data hiding using Pixel Value Differencing

Shruti
Deptt. Of Computer Science
Guru Nank Dev University
Gurdaspur, India

Abstract:-This paper presents embedding of data in an image using pixel-value differencing technique. This scheme is used to embed large amount of data by changing the difference between two pixels so that we are able to increase the embedding capacity. There is another technique that is pixel value shifting which also increase the embedding capacity but according to this scheme capacity will increase at edge areas of image.

Keywords: - Capacity; Pixel-value differencing; Image quality

1. INTRODUCTION

Nowadays, the Internet has become a common communication channel. Communicating in public system means that some problems need to be faced, such as data security, copyright protection, etc. Ciphering is a well-known method for security protection but it has the disadvantage of making a message unreadable thereby attracting the attention of eavesdroppers. This makes steganography which hides data within data a good choice for secret communications. One of the best-known steganographic methods is the least significant- bit (LSB) substitution[1]. The simple LSB substitution method replaces the length-fixed LSB with the fixed length bits. Although the technique is efficient, it is rather easy to create a noticeable distortion for the human eye or can be detected by some programs. Therefore, several adaptive methods have been proposed for steganography in order to decrease the distortion caused by the LSB substitution[3]. In addition, some methods use the concept of human vision to avoid the detection of Programs. Now a technique "pixel-value differencing" steganographic method that used the difference value between two adjacent pixels in a block in order to determine the number of embeddable secret bits . This difference value is adjusted so as to embed the secret bits, and the difference between the original and new difference values is adjusted between the two pixels. To check the proposed method, author applies the dual statistics method , called as RS-diagram, to detect the function of embedding method. In RS-diagram, first of all, the discrimination and flipping functions are applied to define pixel groups: Regular (R), Singular (S), and Unusable (U). Then, the percentages of all groups of Regular and Singular with masks m= [0110] and −m= [0−1−1 0] are computed, in which they are represented as Rm, R−m, Sm, and S−m, respectively. Finally, the RSdiagram applied hypotheses of Rm ~= R−m and S ~= S−m S−m to present the detected resultant. There is a method which combines the pixel-value differencing and LSB replacement method. This approach provides higher capacity[2] than pixel-value differencing, but it does not pass the detection of RS-diagram. But pixel-value differencing method not only provides high capacity but also passes the detection of RS-diagram:-

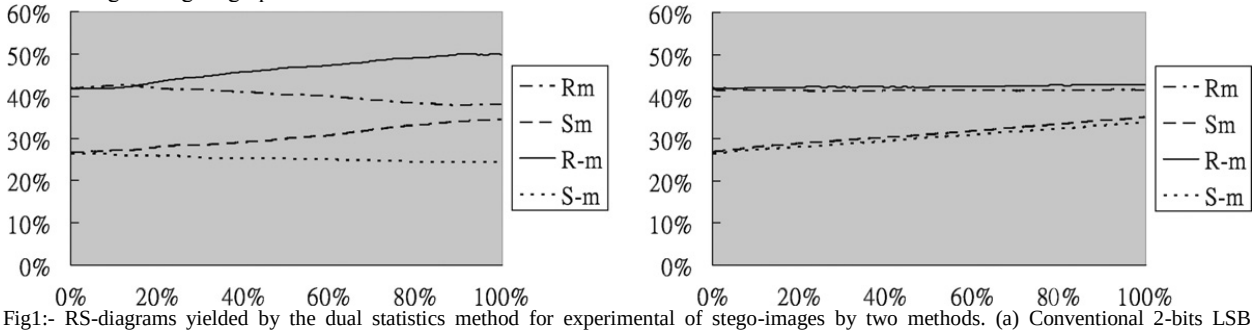


Fig1:- RS-diagrams yielded by the dual statistics method for experimental of stego-images by two methods. (a) Conventional 2-bits LSB substitution; (b) pixel-value difference method.

In this paper, we propose an efficient steganographic scheme to hide data imperceptibly in gray-level images. This scheme is based on the property of human eye, which is more sensitive to the change in the smooth area than the edge area. In this paper, we process a block of four neighboring pixels simultaneously. The number of secret bits to be embedded in a block depends on the degree of smoothness or sharpness. Each four-pixel block is divided into two two-pixel groups, and each group is processed using the approach of pixel-value differencing. In order to extract the embedded data correctly, some schemes are designed cleverly to differentiate which pixels belonging to different groups.

Some conditions will cause a block to be abandoned without embedding. To overcome this obvious drawback, a new technique known as pixel-value shifting is proposed.

2. LITERATURE REVIEWS

Steganographic method hides secret data in graylevel images by pixel-value differencing[4]. First, the host image is partitioned into non-overlapping consecutive two-pixel blocks by scanning all the rows of the host image in a zigzag manner. A difference value d is calculated from the two pixels, say pi and pi+1, of each block. By symmetry, only the possible absolute values of d (0 through 255) are considered and they are classified into a number of contiguous ranges, called as Ri, where i=1, 2, 3, . . ., n. The width of Ri is ui −li + 1, where ui is the upper bound of Ri and li is the lower bound of Ri. The width of each range is taken as a power of 2. This restriction of width facilitates the embedding of binary data. If d falls in smooth area, less secret data will be hidden in the block. On the other hand, if d falls in sharp area, then the block has higher tolerance and thus more secret data can be embedded inside it. Suppose that d falls into the range Rk. The number of embedding bits is determined by the width of Rk. Therefore, the embedding operation is to replace d with a new difference value d_, which is the sum of the embedded value and the lower bound of Rk. Finally, an inverse calculation from d_ is performed to generate the new gray values of the two pixels in the block. Note that the new gray values of the two pixels must lie in between the range [0, 255]. Therefore, if the new gray values are created by value uk, which are the maximally possible value of d_, falling outside the range [0,255], the block must be abandoned for embedding data. In the extracting phase, the secret data are extracted from the blocks of the stego-image in the same order as the embedding phase. The number of secret bits to be embedded in a two-pixel block is determined by the range Rk, which is the range of the difference value between two pixels. In addition, the value of the embedded data in the block is calculated by subtracting the lower bound of Rk from the difference value of the block. Therefore, the embedded bits in the block can be reconstructed. To verify the security of the proposed method, authors apply statistic steganalytic technique which is called RS-diagram, in order to prove that the method is undetected. The results are shown in Fig. 1.

3. OUR APPROACH

In this section, steganographic scheme based on blockwise embedding. We use the idea of pixel-value differencing, but we process four pixels simultaneously in spite of two at a time. In the pixel-value differencing approach[5], each time two pixels are grouped for embedding secret data. Fig. 2(a) shows the only grouping result of Wu and Tsai's method for a four pixel block. However, as shown in Fig. 2(b), there are three kinds of possible grouping results. In order to embed data more efficiently, we have considered different grouping results in our approach. Moreover, new techniques are proposed in order to avoid the additional information needed for recording the selected grouping data. The grouping, embedding, and extracting procedures of our approach are described in the following subections.

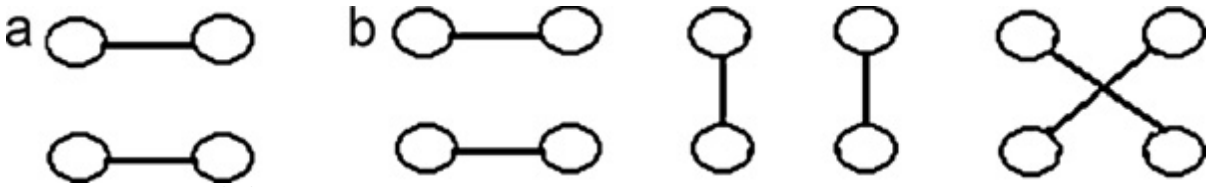


Fig:-2 Grouping results of a four-pixel block: (a) the only grouping result (b) all possible grouping results.

3.1 Pairwise grouping of a block

The host images used in our scheme are 256 gray value. Two difference values d1 and d2 are computed from each non-overlapping block with four neighbor pixels, say pi,j, pi,j+1, pi+1,j+1, and pi+1,j, of a given host image. The strategy of partitioning the host image into four-pixel blocks is to run through all the rows in a raster scan. The four pixels pi,j, pi,j+1, pi+1,j+1, and pi+1,j are then renamed as p1, p2, p3, and p4, and their corresponding gray values g1, g2, g3, and g4 satisfy the condition g1 ≤g2 ≤g3 ≤g4. The four-pixel block is partitioned into two two-pixel groups (g1, g4) and (g2, g3). The group which belongs to pi,j is defined as group1, and the other is defined as group2. In our scheme, the two group differences are computed as (g4 −g1) and (g3 −g2). The group difference of group1 is denoted as d1, and the group difference of group2 is denoted as d2. The difference value di (where i = 1 or 2) may be in the range from 0 to 255.

3.2 Data embedding

The secret message can be seen as a long bit stream. The task is to embed this stream into the four pixels block. The number of bits which can be embedded varies and is decided by the width of the range to which the difference values of the block belongs. Let the group difference di (where i = 1, 2) falls into the range of index ki. Then, the number of bits, say ni, to be embedded in ith group is calculated by ni = log2(uki – lki + 1). We embed n1 bits and n2 bits into group1 and group2, respectively. Let S1 and S2 be the bit streams of the secret message to be embedded, where |S1| = n1 and |S2| = n2. The two new differences d_ 1 and d_ 2 can be computed by:-

d_ I = lki + bi, where i = 1, 2.

In the above equation, b1 and b2 are the decimal values of S1 and S2, and are embedded into group1 and group2, respectively. By replacing d1 and d2 with d_ 1 and d_ 2, respectively, an inverse calculation from d_ 1 and d_ 2 generates the new pixel values g_ 1, g_ 2, g_ 3, and g_ 4 of the pixels p1, p2, p3, and p4. For the following two reasons, the pixel values g_ 1, g_ 2, g_ 3, and g_ 4 need to be modified further. One is that all values g_ 1, g_ 2, g_ 3, and

g_ 4 must fall into the range [0, 255]. The other is that the two groups need to be distinguished after embedding. The key point in order to successfully distinguish between two groups is to maintain the two intervals [g_ 1, g_ 4] and [g_ 2, g_ 3] such that one of them contains the other.

3.3. Data extracting

The process of extracting the embedded message is similar to the embedding process with the same traversing order of visiting all blocks. First,were name the pixels of a block as (p1, p2, p3, p4) and distinguish group1 and group2. Then, all values of di (i = 1, 2), ki (i=1, 2), and ni (i = 1, 2) of this block are found. Note that the calculation of these values is the same as the embedding process, except that the block now comes from stego-images. The bit-stream values b1 and b2, which are embedded in this block are then extracted using the following equations:

bi = di – lki , where i = 1, 2

Finally, each value bi is transformed into a bit stream with ni bits.

4. EXPERIMENTAL RESULTS

(a)

(b)

(c)

Fig:-3. Two of the experimental host images with size 512×512: (a) Peppers; (b) Baboon (c) Leena.

(a) (b)

Fig:-4. Two secret images with size 256×256: (a) Boat; (b) Airplane.

In our experiments, we used two host images with size 512×512. These are shown in Fig. 3. Two sets of widths, which partition the range [0, 255], are used in the experiments. The first experiment selects the widths of 8, 8, 16, 32, 64, and 128, which partitions the range [0, 255] into ranges [0, 7], [16, 31], [32, 63], [64, 127], and [128, 255]. The second experiment is based on the widths of 16, 16, 32, 64, and 128. We used a random bit stream, images "Boat" and "Sailboat" as separate secret messages in the experiments. Images "Boat" and "Airplane" are 256×256 as shown in Fig. 4. If a random bit stream is used as secret data, then the values of PSNR and capacities are the averages of the results obtained by 100 times executing the different random bit streams.

Table 1:- The capacities and PSNRs for embedding random bit stream, image Boat, and image Sailboat by our approach.

Cover images (512 × 512)		Widths of 8, 8, 16, 32, 64, and 128			Widths of 16, 16, 32, 64, and 128		
		Random bits	Boat	Airplane	Random bits	Boat	Airplane
Lena	Capacity	410,854	410,854	410,854	528,966	528,966	528,966
	PSNR	40.54	41.24	41.21	36.75	37.38	37.23
Baboon	Capacity	482,515	482,515	482,515	559,222	559,222	559,222
	PSNR	34.67	35.31	35.21	34.30	34.46	34.32
Peppers	Capacity	408,281	408,281	408,281	528,791	528,791	528,791
	PSNR	40.47	41.19	41.05	36.83	37.38	37.27

Table 2:- The results of our approach using the widths of 8, 16, 32, 64, 128, and 8.

Our scheme		Cover images		
		Lena	Baboon	Peppers
Ranges	Capacity	432,333	524,785	428,909
8–16–32–64–128–8	PSNR	38.86	33.31	39.15

The capacities and PSNR of embedding the results of two sets of widths are shown in Table 1. The capacities of the second experiments are higher than that of the first experiment. This is due to the fact that the second experiment uses larger widths. Consequently, the values of PSNR of the second experiment are worse.

5. ANALYSES AND DISCUSSIONS

There are three conclusions shown in this section. First, our approach finds out more edge areas for various images. In other words, more secret could be embedded into the edge areas by means of our scheme. Second, our approach can avoid the conditions of falling out of the boundary. We propose the skill of pixel-value shifting to shift the pixel values. It can solve the conditions of falling out of the boundary, and hence increase the embedded messages shows the amounts of pair wise pixels falling out the boundary . Although the amounts of falling out of the boundary are not large in this method, however, our approach guarantees that all pair wise pixels are usable. Finally, there is a partition scheme of range, but it is not practical. In their partition, the later ranges are larger for the reason that edge areas can tolerate larger distortions. Therefore, their partition of range is reasonable. We suggest a more practical partition with widths 8, 16, 32, 64, 128, and 8. The experimental results are shown in Table 2 hence capacities increase 17,344–42,270 bits and the PSNR values are still accepted by human eyes.

6. CONCLUSION

This Paper presents the technique for embedding the data keeping human vision in mind. Here I use blockwise approach in which I process four pixel block simultaneously. According to pixel value differncing, if a block in a sharp area then embedding capacity will increases. Therefore, by this approach large amount of secrete data can be embedded in host image.

7. References

[1]. Ker, A.D., 2007. Steganalysis of embedding in two least-significant bits. IEEE Transactions on Information Forensics and Security 2 (1), 46–54.

[2] .Lee, Y.K., Chen, L.H., 2000. High capacity image steganography. IEE Proceedings on Vision Image and Signal Processing 147 (3), 288–294.

[3].Li, X., Yang, B., Cheng, D.F., Zeng, T.Y., 2009. A generalization of LSB matching. IEEE Signal Processing Letter 16 (2), 69–72.

[4].Liu, J.C., Shih, M.H., 2008. Generalizations of pixel-value differencing steganography for data hiding in images. Fundamenta Informaticae 83 (3), 319–335.

[5].Wang, C.M., Wu, N.I., Tsai, C.S., Hwang, M.S., 2008. A high quality steganographic method with pixel-value differencing and modulus function. Journal of Systems and Software 81 (1), 150–158.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 322 - 326, 2015, ISSN:- 2319–8656

Multi Objective Segmentation for Vehicle License Plate Detection with Immune-based Classifier: A General Framework

Musab Mohammed Bagabir
College of Computer Science and Information Technology,
Sudan University for Science and echnology,
Khartoum, Sudan

Siti Mariyam Shamsuddin,
UTM Big Data Centre,
Universiti Teknologi Malaysia,
Johor, Malaysia

Mohammed Elhafiz
College of Computer Science and Information Technology,
Sudan University for Science and Technology,
Khartoum, Sudan

Ali Ahmed
Faculty of Engineering,
Karay Universit,
Khartoum, Sudan

Abstract: Vehicle License Plate Recognition (VLPR) is an important system for harmonious traffic. Moreover this system is helpful in many fields and places as private and public entrances, parking lots, border control and theft control. This paper presents a new framework for Sudanese VLPR system. The proposed framework uses Multi Objective Particle Swarm Optimization (MOPSO) and Connected Component Analysis (CCA) to extract the license plate. Horizontal and vertical projection will be used for character segmentation and the final recognition stage is based on the Artificial Immune System (AIS). A new dataset that contains samples for the current shape of Sudanese license plates will be used for training and testing the proposes framework.

Keywords: Multi Objective Particle Swarm Optimization; Artificial Immune System; Vehicle License Plate Recognition; Connected Component Analysis

1. INTRODUCTION

Therefore, many control and surveillance applications have been used VLPR Vehicle license plate recognition (VLPR) is an important component for automating many control and surveillance systems, such as road traffic monitoring, parking lots, access control, highway electronic toll collection, red light violation enforcement, finding stolen cars, gathering traffic flow statistics [1].

For the last few decades and due to the difference of license plates in formats, styles, colors and size from country to others the field of VLPR and its application has attracted many researchers in many countries to search and develop systems in order to identify their own vehicles license plate numbers. So far, many methods have been proposed for VLPR depending on the country's license plate characteristics.

VLPR is a technique that involves image processing technology and computer vision. Recognition/identification algorithms are generally composed of four major parts; pre-processing, license plate localization/detection, character segmentation and character recognition, and each part may contain several steps.

A considerable amount of literature has been published on VLPR, some of the related work is as follows: Saqib Rasheed, Asad Naeem and Omer Ishaq [2] proposed a technique of automated number plate recognition, which using canny edge detection operator and Hough lines for license plate localization, and template matching in recognition part.

In the recognition system proposed by Alginahi [3] for Saudi Arabian's license plate, in pre-processing, the median filter is used to remove noise, and then the sobel detector is applied. As a next step, the license plate is detected by searching the vertical lines, then the width to length ratio is calculated if more than two vertical are detected. Character segmentation is then performed by using the 8-connected components technique to find all the components (characters/numerals) on the plate. Then the horizontal projection profiles and zoning is used as features extractor. And finally to recognize the plate both a Mahalanobis distance classifier and a Multilayer Perceptron Neural Network classifier are used.

In [4] the Saudi Arabian's license plate detection depends on a black cross that centers the plate, so an edge detector is applied to find the horizontal and vertical maps, then before median filter performed, the binary image is obtained by using the average value of pixels in each map as a threshold. After the numbers and letters were segmented the recognition part is performed using a template matching.

In [5] Radial Basis Function (RBF) neural network was used both for the detection and recognition of Libyan's license plate. The pre-possessing phase starts with obtain the binary scale image, performing edge detection using sobel's mask operator, performing dilation process, using 'flood fill' algorithm to fill the interior gaps, applying the filtering task, and finally the image was smoothened by eroding it twice. The character segmentation phase is based on thresholding and CCA. A matrix of size 4 x 2 contains a character feature, which is obtained by dividing a character image into a sequence of horizontal "scan lines" by using the raster scanning.

In [6] the authors used Canny edge detection operator to locate the license plate in the image, the binary large object (blob) coloring algorithm is used in character segmentation part, the feature vector of a character image was encoded by

using Average Absolute Deviation algorithm, in the classification phase a multi layered perceptron artificial neural network model was used , in addition to, the numbers and the letters were classified by using two separate ANN.

According to image segmentation and the multi-level thresholding problem in image segmentation, many techniques are proposed using Particle Swarm Optimization (PSO) Algorithm such as in [7,9].

In [7] an approach on the basis of the PSO with wavelet mutation is used. The optimization of the multi-level thresholds for the images is performed by maximizing the total entropy of the image. In [8] the algorithm is used PSO to maximize the entropy criterion (Kapur's) and between-class variance (Otsu's) objective functions.[9] presents a hybrid of PSO and Genetic algorithm for multi-level thresholding base on maximum entropy criterion .

Other techniques are used to solve the multi-level thresholding problem in image segmentation as Genetic Algorithm in [10], Hill Climbing in [11], and Artificial Bee Colony (ABC) Algorithm in [12] which approximates the 1-D histogram of an image by calculating the Gaussian mixture model parameters.

In the recognition and classification phase, AIS is the quite popular and active research topic in recognition domain. As [13] and [14] describe the implementation of AISs in solving an image classification problem. In [13] the clonal selection algorithm proposed in a shape recognition problem beside hamming distance to calculate the affinity. In [14] the authors proposed a handwritten character recognition algorithm based on AIS by using clone selection principle .

While this study will focus on Sudanese vehicles license plates, each country has its own rule on vehicle license plate designing. The following section will give some important information about the type, shape and design of Sudanese vehicles license plates.

The paper is organized as follows. The next section constitutes an introduction about the Sudanese vehicle license. Section III introduces the Multi-objective optimization and Multi-objective Particle Swarm optimization methods .The complete framework is presented in details in Section IV. Section V present a discussion, the conclusion is presented in Section VI.

2. SUDANESE VEHICLE LICENSE PLATE

The Sudanese vehicles license plates are categorized in a number of types, that categorization was based on the differences of plates background color and characters color, some of types are mentioned in Table 1 below:

Table 1. Sudanese Vehicles Plates Types

Type	Background Color	Characters Color
Private vehicles	White	Black
Commercials (Passenger)	White	Green
Commercials (Goods)	Black	White
Police	Blue	White
Government	Yellow	Black

The size of all plate types is 32 ×16 centimetres (See Figure1). The plate has been divided into three regions; one region in the upper part, that contains the name of the country 'SUDAN' written in English and Arabic. The other two regions in the lower part, which divided by a silver metallic bar, one on the right this region contains numerals (1 to 5 numbers) written in English and Arabic, and the other region on the left side contains characters or character and number written in English and Arabic, the characters are an abbreviation of Sudan states' names, and the number to keep the sequence of the numbering. This study will focus on the first type: private vehicles as show in Figure 1.

Figure 1. Sudanese License Plate

3. MULTI OBJECTIVE PARTICLE SWARM OPTIMIZATION

Multi Objective Optimization (MOO) also known as multi criterion extends the optimization theory by permitting several design objectives to be optimized simultaneously, the goal is to find a set of values for the design variables that simultaneously optimizes several objective (or cost) functions [16]. In mathematical terms, A MOO problem can be formulated as :

$$\text{Minimize } \vec{f}(\vec{x}) := [f_1(\vec{x}), f_2(\vec{x}), \ldots . f_k(\vec{x})] \quad (1)$$

Subject to the constraints

$$g_i(\vec{x}) \le 0, \quad i = 1, 2, \ldots, m \quad (2)$$

$$h_j(\vec{x}) = 0, \quad j = 1, 2, \ldots, p \quad (3)$$

where $g_i(\vec{x}) = [x_1, x_2, \ldots, x_n]^T$ is the vector of design variables, $f_i : \Re^n \to \Re$, i = 1, ..., k are the objective functions and $g_i, h_j : \Re^n \to \Re$, i = 1, ..., m, j = 1, ..., p are the constraint functions of the problem. The objectives (2) and (3) often conflict with each other. Improvement of one objective may lead to aggravation of another. Thus, a single solution which can optimize all objective simultaneously does not exist. As an alternative, the best trade off solutions is called the Pareto Optimal Solution.

There are different approaches to solving MOO problems such as aggregating, population based non-pareto and pareto-based techniques [20]. A comprehensive review of the

different approaches to solving MOO problems can be found in [20].

Particle Swarm Optimization (PSO) is developed by Kenney and Eberhart in 1995 [17], it's a population-based metaheuristic inspired on the social behavior of birds within a flock. In a PSO algorithm each potential solution to the problem is called a particle and the population of solutions is called a swarm. The way in which PSO updates the particle x_i at the generation *t* is through the formula[22]:

$$x_i(t) = x_i(t-1) + v_i(t) \quad (4)$$

where the factor $v_i(t)$ is known as velocity and it is given by

$$v_i(t) = w * v_i(t-1) + C1 * r1 * (x_{pbest_i} - x_i) + C2 * r2 * (x_{gbest_i} - x_i) \quad (5)$$

In this formula, x_{pbest_i} is the best solution that x i has viewed, x_{gbest_i} is the best particle (also known as the leader) that the entire swarm has viewed, w is the inertia weight of the particle and controls the trade-off between global and local experience, r1 and r2 are two random numbers within the interval [0, 1], and C1 and C2 are specific parameters which control the effect of the pbest and gbest particles.

The simplicity of PSO have made it a popular optimization approach, and a good candidate to be extended for multi objective optimization (MOO) [21]. This technique called Multi Objective Particle Swarm Optimization (MOPSO) has been introduced in [21] to deal with MOO problems. More details about MOPSO can be found in [20, 21]

Image segmentation based on thresholding techniques can be seen as a multi objective problem. Where, there is no single threshold objective function (criterion) able to produce an optimal thresholds values for all kinds of images [15, 23]. The use of MOPSO aims to obtain a good segmentation results on more kinds of images. Consequently, the segmentation problem can be formulated mathematically as:

Minimize/Maximize

$$f_m(t_1, ..., t_{N-1}) \quad \text{m = 1, 2, ..., M} \quad (6)$$

Subject to

$$0 < t_1 < t_2 < ... < t_{N-1} < L \quad (7)$$

Where *M* is the number of objective functions (criterion) used for the segmentation, t_i the segmentation thresholds, and L the number of gray levels.

In our work, two broadly used optimal thresholding methods namely entropy criterion [19] method and between-class variance [18] method will be used as our objective functions.

4. THE FRAMEWORK

The proposed framework is designed for Sudanese vehicle license plate recognition. This system is composed of a number of stages as shown in figure 2 are pre-processing that includes the MOPSO, Plate Detection/Localization, Character Segmentation and AIS Recognizer/Classifier.

The input of the system is the original image of the vehicle in RGB scale of size 769×559 pixels taken from different distances and Angles. The details of other stages are presented in the following subsections.

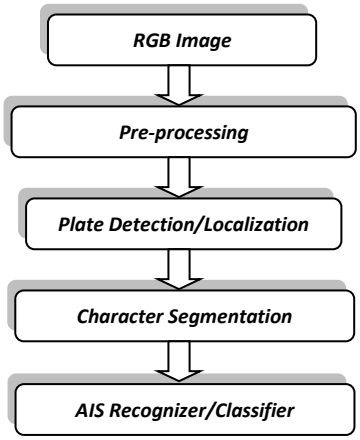


Figure 2. The Proposed Framework Diagram

4.1 Pre-processing

Pre-processing stage plays an important role, which influences the accuracy of the plate detection stages significantly. In this study the first step in pre-processing is the RGB image conversion into gray scale, as the gray scale decreases the computational time; the RGB image consists of three channels red, green and blue, the value of each channel in the range 0-256, whereas the gray scale image contains only one channel. Then, in this framework we apply median filtter to remove noises like random occurrences of black and white pixels, then the image contrast will be adjusted by using histogram equalization technique.

Image segmentation is a process of partitioning a digital image into multiple segments, so it aims to partitioning an image into regions in order that each region groups adjacent pixels having similar attributes as intensity. Multi-level thresholding is one of the most important image segmentation techniques, which decrease the number of intensity levels. The basic idea of multi-level thresholding technique is to divide the pixels into several groups based on a certain number of threshold values. The reason behind the adoption of the multi-level thresholding is to clearly separate the plate region in the image from the background. Therefore, the automatic selection of optimal thresholds has remained a challenge in image segmentation [15].

In this framework, we propose a new image segmentation technique based on multi objective particle swarm optimization MOPSO, that combines the flexibility of multi objective fitness functions with the power of PSO for searching vast combinatorial state spaces, in order to find the optimal thresholds.

4.2 Plate Detection/Localization

The aim of this stage is to extract the license plate from the gray scale image. Therefore the proposed framework provides more importance to this stage whose success guarantees the successful performance of the VLPR system.

As a first step in this stage, the gray scale image resulted by the previous stage will be converted to binary image using the well-known Otsu global thresholding method [18]. Other Other morphological operations may also be used in order to isolate the plate from the background.

The resultant binary image will be labeled (Connected Components Labeling-CCA) in order to identify objects/regions in the binary image (each region of connected pixels is called connected component). Then CCA will be used to analyze the regions independently and measure the property of each region. Since the Sudanese license plate have a unique features such as the following: rectangular shape, the ratio of the width to height of the rectangle is fix and the area of the rectangle, this features will be useful to identify the plate region.

4.3 Character Segmentation

Character segmentation is the procedure of extracting and isolating the characters and numbers (each character and number) on the license plate image. At this phase, the English characters and Arabic numbers would be considered as output target in this phase. The study will focus on segmenting the English characters and Arabic numbers as shown in figure 3.

Figure 3. English characters and Arabic numbers with in Sudanese License Plate

At the beginning of character extraction phase and for accurate character segmentation first need is to remove skew from the extracted license plate image. The whole idea of skew correction depends on finding the angle of rotation, and then the extracted license plate image is rotated using that angle of rotation. Then some enhancement techniques will be used as well as morphological operations in order to overcome noise and light variance problems on extracted license plate image.

Then the horizontal projection and vertical projection techniques were proposed to perform the character segmentation process [3][4]. First removing the upper part that containing the words 'SUDAN السودان', then dividing the plate into two images; character on left part(left image) and numbers on right part(right image). Then each image is segmented into isolated characters and numbers after removing the lower part form the right image and upper part from the left image. After this phase all the identification contents of the license plate were prepared for the recognition phase.

4.4 AIS Recognizer/Classifier

The final phase in the license plate recognition process is character recognition. Although there are many techniques presented and applied for character recognition, in this work, character recognition is performed based on AIS [14].

First, all segmented characters images will be normalized and resized to a fixed length and width. The character recognition will carried out using Clonal Selection Algorithm [13].

In our proposed AIS the segmented characters and templates of English character and Arabic numbers will considered as antigens and antibodies respectively. The affinity will be calculated using Hamming Distance.

5. DISCUSSION

The data set that will be used in the experiments contains 200 color images with a size of 640 × 480 pixel, acquired from Sudanese Traffic Police. The tested images have been acquired from the front of vehicles under various illumination and weather conditions (sunny, cloudy ,daytime, nigh time, rainy days...etc).

The recognition system receives the image of the target vehicle in RGB format, the meadian filter will be used after the image is converted into gray scale to remove noise.

By performing MOPSO technique on the gray scale image the plate region will be clearly separated from the background, then before applying morphological operations the image is converted to black and white image.

The plate region in the resultant image is determined by applying CCA on the binary image, and the outputs of this process are rectangular shaped regions. In order to identify the plate region, certain features, which are unique to plate, such as mentioned earlier will be used. The final output of plate detection phase is the license plate coordinates, which are used to crop the license plate from the original gray image.

The character segmentation phase receives the extracted vehicle plate image as input. A pre-processing step is required to improve and enhance the plate image, such as skew correction and noise removal. Then horizontal projection and vertical projection techniques are used for segmentation and extraction process. The outputs of this phase are images of each English character and each Arabic number within the license plate.

After the individual characters are extracted, these extracted characters will be passed to the AIS as antigens for classification and recognition task.

6. CONCLUSION

The purpose of this paper is to presents a framework for an automatic vehicle license plate recognition system. It describes the proposed framework that will be use to recognize the Arabic numbers and English characters within the Sudanese private vehicle license plates. According to the state of the art of VLPR system a suitable collection of that techniques and methods has been chosen to implement the relevant part of this the proposed framework: MOPSO and CCA are chosen for plate detection and extraction; horizontal and vertical projection are found suitable for segmentation the plate character; AIS for character recognition. The proposed framework is expected to be able to succeed in recognizing the plates efficiently and accurately.

7. REFERENCES

[1] Hu, H., Zhang, Z., & Bai, Y. (2012). Car License Plate Location Based on Mathematical Morphology. In Recent Advances in Computer Science and Information Engineering . Springer Berlin Heidelberg: 415-420.

[2] Rasheed, S., Naeem, A., & Ishaq, O. (2012). Automated Number Plate Recognition using hough lines and template matching. In Proceedings of the World Congress on Engineering and Computer Science Vol. 1: 24-26.

[3] Alginahi, Y. M. (2011). Automatic arabic license plate recognition. International Journal of Computer and Electrical Engineering 3.3: 454-460.

[4] Basalamah, S. (2013). Saudi License Plate Recognition. International Journal of Computer & Electrical Engineering 5.1.

[5] Abulgasem, Nureddin A., Dzulkifli Mohamad, and Siti Zaiton Mohamad Hashim. (2011). Automatic License Plate Detection and Recognition Using Radial Basis Function Neural Network. International Journal of Computer Vision and Applications (IJCVA) 1.1.

[6] Erdinc Kocer, H., & Kursat Cevik, K. (2011). Artificial neural networks based vehicle license plate recognition. Procedia Computer Science 3: 1033-1037.

[7] Jiang, F., Frater, M. R., & Pickering, M. (2012). Threshold-based image segmentation through an improved particle swarm optimisation. In Proceedings of the International Conference on Digital Image Computing Techniques and Applications (DICTA):1-5. IEEE.

[8] Duraisamy, S. P., & Kayalvizhi, R. (2010). A new multilevel thresholding method using swarm intelligence algorithm for image segmentation. Journal of Intelligent Learning Systems and Applications 2.3: 126.

[9] Baniani, E. A., & Chalechale, A. (2013). Hybrid PSO and Genetic Algorithm for Multilevel Maximum Entropy Criterion Threshold Selection. International Journal of Hybrid Information Technology 6.5.

[10] Banimelhem, O., & Yahya, Y. A. (2011). Multi-Thresholding Image Segmentation Using Genetic Algorithm. In World Congress in Computer Science, Computer Engineering, and Applied Computing.

[11] Nath, S., Agarwal, S., & Kazmi, Q. A. (2011). Image histogram segmentation by multi-level thresholding using Hill climbing algorithm. International Journal of Computer Applications 35.1.

[12] Cuevas, E., Sención, F., Zaldivar, D., Pérez-Cisneros, M., & Sossa, H. (2012). A multi-threshold segmentation approach based on Artificial Bee Colony optimization. Applied Intelligence 37.3: 321-336.

[13] Isa, N., Sabri, N. M., Jazahanim, K. S., & Taylor, N. K. (2010). Application of the Clonal Selection Algorithm in artificial immune systems for shape recognition. In Proceedings of the International Conference on Information Retrieval & Knowledge Management,(CAMP) : 223-228. IEEE.

[14] Chen, Y., Liang, C., Yang, D., Peng, L., & Zhong, X. (2010). A handwritten character recognition algorithm based on artificial immune. In Proceedings of the International Conference on Computer Application and System Modeling (ICCASM) Vol. 12 : V12-273. IEEE.

[15] Duraisamy, S. P., & Kayalvizhi, R. (2010). A new multilevel thresholding method using swarm intelligence algorithm for image segmentation. Journal of Intelligent Learning Systems and Applications 2.3: 126.

[16] Nakib, A., Oulhadj, H., & Siarry, P. (2010). Image thresholding based on Pareto multiobjective optimization. Engineering Applications of Artificial Intelligence 23.3 : 313-320.

[17] James, K., & Russell, E. (1995). Particle swarm optimization. In Proceedings of International Conference on Neural Networks :1942-1948. IEEE.

[18] Otsu, N. (1975). A threshold selection method from gray-level histograms.Automatica, 11(285-296): 23-27.

[19] Kapur, J. N., Sahoo, P. K., & Wong, A. K. (1985). A new method for gray-level picture thresholding using the entropy of the histogram. Computer vision, graphics, and image processing 29.3 : 273-285.

[20] Reyes-Sierra, M., & Coello, C. C. (2006). Multi-objective particle swarm optimizers: A survey of the state-of-the-art. International journal of computational intelligence research 2.3 : 287-308.

[21] Coello, C. A. C. (2011). An introduction to multi-objective particle swarm optimizers. In Soft Computing in Industrial Applications : 3-12 . Springer Berlin Heidelberg.

[22] Durillo, J. J., García-Nieto, J., Nebro, A. J., Coello, C. A. C., Luna, F., & Alba, E. (2009). Multi-objective particle swarm optimizers: An experimental comparison. In Evolutionary Multi-Criterion Optimization : 495-509 . Springer Berlin Heidelberg.

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 327 - 330, 2015, ISSN:- 2319–8656

Preventing Disclosure Attacks by Secured Traffic Aware Protocol in Manets

S.Brindha Devi
Department of CSE
Panimalar Institute of Technology
Chennai,India

A. Porselvi
Department of CSE
Panimalar Institute of Technology
Chennai,India

Abstract: In this paper we propose a system that allows a safe and secure data transfer in MANETs between the source and the destination. As MANETs are unplanned networks and networks of instant communication, they are prone to attacks like disclosure, brute force attacks etc. In this paper we mainly concentrate on limiting the disclosure attacks in MANETs. Disclosure attack means that the network is monitored quietly without modifying it. The monitoring of network is possible only if the traffic is known. Hiding of traffic between the source and destination would prevent disclosure attacks in MANETs. To hide the traffic between the source and destination we must identify it. The traffic is identified using STARS(Statistical Traffic Pattern Discovery System for MANETs) technique. Using this technique, the traffic is made observable only for the intermediary nodes and the data is sent via intermediary nodes to the destination as single hop. The data which is sent as single hop by hop via intermediary nodes prevents the malicious node from knowing the original source and destination and thus preventing MANETs from disclosure attack.

1. INTRODUCTION

As networking is becoming an increasingly important technology for both military and commercial applications, security is an essential
requirement. MANETs are one of the network which is been widely used in the military environments. MANETs are used to start a instant communication between a source and the destination. MANETs are vulnerable to security attacks due to the lack of trusted authority and limited resources.

In mobile wireless networks there is no infrastructure, so it becomes even more difficult to efficiently detect the malicious activities that takes place inside the network. Because of this it becomes easy for the malicious nodes to flood the network withjunk packets or reveal the information by monitoring the network.

MANETs are usually prone to routing attacks because of their dynamic topology and less infrastructure .Attacks on network is of two categories

1.Active attacks.

2.Passive attacks.Active attacks are those which disrupt the normal functionality of MANETs such as doing data interruption, modification etc. Example: DoS, jamming. Passive attacks are those that do not disturb the functionality of MANETs but obtains the data that has been exchanged in the network. Example: traffic analysis, monitoring.Here, we majorly deals with passive attacks on MANETs such as traffic analysis and disclosure attacks. Traffic analysis in the MANETs are nothing but identifying the communication parties between whom and whom the communication is taking place and also finding their functionalities. This attacks takes place on data link layer of the network. This attacks can be prevented by encryption. But still nodes should continuously monitor time to time and look up for the malicious nodes to prevent from their misbehavior.

The goal of this paper is to prevent disclosure attacks in MANETs. Here in this attack, the malicious nodes
which is present as legitimate nodes in a network leaks confidential information to unauthorized nodes in the network. The best way to overcome this attack is by secure data transmission. For this secure data transmission, we use AOMDV protocol. This protocol discovers multiple path in a single route discovery and also uses hop- by- hop routing approach which hides the information about the actual source and the destination.

2. RELATED WORK

As the network evolves and the components in the network becomes bigger and bigger, network security becomes one of the important factor to be considered .By increasing the network security we can decrease the chance of piracy, spoofing , information theft etc.

In paper [1] proposed by David L.Chaun, they used a public key cryptography technique to hide the participant who is communicating with whom in a e- mail system.

In paper [2], Michael K.Reiter and Aviel D.Rubin introduced a system called CROWDS, where the users are grouped together to form a large group so that the web servers are unable to learn the true source of request.

In paper [3] Michael G.Reed, Paul F.Syverson proposed onion routing technique where data are encapsulated as layers to provide secure communication over public network.

In paper [4] Azzedine Boukreche, Khalil El-khalib, Lary Korba proposed a distributed routing protocol which makes sure that only the trustworthy intermediary nodes participate in the communication between source and the destination. In paper [5], Ronggong song, Lary korba, George Yee proposed a anonymous dynamic source routing
(AnonDSR) to provide user security.

In paper [6] Yanchao Zhang,Wei Liu proposed a anonymous

on- demand routing protocol termed as MASK which can accomplish MAC layer and Network layer communication without disclosing the real ID's.

In paper [7] S.Seys and B.Preneel proposed ARM(Anonymous Routing protocol) for MANETs that hides the routes.

In paper [8] Jiejun Kong, Xiaoyan Hong and Mario Gerla proposed two techniques namely identity-free routing and On - demand routing.

In paper [9] Reza Shokri, Maysam Yabandeh, Nasar Yazdani proposed a AODV protocol to provide sender and recipient relationship with least computational overhead.

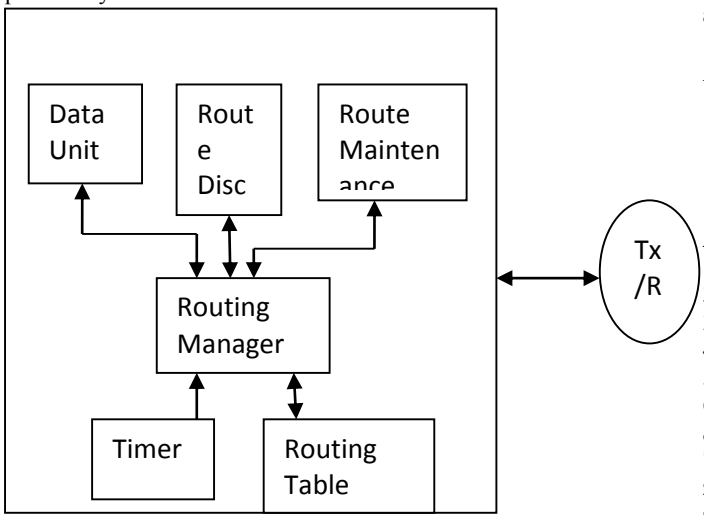
In paper [10] Xinyuan Wang, Shiping Chen and SushilJajodia proposed a water marking technique to identify the long flow between the original packet and anonymized waterflow.

3. PROPOSED WORK

To disclose the hidden traffic pattern in mobile networks, Statistical Traffic pattern discovery system (STARS) is used. STARS uses two major steps.

1.Point-to-point matrix to derive End-to-End matrix.

2. Analysing the end-to-end matrix to calculate the probability of each node to be a source or destination.



ROUTE DISCOVERY AND ROUTE MAINTAINENCE

In the intermediate node ,after receiving RREQ a reverse entry is set up by the node and it consists of all the details of it.when it reaches the destination and if both the IP address and the conditions are met then the RREQ gets response from the node to the source by means of unicasting RERR(route maintenance) is initiated by the node upstream (closer to the source) of the break, Its propagated to all the affected destinations. RERR lists all the nodes affected by the link failure Nodes that were using the link to route messages (precursor nodes). When a node receives an RERR, it marks itsroute to the destination as invalid Setting distance to the destination as infinity in the route table.When a source node receives an RRER, it can reinitiate the route discovery.

POINT-TO-POINT MATRIX:

With the captured point-to-point traffic(one hop) in a certain period T, we build up point-to-point traffic matrices. We construct this point to point matrices such that each traffic matrix contain only independent one hop packets. To avoid a single point-to-point traffic matrix from containing two dependent packets, we apply slicing technique. That is we take snapshots of the network and each snapshot is triggered by a captured packet. A sequence of snapshots during a time interval (Δtc) construct a slice represented by traffic matrix which is N*N one hop matrix. When calculating length of the time interval we consider two important criteria.

1. In the time interval node can either be sender or receiver.
2. Each matrix must represent one hop transmission during the time interval.

The time slicing technique which is done here is to make sure that packets captured in time interval are independent of each other. Each packet p in Wc(i,j) has three features P.vsize,P.time and P.hop. A packet hop count is set to 1.

W1=[0 1 0,0 0 0,0 0 0]

END-TO-END MATRIX:

From the given sequence of point-to-point matrix we derive end-to-end matrix R=(r(i,j)) N*N where r(i,j) is the accumulative traffic volume from node i to node j.

Algorithm 1. —f(W| 1*K).
1: R = W1
2: for e = 1 to K-1 do
3: R = g(R;We+1) + We+1
4: end for
5: return R

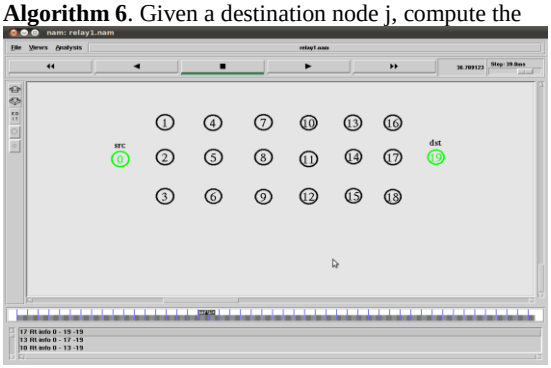
Algorithm 2. —g(R,We+1)
1: R'= R
2: for i = 1 to N do
3: for k = 1 to N and k ≠ i do
4: for j =1 to N do
5: for each x € we+1(j, k).pkt do
6: if y € r(i, j).pkt s.t. x.time – y.time < T and y.hop < H then
7: create z with z.time =x.time
z.hop =y.hop + 1
z.vsize = min{x:vsize, y:vsize}
8: r'(i, k).pkt = r'(i; k).pkt U {z}
9: r'i; k) = r'i, k) + z.vsize
10: end if
11: end for
12: end for
13: end for
14: end for
15: return R'

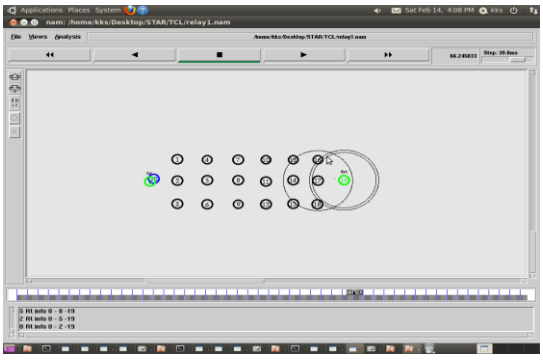
Algorithm 3. —Src(R).
1: S0 = (1/N, 1/N, . . . , 1/N)
2: n = 0
3: do
4: Sn+1 = (¢(R) . ¢ T(R)) . Sn
5: normalize Sn+1
6: n =n + 1
7: while Sn ≠Sn-1
8: S =Sn
9: return S

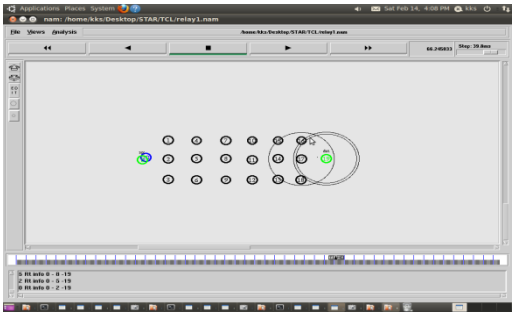
Algorithm 4. —Dest(R) .

1: D0 =(1/N, 1/N; . . . , 1/N)
2: n = 0
3: do
4: Dn+1 = (¢T(R). ¢ (R). Dn
5: normalize Dn+1
6: n = n + 1
7: while Dn ≠ Dn-1
8: D = Dn
9: return D
Algorithm 5.

Given a source node i, compute the
probability distribution vector Ls-d(i) for each node
to be the intended destination of i.
1: R = f(W|1 *K);
2: D = Dest(R);
3: W '|1*K = Suppress-Sender(i);
4: R0 = f(W '|1*K);
5: D-= Dest(R');
6: Calculate the probability reduction vector as:
L's-d(1)= D -D-. If negative elements exist in L's-d (1)
increase each element by the absolute value of the
smallest negative element;
7: Normalize L's-d (1) to generate the probability vector
L's-d (1) for each node to be the intended destination of i;
8: Return L's-d (1)
Algorithm 6. Given a destination node j, compute the







probability distribution vector Ld-s (j)for each node to be the corresponding source of j.
1: R =f(W|1*K);
2: S = Src(R);
3: W'|1*K = Suppress -Receiver(j);
4: R0 = fõW '|1 *K);
5: S- = Src(R ');
6: Calculate the probability reduction vector as:
L'd-s (j)S – S-. If negative elements exist in L'd-s (j)increase each element by the absolute value of the
smallest negative element;
7: Normalize L'd-s (j) to generate the probability vectorLd-s (j) for each node to be the corresponding source of j;
8: Return Ld-s (j)
Once we find the hidden traffic pattern between the source and destination using STARS we hide the traffic from the malicious nodes. This is done by making the nodes in the network observable or visible only for trustworthy nodes. The data is send from source to destination in a hop by hop manner. That is the information about the actual source and destination is hidden in the routing table. The routing table consists of source and the next intermediary node as destination. After one hop the destination node becomes the source and the next intermediary node in the path that reaches the final destination become the next destination. In this manner the data sent reaches the final destination. Even if the malicious nodes are present in the actual traffic, it cannot track or disclose the traffic to the unauthorized nodes which in turn leads hacking of the data.

4. CONCLUSION

Here we use STARS which is generally an attacking system that discloses the traffic pattern to the malicious nodes. Using the above system we find the actual traffic and hide the traffic to the untrustable nodes. Thus, this helps in preventing disclosure attacks in MANET

5. REFERENCES

[1] D. Chaum, "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability," J. Cryptology, vol. 1, no. 1, pp. 65-75, 1988.
[2] M. Reiter and A. Rubin, "Crowds: Anonymity for Web Transactions," ACM Trans. Information and System Security, vol. 1, no. 1,pp. 66-92, 1998.
[3] M. Reed, P. Syverson, and D. Goldschlag, "Anonymous Connec- tions and Onion Routing," IEEE J. Selected Areas in Comm., vol. 16, no. 4, pp. 482-494, May 2002.
[4] A. Boukerche, K. El-Khatib, L. Xu, and L. Korba, "SDAR: A Secure Distributed Anonymous Routing Protocol for Wireless and Mobile Ad Hoc Networks," Proc. IEEE 29th Ann. Int'l Conf. Local Computer Networks (LCN '04), pp. 618-624,2004.
[5] R. Song, L. Korba, and G. Yee, "AnonDSR: Efficient Anonymous Dynamic Source Routing for Mobile Ad-Hoc Networks," Proc. Third ACM Workshop Security of Ad Hoc and Sensor Networks (SASN '05), pp. 33-42, 2005.
[6] Y. Zhang, W. Liu, W. Lou, and Y. Fang, "MASK: Anonymous On- Demand Routing in Mobile Ad Hoc Networks," IEEE Trans. Wireless Comm., vol. 5, no. 9, pp. 2376-2385, Sept. 2006.
[7] S. Seys and B. Preneel , "ARM: Anonymous Routing Protocol for Mobile Ad Hoc Networks,"
Proc. IEEE 20th Int'l Conf. Advanced Information Networking

and Applications Workshops
(AINA Work- shops ’06), pp. 133-137, 2006.
[8] J. Kong, X. Hong, and M. Gerla, “An Identity-Free and On-Demand Routing Scheme against Anonymity Threats in Mobile Ad Hoc Networks,” IEEE Trans. Mobile Computing, vol. 6, no. 8, pp. 888-902, Aug. 2007.
[9] R. Shokri, M. Yabandeh, and N. Yazdani, “Anonymous Routing in MANET Using Random Identifiers,” Proc. Sixth

International Journal of Computer Applications Technology and Research
Volume 4– Issue 4, 331 - 335, 2015, ISSN:- 2319–8656

Ontology based design for M- learning system

M.Thangaraj
Department of Computer Science
Madurai Kamaraj University, Madurai
Tamil Nadu, India

S.Vanathi
Department of Computer Science
Govt. Arts College, Melur
Tamil Nadu, India

Abstract: Information and communication Technology is a gateway through which large population of students has been addressed. Mobile learning technology the latest arrival highly changing the way the students learn, interact, access up to data information. It mainly satisfies the current and future generation which needs information at the earliest rather than later few touches. The World Wide Web acts as an interface in E- learning as well as in mobile learning (M-learning) environments. It supports and facilitates the delivery of teaching and learning materials. M - l e a r n i n g provides quality educational content with the help of semantic web technologies like Ontology. This study presents Mobile Learning framework for making efficient learning with a case study on cyber security.

Keywords: Information and Communication Technology; Ontology; M-learning; E-learning; cyber security.

1. INTRODUCTION

Education is a process which enables a person's holistic development of personality through knowledge acquisition. Learning environment has the influence on knowledge improvement or knowledge enhancement [8]. Teaching learning processes delivered through mobile devices are based on wired and wireless communication technologies. Mobile learning creates a new learning environment using handheld devices by interfacing world wide web and the learner [10].

Mobile learning, recent form of distant learning is an extension of E-learning application is shown in Figure 1, which also has the use of audio, visual, cognitive, cooperative and interactive contents delivered via smart digital electronic devices in an attempt to create a direct, dynamic, ongoing learning environment. Such form of learning enables the individual learner to move freely in the learning material, at the same time can access to knowledge sources wherever and whenever the learner desires.

In the world of education and training there exist many definitions for mobilE-learning. Among them ADL defines mobile learning or "M- Learning" as the use of handheld computing devices to provide access to learning content and information resources.

Thus M-learning overcomes the time and distance barriers of traditional learning and cost barrier of E-learning.

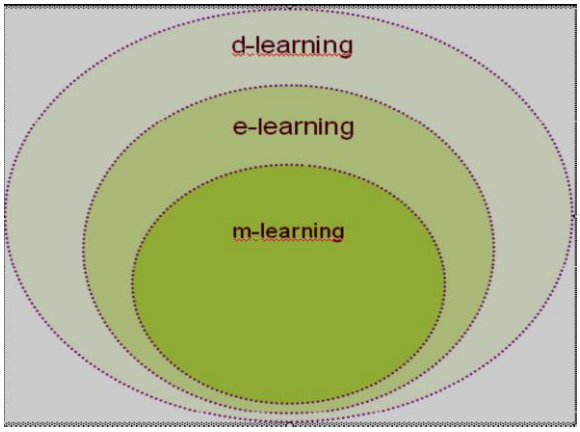


Figure 1. Various Dimensions of latest learning/ Evolution of distant learning

Ontology is a conceptual system a semantic account, which expresses a meta-level specification [5], and it owns the features like expressiveness, extensibility, interoperability, sharing and re-usability.

The various directions of research in ontology based M-learning are discussed as related work in section 2. Section 3, describes the Modified Intelligent Recommender System for Effective E-learning Environment Architecture, which encompasses various activities of M-learning in detail. Section 4 explains evaluation and experimental results on the efficiency of the system. Section 5, concludes and guides future work.

2. RELATED WORK

The changes in the educational scenario are wider and fast. Mobile learning is a personalized as well as interactive mode of learning. The implementation of mobile learning using android OS is illustrated in [3].

The meaning of mobile learning, its key issues, solutions and the knowledge transformation due to the development of mobile learning are clearly explained in [6].

E-learning fulfils the user context with the help of semantic web technology like ontology. Ontology based applications improve the understanding of concepts, critical thinking and creative thinking using the semantic learning environment is discussed [9].

IRS-EEE presents the effective E-learning environment using concept oriented, ontological content management. The essentials of semantic web are emphasized in [11].

The cost and portability while using laptops for E-learning are high. This paper tries to overcome the cost and time barriers.

3. MODIFIED IRS-EEE ARCHITECTURE

The Modified IRS-EEE Architecture (Figure 2) is based on the semantic structure, promises a powerful approach to satisfy the M-learning requirements. This Semantic architecture of M-learning has four layers such as User Layer (UL), Service Layer (SL), Content Management Layer (CML) and Database Layer (DBL).

The semantic based IRS-EEE architecture [11] is structured to satisfy E-learning requirements, of the user which is tailored to deliver the content through mobile devices result into the **modified IRS-Architecture** is an extension of the IRS-EEE. This architecture has four layers such as User Layer (UL), Service Layer (SL), Content Management Layer (CML) and Database Layer (DBL).

1. *User layer*

The two types of users of this layer are Provider and student .This system revolves around the users. The user profile is stored and updated whenever needed.

A. Data Provider

The role of provider in this system is to monitor user's profile, presenting the content, adding and updating the contents, questions, assessment of outcome and learner's performance and tracking the performance of the entire system. Provider has the facility to convert the web pages into compatible mobile device content.

B. Learner

The learner is the next stage user. The learners are the Persons/Students using advanced cellular phones or smart phones having the link to internet. Using such phones is regarded as more trendy, fashionable and prestigious among the young education by means of mobile devices is gaining efficiency in its design methods.

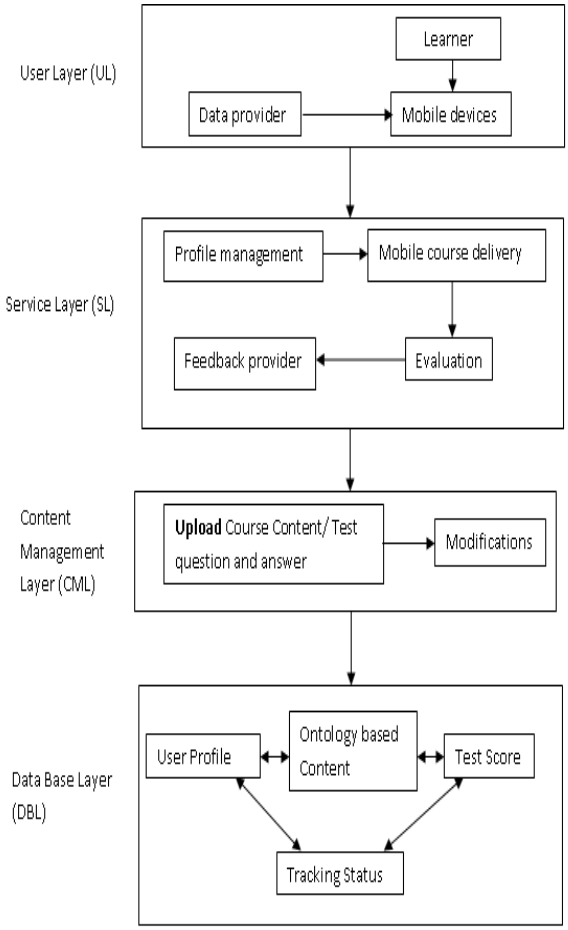


Figure 2. Modified IRS-EEE Architecture

C. Mobile devices

Typical examples of mobile devices include smart phones, tablets, laptops, and personal media players. We limit our implementation in smart phones and tablets only because they are highly portable (than laptop etc.,) and easy to handle. These mobiles devices have the property to produce unique educational affordances: portability, social interactivity, context sensitivity, connectivity and individuality [4**].**

2. Service Layer

A. Profile management

The function of the Profile management module is to maintain all the related information about the learners, and there interactions etc. it also, updates and keeps track of the consistency of data by proper updating process.

B. Mobile course delivery

Presently mobile learning is regarded as a core pedagogical activity in higher institution of learning and the instructional technology insist the content transmitted through mobile technology is mostly social, to a lesser extent and economic [6]. The content of a particular concept or sub-concept is observed and learned by the learner. An effective learning environment is provided with the help of ontology and sub-ontology based concepts. The learners should concentrate an assessment before beginning to learn a concept known to be pretest.

C. Evaluation and Feedback Provider

Once the student has finished the course, he/she would take up a post-test. During the test multiple-choice questions related to the topic are presented in slides and answers are gathered using SMS facilities of the mobile phones of the students that are illustrated in Figure 3.The results are compared with pretest to assess the improvement of the user [1] and the skills gained by the user. Test results are Tracked and stored to assess their improvement, and it is used as a measure of input of M-learning. It provides the facility that the learner can give feedback about the presentation of the content and about the ease of access.

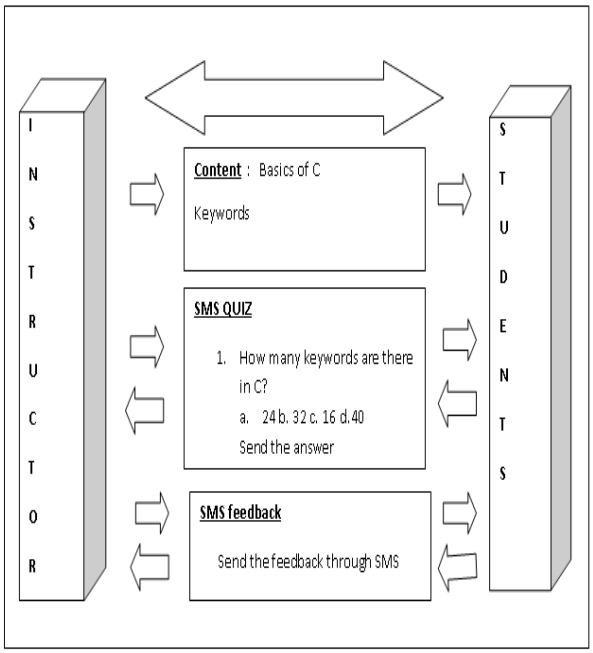


Figure 3. Evaluation and feedback process

The design of CML presents and preserves the hierarchical course structure with their semantic relationship between the concepts.

3. Database layer

The database components of user layer contain all the required information about learners. This layer maintains the databases with users' profile, ontology based content, test score and tracking status of the every learner.

4. EVALUATION AND EXPERIMENTAL RESULTS

The mobile platforms targeted for the study included android, Smartphone platforms provide a mobile web browser with support for HTML5. These direct-touch Smartphone platforms are specifically targeted as they provide a superior user experience compared to non-touch mobile devices. In addition, we need to provide the best possible user experience on mobile to see its confirmation.

The course on CYBER SECURITY is selected as sample course for M-learning to provide awareness of Information Security and give an exposure as a spectrum of security activities, methods, methodologies, and procedure. Our M-learning course includes the topics like security principles, threats, attacks, security models, security policies, overview of authentication, encryption, and certification, security detection, business risk analysis, protection of information assets.

This course is developed and pilot-tested. For this study, Bachelor of Science in computer science students are chosen from different colleges which are geographically distributed from urban, semi-urban and rural colleges of Madurai Kamaraj University and they have participated in this study with the condition that

1. Learner should have knowledge of computers.
2. Learner should have smart phones or tablet Pc linked with internet and sizable storage.

Once the learner enters into modified IRS-EEE he/she has to sign-in and enter his profile. The admin has approved the learner .The learner selects the course. Before the course content is presented there is pretest. The selected course is divided into blocks of related concepts. After careful learning of the concepts to estimate their capture before the course MCQ (Multiple Choice Questions) is given. MCQ assessment is one of the best and often used Formative Assessment tool [7] post-test is conducted to measure the knowledge improvement in the domain and the efficiency of the system.

The efficiency of the system is measured using the parameters like ease of interactivity, independency, instant access, storage and maintenance which are measured against Learners satisfaction in percentage. The results of the study shows the acceptance of mobile learning that is following shown in Figure 4.

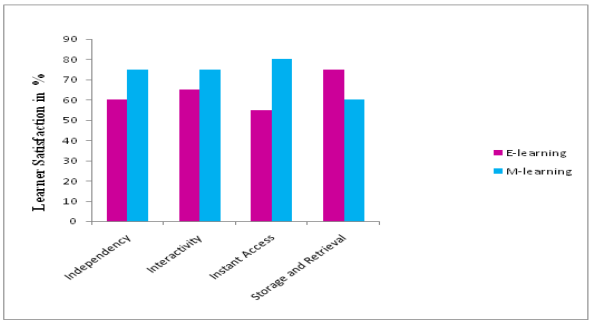


Figure 4. Comparisons of E-learning and M-learning

Compared to traditional learning environment E-learning changes the environment from classroom to computers that to be transferred to just a touch in M-learning.

The new generation prefers much than E-learning for its features like cost effectiveness. Personalization but the system maintenance is less in E-learning than M-learning which is shown in Figure 5.

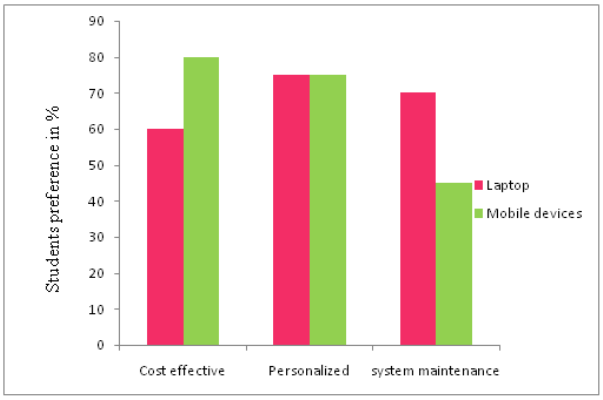


Figure 5. Student Access mobile vs laptops

5. CONCLUSION

As computer and Internet become essential tools for education; technology becomes more mobile, affordable, effective and easy to use. This offers many opportunities to widen participation and access to ICT, particularly the Internet (infoDev, 2010). Mobile devices such as phones and PDAs are much more affordable than desktop computers, and therefore represent a less expensive access to the internet (even if the cost of the connection may be higher) (InfoDev, 2010). The introduction of the Tablet PC can now access mobile internet with much functionality than desktop computers.

Mobile devices can be used anywhere, anytime, including at home, on the train, in hotels-this is invaluable for work-based training.

It is much easier to accommodate several mobile devices in a classroom than several desktop computers.

Future mobile technologies may be able to present textbooks, create database, aid in library utilization, and foster contextual learning. Mobile learning has become an integral part of education in many parts of the world, and new research advances in design and implementation will ensure its increasing importance.

6. REFERENCE

[1] Fernando A. Mikic Fonte, Juan C. Burguillo-Rial,Martin Llamas-Nistal, and David Fernandez-Hermida "A BDI based Intelligent Tutoring Module for the E-learning Platform INES" 40thASEE/IEEE frontiers in education Conference, October 27-30,2010, Washington, DC.

[2] Giselher H. J. R, "An Educational Taxonomy for Learning Objects", Third IEEE International Conference on Advanced Learning Technologies (ICALT'03), pp250, 2003.

[3] Hafizul Fahri Hanafi, Khairulanuar Samsudin,'' Mobile learning Environment System(MLES)", The case of Android based Learning Application on Undergraduates' Learning (IJACSA) International Journal of Advanced Computer Science and Applications Vol.3, No 3, 2012.

[4] Klopfer, E., and Squire, K. (2008),''Environmental Detectives: the development of an augmented reality platform for environmental simulations''. Educational Technology Research and Development.

[5] Ljiljana Stojanovic, Steffen Stab, Rudi Studer,"E-learning based on the Semantic Web", The Electronic JournalofE-learningvolume4.issue 2, pp111=118.

[6] Mohamed Osman M. El-Hussein and Johannes C. Cronje, "Defining Mobile Learning in the Higher Education Landscape, Educational Technology &Society'', 13 (3), 12–21.(2010).

[7] Noorminshah lahad, Georgios A. Dafoulas, Maya Milankovic-Atkinson, Alan Murphy, "E-learning in Developing Countries in Suggesting a Methodology for Enabling Computer Aided Assessment" IEEE International Conference on Advanced Learning Technology(ICALT) 2004.

[8] Shaileshkumar K. Patel et al, "Semantic Web Technology and Ontology designing for E-learning Environments'', International Journal of Computer Science and Information Technologies, (IJCSIT), Vol. 6 (1) , 2015, 48-51.

[9] Sohaib Ahmed and David Parsons, "An ontology-Driven Mobile Web Application for Science Enquiry Based Learning", The 7th International

Conference on Information Technology and Applications (ICITA 2011).

[10] Tayseer Andrawes Saleem, "Mobile learning technology: A new step in E-learning, Journal of Theoretical and Applied Information Technology", 31st December 20011, Vol 34 No.2, ISSN: 1992-8645.

[11] Thangaraj M, Vanathi S, " An Intelligent Recommender system for Effective E learning Environment, International Journal of Engineering Research and Technology (IJERT) ISSN: 2278-0181,Vol 3 Issue 11, November-2014.