*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 13, Issue 2 (Jul. - Aug. 2013), PP 01-07*
*www.iosrjournals.org*

# Privacy Management of Multi User Environment in Online Social Networks (OSNs)

P. Amrutha[1], R. Sathiyaraj[2]
*M.Tech Scholar , Dept.of CSE, Madanapalle Institute of Technology and Sciences, Madanapalle, JNTUA[1]*
*Assistant professor, Dept.of CSE, Madanapalle Institute of Technology and Sciences, Madanapalle, JNTUA[2].*

***Abstract:*** *Online Social Networks (OSNs) are inherently designed to enable people to share personal and public information and make social connections with others. These OSNs provides digital social interactions and social as well as personal information sharing, but in sharing a number of security and privacy problems raised. While OSNs allow users to restrict access to shared data, they currently do not provide any mechanism to totally enforce privacy issue solver associated with multiple users. To this end, we propose an approach to enable the protection of shared data associated with multiple users in OSNs. We formulate an access control model to capture the essence of multiparty authorization requirements, along with a multiparty policy specification scheme and a policy enforcement mechanism. Besides we also implement a proof-of-concept prototype which is called as MController (multi controller) having contributor, stakeholder and disseminator controllers along with owner controller.*

***Index Terms*** *--- social network, multi party access control, MController, decision voting*

## I. Introduction

Many people interested to share personal and public information and make social connections with friends, family, colleagues, coworkers and even with strangers through Online Social Networks(OSN) such like Facebook, Twitter, Google+ and etc,. OSN provide some space to each user for basic profile and sharing photos and videos with others. In photo sharing unfortunately some privacy and security problems are raised. Presently there is no mechanism to totally avoid these privacy issues. The main problem is collaborative authorization management, means if user tags the photo to his friend only. But the updates of photo are presented in both user as well as friends profiles. Then friend of friends or others may share that photo. So here the user expected privacy was spoiled. The existing protection for photos is binary condition either put or delete in profile space. If the photo was deleted after tagging, the content may loss in space, else the privacy was spoiled.

### 1.1. OSNs Privacy

In OSNs privacy restrictions form a spectrum between public and private data. On the public end, users can allow every particular OSN member to view their personal content. On the private end, users can restrict access to a specific set of trusted users. Despite the spectrum of available privacy settings, users have no control over information appearing outside their immediate profile page, when a user comment on friend's image, user and friend both cannot restrict the comment from other viewers. Similarly, if a user posts a photo and indicates the name of a friend in the photo, the friend cannot specify which users can view the photo. For both of these cases, Facebook currently lacks a mechanism to satisfy privacy constraints when multiuser is involved, So that the user's privacy may be violated. Privacy conflicts publicly expose personal information, slowly decreasing a user's privacy.

The user would have more control over his photos where a set of malicious users may want to make a shared photo available to a wider audience. If the malicious users can access the photo from original user then they tag photo with fake identities to others. Those may further share with other users. This continuous process, by this the original photo may change totally and shared with number of persons. At that time the privacy of photo which was expected by original user may collude totally. To prevent such an attack, three conditions need to be satisfied:

- No Fake Identity in OSNs.
- All Tagged Users are Real Users for the Photo.
- All Controllers are Honest to specify their Privacy policies for the photo.

## II. MController

OSN is mainly relationship network including set of users as well as their data. So that OSN represented with directed labeled graph where each node represents user and edge denotes relationship between two users. The edge direction denotes the relationship from initial to terminal node. The profile space of the user managed himself with his privacy data and content. For that privacy data to maintain security several schemes are introduced. But no scheme gives totally security, mainly all those schemes have only one controller that is owner. By this single controller security and privacy issues may be raised on data which was personal to the owner.

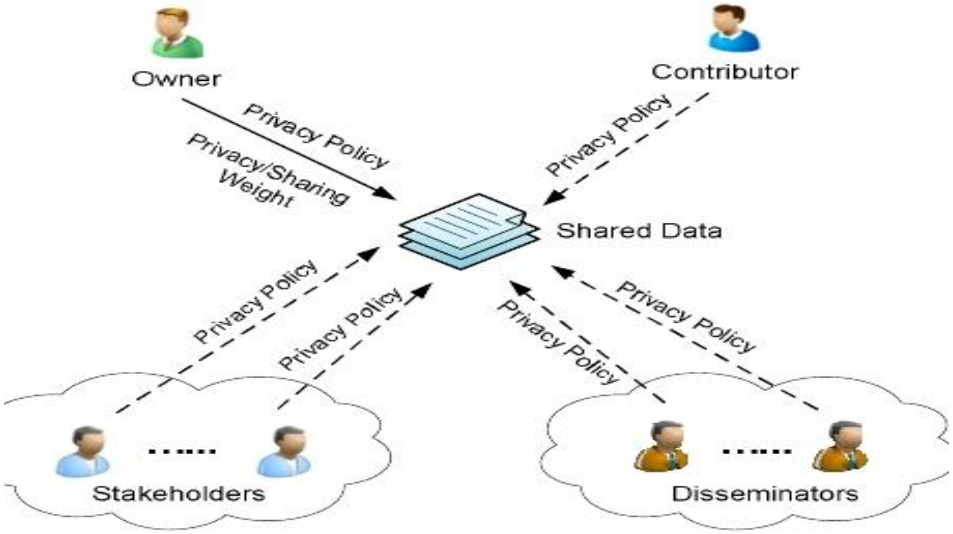



Figure.1. MController Architecture

So that rather than the owner controlling additional controllers are need for the flexible privacy mechanisms in OSN. The additional controllers are contributor, stakeholder and disseminator which provide their own privacy policies on shared data by giving the permission either permit or deny to unauthorized user on shared data. Figure 1 illustrates different controllers providing their privacy policies on shared data. We define multi controllers as follows:

- **Owner (O):** In the social network the user u is called the owner of the data item d, if d presents in the space m of user u. The user u is also called as contributor of d, when that user share data item d. The owner share data in three types, they are profile sharing, content sharing and relationship sharing. It enables the owner to discover potential malicious activities in collaborative control.
- **Contributor (C):** In the social network the user u is called the contributor of the data item d, if d published by user u in someone else's space. The contributor tags content to other's space and the content may also have multiple stakeholders (e.g., tagged users). The memory space for the user will be allotted according to user request for content sharing.
- **Stakeholder (S):** In the social network the user u is called a stakeholder of the data item d, if user u is tagged user T for d. A shared content has multiple stakeholders.
- **Disseminator (D):** In the social network, let d be a data item shared by a user u from someone else's space to his/her space. The user u is called a disseminator of d. the real content sharing starts with the owner, then disseminator views the content and shares with others. This disseminated content may be re-disseminated again and again by others.

## III. Multi Party Access Control (MPAC) Model

### 3.1. MPAC Specification

It is very essential for MPAC policies to regulate access and representing authorization requirements from multiple associated users to enable a collaborative authorization management of data sharing in OSNs.

- **Accessor Specification:** Accessor is the set of users who granted to access the shared data. Accessor can be represented with a set of user names, relationship names and group names in OSNs.

The accessor specification is defined as a set, accessors = {a1, a2. . . , an}, where each element is a tuple < ac,at >. where ac ∈ U ∪ RT ∪ G be a user u ∈ U, a relationship type rt ∈ RT, or a group g ∈ G. at ∈ {UN,RN,GN} be

the type of the accessor specification, where UN,RN,GN represents user name, relationship name, and group name.

- **Data Specification:** The data specification represented in three ways; profile, relationship and content sharing. For effective privacy the different controllers provide sensitivity levels on data.
  Let dt ∈ D be a data item, sl be a sensitivity level (range 0.00 to 1.00) for data item dt. The data specification is defined as a tuple < dt, sl >.

### 3.2. MPAC Policy

To summarize the above-mentioned specification elements, we introduce the definition of a multiparty access control policy as follows:
The multi party access control policy is a 5 - tuple
P = < controller, Ctype, accessor, data, effect >
where

- Controller is a user who can regulate the access of data.
- Ctype is the type of the controller.
- Accessor is the set of users who granted to access the shared data.
- Data is represents a data specification.
- Effect ∈ {permit, deny} is the authorization effect of the policy. Suppose a controller can leverage five sensitivity levels: 0.00 (none), 0.25 (low), 0.50 (medium), 0.75 (high), and 1.00 (highest) for the shared data.

### 3.3. MPAC Evaluation

Multi party access control is evaluated in two steps. In step-1, the individual decision are collected from different controllers, and in step-2, individual decision are aggregated and makes final decision for the access request.

Figure 2 illustrates that how MPAC evaluated in step by step. Initially an access request goes to under policy evaluation, which is done under four controllers. The four controllers provide their own privacy policies in the form of decision either permit or deny in step-1 process. After giving decisions by individual controllers, they are aggregated and make final decision by using decision voting schemes in step-2 process. The final decision making decides whether the access request is allowed or refused.

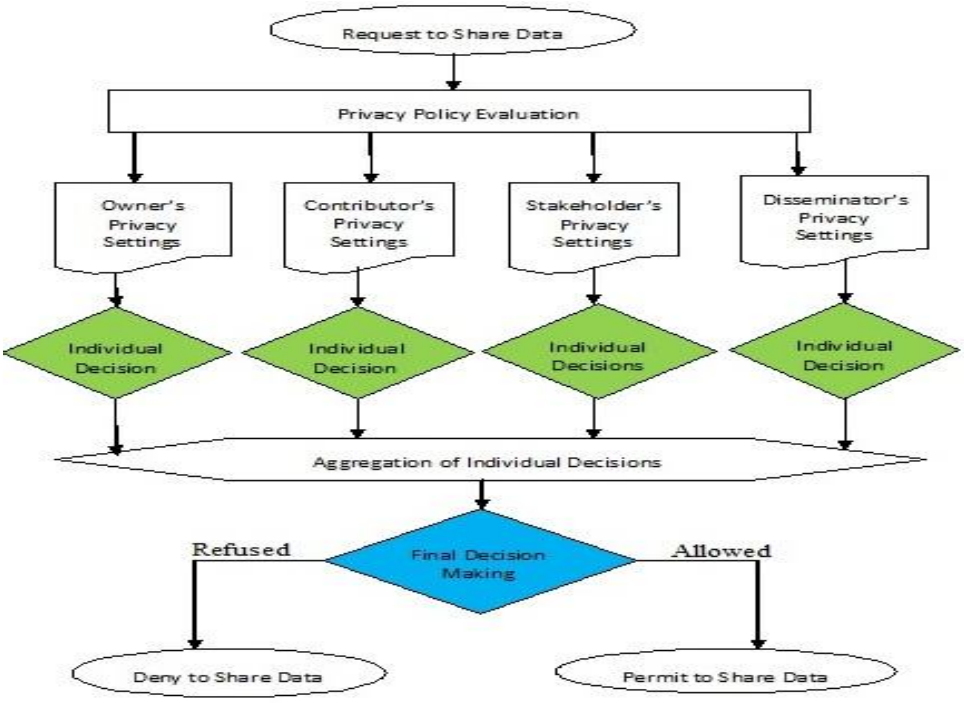



Figure.2. MPAC Evaluation

From the process of evaluation in MPAC policies, the controllers give different decision for an access request. There may be a chance of occurring conflicts. So that a mechanism is needed to resolute the conflicts

for taking an unambiguous decision for each access request. For the better privacy, a strong resolution for conflict may need. So it is better to consider tradeoff between privacy and utility in resolution of conflict. For this conflict issue, we introduce decision voting schemes resolving the MPAC conflicts which is simple and flexible.

## IV. Final Decision Making Schemes

### 4.1. Decision Voting Mechanism

Decision making mainly depends on majority. For such decision making, we introduce a voting scheme for conflict resolution. In voting mechanism each controller's individual decision effects the final decision. Mainly this voting scheme is described in two voting mechanisms; they are decision voting and sensitivity voting.

**4.1.1. Decision Voting:** the policy evaluation derives the decision voting value (DV) either permit or deny as follows,

Where Evaluation(p) represents the policy p decision:

$$DV = j \begin{cases} 0 & \text{if Evaluation(p) = Deny} \\ 1 & \text{if Evaluation(p) = Permit} \end{cases} \quad (1)$$

Assume that all controllers are equally important, an aggregated decision value (DVag) (range 0.00 to 1.00) from multiple controllers including the owner (DVow), the contributor (DVcb) and stakeholders (DVst), is computed with following equation:

$$DVag=(DVow+DVcb+\sum_{i\in SS} DV^{i}_{st})\times 1/m \quad (2)$$

Where SS is the set of stakeholders for shared data item, and m is total number of controllers for shared data item.

For the shared data item each controller may have (i) a different trust level over the data owner and (ii) a different reputation value in terms of collaborative control. So we need to introduce weights for decision voting scheme. Weights for different controllers can be calculated by aggregating trust levels and reputation values. The weight of controller x is "weightx / sum of weights". Suppose ωow, ωcb and ωi sh are weights for owner, contributor and stakeholder controllers, respectively, and n is the number of stakeholders of the shared data item. A weighted decision voting scheme is as follows:

$$DVag=(\omega ow\times DVow+\omega cb\times DVcb+\sum_{i=1ton}(\omega i_{st}\times DV^{i}_{st}))\times 1/(\omega ow+\omega cb+\sum_{i=1ton}\omega ist) \quad (3)$$

**4.1.2. Sensitivity Voting:** Each controller assigns a sensitivity level (SL) to the shared data item to reflect her/his privacy concern. A sensitivity score (SC) (range 0.00 to 1.00) for the data item can be calculated based on following equation:

$$SC=(SLow+SLcb+\sum_{i\in SS} SL^{i}_{st})\times 1/m \quad (4)$$

### 4.2. Threshold-Based Conflict Resolution

A basic idea of our approach for threshold-based conflict resolution is that the sensitivity score (SC) can be utilized as a threshold for decision making. Obviously, if SC increased, then the chance of final decision to deny is increased, so that the utility of OSN services cannot be affected. The threshold-based conflict resolution calculates final decision as follows:

$$Decision = j \begin{cases} Permit & \text{if } DVag > SC \\ Deny & \text{if } DVag \leq SC \end{cases} \quad (5)$$

It is worth noticing that our conflict resolution approach has an adaptive feature which reflects the changes of policies and sensitivity levels. If any controller changes his privacy policy or sensitivity level on the shared data item, then the aggregated decision value (DVag) and the sensitivity score (SC) will be recomputed and accordingly the final decision may be changed.

### 4.3. Strategy-Based Conflict Resolution

If we treat all controllers equally important, then above threshold-based conflict resolution provides a simple mechanism for making final decision. But in practical, different controllers may have different priorities making final decision. Especially the owner has highest priority in the control of shred data item. So that we provide strategy-based conflict resolution mechanism to satisfy owner authorization requirements of shared data.

Here the sensitivity score (SC) considered as guideline in selecting appropriate strategy for conflict resolution of shared data item. We introduce following strategies for the purpose of resolving multiparty privacy conflicts in OSNs.

- **Owner−overrides:** In final decision making, the highest priority goes to owner's decision. This strategy is totally owner controlling mechanism in data sharing. Based on the weighted decision voting scheme, we set ωow = 1, ωcb = 0 and ωst = 0,1 and the final decision can be made as follows:

$$\text{Decision} = j \begin{cases} \text{Permit} & \text{if DVag} = 1 \\ \text{Deny} & \text{if DVag} = 0 \end{cases} \quad (6)$$

- **Full−consensus−permit**: The final decision is deny, if any controller deny the access. This strategy can achieve the naive conflict resolution. The final decision can be derived as:

$$\text{Decision} = j \begin{cases} \text{Permit} & \text{if DVag} = 1 \\ \text{Deny} & \text{otherwise} \end{cases} \quad (7)$$

- **Majority−permit:** This strategy permits (denies, resp.) a request if the number of controllers to permit (deny, resp.) the request is greater than the number of controllers to deny (permit, resp.) the request. The final decision can be made as:

$$\text{Decision} = j \begin{cases} \text{Permit} & \text{if DVag} \geq ½ \\ \text{Deny} & \text{if DVag} < ½ \end{cases} \quad (8)$$

## V. Logical Representation of Multiparty Access Control

We introduce an ASP program for multiparty authorization specification.

### 5.1. Logical Definition of Controllers and Relationships

The basic components and relations in our MPAC model can be directly defined with corresponding predicates in ASP. We have defined UDct as a set of user-to-data relations with controller type ct ∈ CT. Then, the logical definition of multiple controllers is as follows:

• The owner controller of a data item can be represented as:
OW(controller,data) ← $UD_{OW}$(controller, data) ∧ (controller) ∧ D(data).
• The contributor controller of a data item can be represented as:
CB(controller, data) ← $UD_{CB}$(controller, data)∧ U(controller) ∧ D(data).
• The stakeholder controller of a data item can be represented as:
ST(controller, data) ← $UD_{ST}$ (controller, data)∧ U(controller) ∧ D(data).
• The disseminator controller of a data item can be represented as:
DS(controller, data) ← $UD_{DS}$(controller, data)∧ U(controller) ∧ D(data).
Our MPAC model supports transitive relationships. Then, friends-of-friends can be represented as a transitive closure of friend relation with ASP rule as follows:
friendsOFfriends(U1, U2) ← friendOf(U1, U2).
friendsOFfriends(U1, U3) ← friendsOFfriends(U1, U2), friendsOFfriends(U2, U3).

### 5.2. Logical Representation of Decision Voting Schemes

decision voting(C) = 1 ← decision(C, permit).
decision voting(C) = 0 ← decision(C, deny).
aggregation weight(K) ← K = sum{weight(C) : controller(C)}.
aggregation decision(N) ← N = sum{decision voting(C) × weight(C) : controller(C)}.
aggregation sensitivity(M) ← M = sum{sensitivity voting(C) × weight(C) : controller(C)}.

### 5.3. Logical Representation of Threshold-Based Conflict Resolution

decision(controllers, permit) ←N >M ∧ aggregation decision(N) ∧ aggregation sensitivity(M).
decision(controllers, deny) ← not decision(controllers, permit).

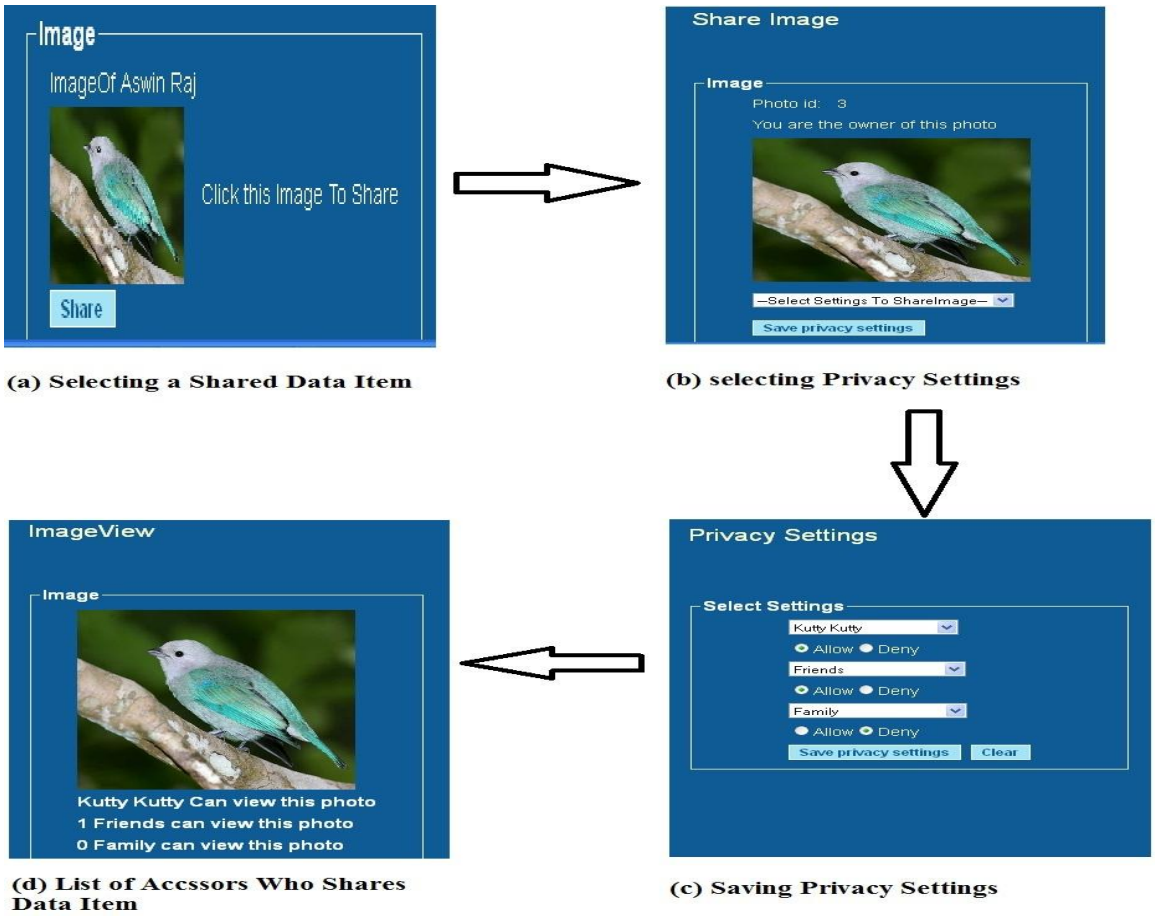
### 5.4. Logical Representation of Strategy-Based Conflict Resolution

**• The conflict resolution strategy for Owner−overrides is represented as:**
weight(controllers) = 1 ← OW(controller, data).
weight(controllers) = 0 ← CB(controller, data).
weight(controllers) = 0 ← ST(controller, data).
decision(controllers, permit) ← N/K == 1∧ aggregation weight(K) ∧aggregation decision(N).
decision(controllers, deny) ← not decision(controllers, permit).
**• The conflict resolution strategy for Full−consensus−permit is represented as:**
decision(controllers, permit) ← N/K == 1∧ aggregation weight(K) ∧aggregation decision(N).
decision(controllers, deny) ← not decision(controllers, permit).

**• The conflict resolution strategy for Majority−permit is represented as:**
decision(controllers, permit) ← N/K > 1/2 ∧ aggregation weight(K) ∧ aggregation decision(N).
decision(controllers, deny) ← not decision(controllers, permit).
**• The conflict resolution strategy for Deny−overrides for dissemination control is represented as:**
decision(deny) ← decision(controllers, deny).
decision(deny) ← decision(disseminator, deny).
decision(permit) ← not decision(deny).

## VI. Implementation

MController is third-party application development for Facebook. This is hosted in an Apache Tomcat application server supporting PHP and MySQL database. MController application is based on the iFrame external application approach. Using the Javascript and PHP SDK, it accesses users' Facebook data through the Graph API and Facebook Query Language. Once user install MController in his Facebook space and accepts the terms and conditions, then MController access the content and basic information of user. Mainly, it retrieves the list of all photos owned by user as well as tagged photos and uploaded. Now user access MController privacy settings on shared images and protect from other viewers.




**(a) Selecting a Shared Data Item**

**(b) selecting Privacy Settings**

**(d) List of Accssors Who Shares Data Item**

**(c) Saving Privacy Settings**

Figure.3. Snapshot of MController

A snapshot of main interface of MController is shown in Figure 3 illustrates that how MController runs in each step and execution. Initially the user selects the image which he needs share and click on share button as showing in figure 3.a. the figure 3.b shows share image and privacy setting option. This privacy setting option is the main aim of MController system. If the user selects the privacy settings option, settings page appeared as like figure 3.c. the figure 3.c shows the options of individual persons as well as groups. Here the user selects access or deny option for different groups like family, friends and coworkers. After settings completed, the user click on save button or else click on cancel button to reset the settings. Once the settings are saved by user, under the shared image, the list of visitors can appeared according to the user privacy settings as shown in figure 3.d. The visitors list informed that who can only see and share the user's image.

## VII. Related Work

Access control for OSNs is still relatively a new research area for privacy issues. Presently several access control models for OSNs have been introduced. Fong et al. proposed an access control model that formalizes and generalizes the access control mechanism implemented in Facebook, which admitting arbitrary policy vocabularies that are based on theoretical graph properties. Fong recently formulated this paradigm called a Relationship- Based Access Control (ReBAC) model that bases authorization decisions on the relationships between the resource owner and the resource accessor in an OSN. Carminati et al. recently introduced collaborative security policies, a new class of security policies, that basically enhance topology-based access control with respect to a set of collaborative users.

## VIII. Conclusion

In this paper, we found the need of privacy for OSN and solution of collaborative authorization management of the shared data. We introduced MController technique to provide their own privacy preferences on a shared data by the different controllers. Additionally MPAC model evaluated providing decision voting schemes and the privacy evaluation. In the future work, we are planning to investigate advanced MController technique to provide privacy settings for the group of photos at a time, because users may be involved to put privacy setting for the number of photos at a time. By this MPAC model it is time consuming process. So that we would study advanced MController for shared data to automatic configure the privacy.

## References


[1] Hongxin Hu, Gail-Joon Ahn, Senior Member, IEEE, and Jan Jorgensen "**Multiparty Access Control for Online Social Networks: Model and Mechanisms"** IEEE transactions,2012.

[2] Besmer and H. Richter Lipford. **"Moving Beyond Untagging: Photoprivacy in A Tagged World"**. pages 1563–1572. ACM, 2010.

[3] L. Bilge, T. Strufe, D. Balzarotti and E. Kirda. **"All Your Contacts are Belong to Us: Automated Identity Theft Attacks On Social Networks"**. pages 551–560. ACM, 2009.

[4] Carminati and E. Ferrari. **"Collaborative Access Control in Online Social Networks"**. pages 231–240. IEEE, 2011.

[5] Carminati, E. Ferrari, and A. Perego. **"Rule-Based Access Control for Social Networks"**. pages 1734–1744. Springer, 2006.

[6] B. Carminati, E. Ferrari, and A. Perego. **"Enforcing access control in web-based social networks."** ACM Transactions on Information and System Security (TISSEC), 13(1):1–38, 2009.

[7] Carrie. **"Access Control Requirements for Web 2.0 Security and Privacy."** In Proc. of Workshop on Web 2.0 Security & Privacy (W2SP). Citeseer, 2007.

[8] J. Choi, W. De Neve, K. Plataniotis, and Y. Ro. **"Collaborative face recognition for improved face annotation in personal photo collections shared on online social networks."** Multimedia, IEEE Transactions on, 13(1):14–28, 2011.

[9] P. Fong. **"Preventing sybil attacks by privilege attenuation: A design principle for social network systems."** In Security and Privacy (SP), 2011 IEEE Symposium on, pages 263–278. IEEE, 2011.

[10] P. Fong.**"Relationship-based access control: Protection model and policy language".** In Proceedings of the first ACM conference on Data and application security and privacy, pages 191–202. ACM, 2011.

[11] P. Fong, M. Anwar, and Z. Zhao. "**A privacy preservation model for facebook-style social network systems".** In Proceedings of the 14th European conference on Research in computer security, pages 303–320. Springer-Verlag, 2009.

[12] J. Golbeck. "**Computing and applying trust in web-based social networks"**. Ph.D. thesis, University of Maryland at College Park College Park, MD, USA. 2005.

[13] M. Harrison, W. Ruzzo, and J. Ullman. "**Protection in operating systems"**. Communications of the ACM, 19(8):461–471, 1976.

[14] H. Hu and G. Ahn. "**Enabling verification and conformance testing for access control model**". In Proceedings of the 13th ACM symposium on Access control models and technologies, pages 195–204. ACM, 2008.

[15] H. Hu and G. Ahn. "**Multiparty authorization framework for data sharing in online social networks**". In Proceedings of the 25th annual IFIP WG 11.3 conference on Data and applications security and privacy, pages 29–43. Springer-Verlag, 2011.

[16] H. Hu, G. Ahn, and K. Kulkarni. "**Anomaly discovery and resolution in web access control policies**". In Proceedings of the 16th ACM symposium on Access control models and technologies, pages 165– 174. ACM, 2011.

[17] H. Hu, G.-J. Ahn, and J. Jorgensen. "**Enabling Collaborative Data Sharing in Google+**". Technical Report ASU-SCIDSE-12-1, April 2012.

[18] H. Hu, G.-J. Ahn, and J. Jorgensen. "**Detecting and resolving privacy conflicts for collaborative data sharing in online social networks**". In Proceedings of the 27th Annual Computer Security Applications Conference, ACSAC '11, pages 103–112. ACM, 2011.

[19] H. Hu, G.-J. Ahn, and K. Kulkarni. "**Detecting and resolving firewall policy anomalies. IEEE Transactions on Dependable and Secure Computing**", 9:318–331, 2012.

[20] L. Jin, H. Takabi, and J. Joshi. "**Towards active detection of identity clone attacks on online social networks**". In Proceedings of the first ACM conference on Data and application security and privacy, pages 27–38. ACM, 2011.

[21] L. Lam and S. Suen. "**Application of majority voting to pattern recognition: an analysis of its behavior and performance**". Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 27(5):553–568, 2002.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 13, Issue 2 (Jul. - Aug. 2013), PP 08-13*
*www.iosrjournals.org*

# Usage and Research Challenges in the Area of Frequent Pattern in Data Mining

P. Alagesh Kannan[1], Dr. E. Ramaraj[2]
[1] *Assistant Professor, Department Of Computer Science, MKU College, Madurai, Tamil Nadu, India*
[2]*Professor, Department of Computer Science and Engg, Alagappa University, Karaikudi Tamil Nadu, India*

***Abstract:*** *Frequent pattern mining is an important chore in the data mining, which reduces the complexity of the data mining task. The usages of frequent patterns in various verticals of the data mining functionalities are discussed in this paper. The gap analysis between the requirements and the existing technology is also analyzed. State of art in the area of frequent pattern mining was thrashed out here. Working mechanisms and the usage of frequent patterns in various practices were conversed in the paper. The core area to be concentrated is the minimal representation, contextual analysis and the dynamic identification of the frequent patterns.*
***Keywords:*** *Frequent pattern, Association, Clustering, Classification*

## I. Introduction

Frequent patterns are the subset of the given dataset with the occurrence frequency that satisfies the user specified threshold and above [1]. Identification of the frequent patterns is a thrust area in the field of data mining which has the applications in association mining, correlation analysis. It is an important part of finding the interesting relationships within the data given. Other functional areas of data mining such as clustering, classification indexing also use the frequent patterns.

Frequent pattern mining was first proposed by Agrawal et al [2] for market basket analysis in the form of association rule mining. It analyses customer buying habits by finding associations between the different items that customers place in their "shopping baskets". The concept is explained as

Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of all items. A k-itemset $\alpha$, which consists of k items from I, is frequent if $\alpha$ occurs in a transaction database D no lower than $\theta|D|$ times, where $\theta$ is a user-specified.

In this paper the usage of frequent patterns in the area of finding the association, classification and clustering is discussed thoroughly. A deep review of the state of art techniques was reviewed. In the section 5 the research challenges posed in the field of frequent pattern mining is discussed.

## II. Fundamental Approaches for Mining Associations Using Frequent Patterns

The basic concepts which used frequent patterns to mine the association between data are discussed in this section. Apriori, FP-growth and Eclat are the three basic approaches emerged by using frequent patters.

### II.A Algorithms using Apriori approach

Agrawal and Srikant [3] observed an interesting downward closure property, called Apriori, among frequent k item sets: A k-itemset is frequent only if all of its sub-item sets are frequent. This implies that frequent item sets can be mined by first scanning the database to find the frequent 1-itemsets, then using the frequent 1-itemsets to generate candidate frequent 2-itemsets, and check against the database to obtain the frequent 2-itemsets. This process iterates until no more frequent k-item sets can be generated for some k. This also motivated its alternative [4] where the algorithm uses all existing information between database passes to avoid checking the coverage of redundant sets.

It follows a level wise search or the breath first search. The extension of this technique reflected in many of the algorithms with some deviations for efficiency. [5] Used the direct hashing and pruning for efficient large item set generation with effective reduction on transaction database size, [6] used the partitioning technique to read the database at most two times to generate all significant association rules. Toivonen [7] suggested an algorithm which uses random sample to form association and later check with the whole database to avoid approximation. The approach is probabilistic and some time requires a second pass.

Dynamic item set counting and implication rules are used in [8] to reduce the number of passes. An incremental updating technique is proposed for efficient maintenance of discovered association rules when new transaction data are added to a transaction database in [9]. Parallelization process is done for finding the association rules in [10], which employs a clue and poll technique to address the uncertainty due to partial knowledge. In the paper [11] provides a hard and tight combinatorial upper bound to answer the question of the maximal number of candidate patterns that can be generated in the passes.

Even though the Apriori throws new light in the frequent pattern generation by the reduction of the size of the candidate sets, the problem with which it suffers is enormous number of the candidate set generation and the repeated scanning of database, which is costly for the large databases.

**II.B Algorithms using Frequent Pattern growth approach**

Han et al. [12] presented an approach to generate frequent patterns without candidate set generation, which is a bottleneck in apriori approach. It built a compact frequent pattern tree structure and from that the frequent patterns are extracted by traversing recursively the tree. The pattern growth is achieved by the concatenation of the suffix pattern with the frequent patterns generated from a conditional frequent pattern tree.

This approach tackles the problem by identification of shorter frequent patterns and from that the long patterns are constructed recursively by concatenation. [13] and [14] uses the hyper-structure mining of frequent patterns for building alternative trees. An array-based implementation of prefix-tree-structure for efficient pattern growth mining by Grahne and Zhu is suggested in [15].

This model reduces the irrelevant information by the deletion of the infrequent items and the more frequently occurring, the more likely to be shared. It recursively work for the frequent pattern and it is done in an incremental mode. It doesn't search for the pattern and match instead it count local frequent pattern and then built the tree.

This technique lacks in few conditions like the tree may not fit into the memory, it takes time to build, and when the support threshold is high time is wasted as pruning could be done on single items. Another important problem is support can only be calculated once the entire data-set is added to the tree.

**II.C Algorithms using the vertical exploration approach**

Zaki [16] proposed Equivalence CLAss Transformation (ECLAT) algorithm by exploring the vertical data format. It uses the depth first search strategy with set intersection. It states that, when the database is stored in the vertical layout, the support of a set can be counted much easier by simply intersecting the covers of two of its subsets that together give the set itself [17].

This algorithm generates the candidate item sets using the join step proposed in the apriori. After this all the items in the database are reordered in ascending. This reduces the number of intersections to be computed. It doesn't fully exploit the monotony property; the number of candidate itemsets that are generated is much larger as compared to a breadth-first approach such as Apriori.

CHARM [18] is an efficient algorithm for enumerating the set of all closed frequent itemsets. CHARM is unique in that it simultaneously explores both the itemset space and transaction space, unlike all previous association mining methods which only exploit the itemset search space. It avoids enumerating all possible subsets of a closed itemset when enumerating the closed frequent sets, which rules out a pure bottom-up search.

## III. Use of Frequent Patterns for Classification

The application of frequent patterns in classification appeared in sporadic studies and achieved initial success in the classification of relational data, text documents and graphs [19]. Frequent patterns reflect strong associations between items and carry the underlying semantics of the data. They are potentially useful features for classification. Frequent pattern-based classification could exploit the state-of-the-art frequent pattern mining algorithms for feature generation, thus achieving much better scalability than the method of enumerating all feature combinations.

Frequent pattern based classification could be employed with feature generation, feature selection, and model learning. Frequent patterns form the efficient features since each pattern is a combination of single features and they are frequent. Frequent pattern is a form of non-linear feature combination over the set of single features. With inclusion of non-linear feature combinations, the expressive power of the new feature space increases.

Direct Discriminative Pattern Mining (DDPMine) performs a branch-and bound search for directly mining discriminative patterns without generating the complete pattern set. Instead of selecting best patterns in a batch, a "feature-centered" mining approach is proposed that generates discriminative patterns sequentially on a progressively shrinking FP-tree by incrementally eliminating training instances. The instance elimination effectively reduces the problem size iteratively and expedites the mining process [20].

Discriminative Pattern Mining Approach for the Classification of Software Behaviors for Failure Detection is suggested in [21], which first mines a set of discriminative features capturing repetitive series of events from program execution traces. It then performs feature selection to select the best features for classification. These features are then used to train a classifier to detect failures.

Frequent patterns allow the construction of high-level sets of compound features which can, in many cases, capture more discriminative information [19]. Frequent patterns are used exhaustively in image classification in [22][25], understanding the scene [26], object recognition and object-part recognition [27].

Effective method for using item set mining to discover a new set of mid-level features called Frequent Local Histograms is suggested by Basura Fernando et. Al [28].

PatMat a classifier based on frequent pattern is built with the use of trie data structure [29]. Monowar et.al [30] suggested an approach for fingerprint classification using data mining approach. Initially, it generates a numeric code sequence for each fingerprint image based on the ridge flow patterns. Then for each class, a seed is selected by using a frequent item sets generation technique. These seeds are subsequently used for clustering the fingerprint images.

Jeroen De Knijf describes a classification method for XML data based on frequent attribute trees. From these frequent patterns the emerging patterns are selected, and use these as binary features in a decision tree algorithm [31]. In this method they used k different minimum support values; one for each class label. To determine an appropriate minimum support value for each class, they started with a high support value, and lowered it gradually until a sufficient number of high quality patterns were produced.

A unified framework for mining multiple domain datasets and design an iterative algorithm called HTMiner[32] capture the associations among different kinds of data. HTMiner discovers essential heterogenous patterns for classification and performs instance elimination. This instance elimination step reduces the problem size progressively by removing training instances which are correctly covered by the discovered essential heterogeneous pattern.

Iyad Batal [33] said the even though some frequent patterns can be important predictors, using all frequent patterns in the classifier is not a good option for the following reasons like large number of patterns, many non predictive patterns and many spurious patterns. It encompasses that the research opening is present in the frequent pattern mining for classification.

## IV. Use of Frequent Patterns for Clustering

Clustering is the unsupervised approach for learning where the process is used for finding the data distribution and patterns in the datasets where their class labels are not known. Pattern based clustering does not require a globally defined similarity measure. Different clusters can follow different patterns on different subsets of dimensions. On the other hand, the clusters are not necessary exclusive. That is, an object can appear in more than one cluster. The generality and flexibility of pattern-based clustering may provide interesting and important insights in some applications where conventional clustering methods may meet difficulties.

An agglomerative hierarchical clustering algorithm to find clusters among the periodic item sets was suggested in [34]. Since the fuzzy number is invariant with respect to shifting, they define similarity measure using the variance of fuzzy intervals associated with frequent item sets.

Frequent pattern for text clustering was discussed in [35]. Here they have used the measure of mutual overlap of frequent sets with respect to the sets of supporting documents. They have presented two algorithms for text clustering, FTC which creates flat clustering and HFTC for hierarchical clustering. FTC works in a bottom-up fashion. Starting with an empty set, it continues selecting one more element (one cluster description) from the set of remaining frequent itemsets until the entire document collection is contained in the cover of the set of all chosen frequent itemsets. In each step, FTC selects one of the remaining frequent itemsets which has a cover with minimum overlap with the other cluster candidates, i.e. the cluster candidate which has the smallest entropy overlap (EO) value. The documents covered by the selected frequent itemsets are removed from the collection D, and in the next iteration, the overlap for all remaining cluster candidates is recomputed with respect to the reduced collection. A frequent item-based approach of clustering is promising because it provides a natural way of reducing the large dimensionality of the document vector space.

Pattern preserving clustering [36] was discussed by Hui Xion et. al. They suggest a better cluster interpretation than traditional clustering approaches by considering the patterns found in clusters. Pattern-based clustering algorithms determine clusters based on the similarities of the patterns among objects across the relevant dimensions, instead of the absolute distance values among objects [37]. Pattern-based algorithm is the bi cluster model proposed by Cheng et al. [38], which tries to measure the coherence of the genes and the conditions in a sub matrix of a DNA array. P Cluster [39] was introduced to cluster objects by computing maximal candidate attributes set (MCAS) and maximal candidate objects set (MXOS) iteratively.

The Maple method [40] advances further this idea to find the maximal pattern-based clusters. Redundant clusters are avoided completely by mining only the maximal pattern-based clusters. The idea is to report only those non-redundant pattern-based clusters, and skip their trivial sub-clusters. It conducts a depth-first, progressively refining search to mine maximal pattern-based clusters. By mining maximal pattern-based clusters, the number of clusters can be reduced substantially. Moreover, many unnecessary searches for sub-clusters can be pruned and thus the mining efficiency can be improved dramatically as well.

An expression pattern based bi clustering technique, CoBi, for grouping both positively and negatively regulated genes from microarray expression data was discussed in [41]. Regulation pattern and similarity in degree of fluctuation are accounted for while computing similarity between two genes. Unlike traditional bi

clustering techniques, which use greedy iterative approaches, it uses a Bi-Clust tree that needs single pass over the entire dataset to find a set of biologically relevant bi clusters.

Clustering based on Frequent Word Sequence (CFWS) is projected in [42]. CFWS uses frequent word sequence and K-mismatch for document clustering. The difference between word sequence and word item set is that word sequence considers words' order while word item sets ignores words' order. Document Clustering Based on Maximal Frequent Sequences (CMS) is proposed in the paper [43]. A frequent sequence is maximal if it is not a subsequence of any other frequent sequence. The basic idea of CMS is to use maximal frequent sequences (MFS) of words as features in vector space model (VSM) for document representation and then k-means is employed to group documents into clusters.

Frequent Itemset-based Hierarchical Clustering (FIHC) is proposed by B.Fung and K.Wang[44]. Two kinds of frequent item are defined in FIHC: global frequent item and cluster frequent item. FIHC develops four phases to produce document clusters: finding global frequent itemsets, initial clustering, tree construction, and pruning.

Maximum Capturing (MC) for document clustering is to produce natural and comprehensible document clusters was discussed in [45].To produce natural clusters to use frequent itemsets for representation and measure similarities of documents based on cooccurrences of frequent itemsets in documents was the notion. To make document clusters comprehensible, most frequent itemsets in a document cluster were assigned as the topic of the cluster. Because documents with largest number of common frequent itemsets were assigned into a same cluster, cluster topics will be the most representative frequent itemsets in the cluster and thus distinguish clusters from each other.

Fuzzy Frequent Itemset-Based Hierarchical Clustering ($F^2$IHC) approach, which uses fuzzy association rule mining algorithm to improve the clustering accuracy of Frequent Itemset-Based Hierarchical Clustering (FIHC) method, was discussed in [46]. In this approach, the key terms will be extracted from the document set, and each document is pre-processed into the designated representation for the following mining process. Then, a fuzzy association rule mining algorithm for text is employed to discover a set of highly-related fuzzy frequent itemsets, which contain key terms to be regarded as the labels of the candidate clusters. Finally, these documents will be clustered into a hierarchical cluster tree by referring to these candidate clusters.

In order to tackle the challenges in the high dimensional clustering, sub space clustering is introduced. For subspace clustering, the determination of subspaces possibly containing clusters is a critical and time-consuming process. Jihong Guan et. al [37] discussed a new pattern-based subspace clustering algorithm CPT by using Pattern tree. The key point is that CPT adopts the pattern tree to discovering the subspace by scanning the database only once.

## V. Research Challenges in Finding Frequent Patterns

In the previous sections we have discussed about the need, importance and the usage of the frequent patterns in various verticals of data mining functionalities. This section let us discuss about the challenges in the field of research in frequent pattern mining.

The important challenge is the set of frequent patterns derived by most of the current pattern mining methods is too huge for effective usage. There are proposals on reduction of such a huge set, including closed patterns, maximal patterns, approximate patterns, condensed pattern bases, representative patterns, clustered patterns, and discriminative frequent patterns [1]. Much research is still needed to substantially reduce the size of derived pattern sets and enhance the quality of retained patterns.

The real bottleneck of the problem is not at the efficiency but at the usability. Typically, if the minimum support is high, mining may generate only commonsense patterns, however, with a low minimum support, it may generate an explosive number of results. This has severely restricted the usage of frequent-pattern mining [47]. Reducing the number of uninteresting patterns is an active and emerging research area [50]. Mining compressing patterns is NP-Hard and belongs to the class of inapproximable problems. Future work should be concentrated on further improvements to the mining algorithm using ideas from compression but keeping the focus on usefulness for data mining [51].

Unlike mining static databases, mining data streams poses many new challenges. In addition to the one-scan nature, the unbounded memory requirement and the high data arrival rate of data streams, the combinatorial explosion of item sets exacerbates the mining task [48]. The main challenges in the data streams are handling of the continuous flow, modeling changes of mining results over time, Data stream pre-processing, Model overfitting, Data stream mining technology [49].

The main research work on pattern analysis has been focused on pattern composition (e.g., the set of items in item-set patterns) and frequency. A contextual analysis of frequent patterns over the structural information can identify why that particular pattern is frequent. The deep understanding of frequent patterns is essential to improve the interpretability and the usability of frequent patterns [52].

## VI. Conclusion

This paper acts as the literature review in the area of the frequent pattern mining. The paper addresses the various techniques used for the frequent pattern mining and the need of it in the various functionalities of the data mining arena. The research challenges and the area to be concentrated are also discussed.

To conclude the area where the frequent pattern mining is to be concentrated is on the dynamic, Contextual and the compressed frequent pattern mining algorithms. The dynamic nature must address the mining concept with respect to the data streams. The contextual part must focus on the area where the pattern's structural component for the understanding semantics. The compression part must spotlight the minimal requirement of the pattern for the representation.

## References


[1] Jiawei Han, Hong Cheng, Dong Xin,Xifeng Yan, "Frequent pattern mining: current status and future Directions", Data Mining and Knowledge Discovery, Vol. 15 (2007), pp. 55-86.

[2] Agrawal R, Imielinski T, Swami A (1993) Mining association rules between sets of items in large databases. In: Proceedings of the 1993ACM-SIGMODinternational conference on management of data (SIGMOD'93), Washington, DC, pp 207–216.

[3] Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings of the 1994 international conference on very large data bases (VLDB'94), Santiago, Chile, pp 487–499

[4] Mannila H, Toivonen H, Verkamo AI (1994) Efficient algorithms for discovering association rules. In: Proceeding of the AAAI'94 workshop knowledge discovery in databases (KDD'94), Seattle, WA, pp 181–192

[5] Park JS, Chen MS, Yu PS (1995) An effective hash-based algorithm for mining association rules. In: Proceeding of the 1995 ACM-SIGMOD international conference on management of data (SIGMOD'95), San Jose, CA, pp 175–186

[6] Savasere A, Omiecinski E, Navathe S (1995) An efficient algorithm for mining association rules in large databases. In: Proceeding of the 1995 international conference on very large data bases (VLDB'95), Zurich, Switzerland, pp 432–443

[7] Toivonen H (1996) Sampling large databases for association rules. In: Proceeding of the 1996 international conference on very large data bases (VLDB'96), Bombay, India, pp 134–145

[8] Brin S, Motwani R, Ullman JD, Tsur S (1997) Dynamic itemset counting and implication rules for market basket analysis. In: Proceeding of the 1997 ACM-SIGMOD international conference on management of data (SIGMOD'97), Tucson, AZ, pp 255–264

[9] Cheung DW, Han J, Ng V, Wong CY (1996) Maintenance of discovered association rules in large an incremental updating technique. In: Proceeding of the 1996 international conference on data engineering (ICDE'96), New Orleans, LA, pp 106–114

[10] Park JS, Chen MS, Yu PS (1995) Efficient parallel mining for association rules. In: Proceeding of the 4th international conference on information and knowledge management, Baltimore, MD, pp 31–36

[11] Geerts F, Goethals B, Bussche J (2001) A tight upper bound on the number of candidate patterns. In: Proceeding of the 2001 international conference on data mining (ICDM'01), San Jose, CA, pp 155–162

[12] Han J, Pei J, Yin Y (2000) Mining frequent patterns without candidate generation. In: Proceeding of the 2000 ACM-SIGMOD international conference on management of data (SIGMOD'00), Dallas, TX, pp 1–12

[13] Agarwal R, Aggarwal CC, Prasad VVV (2001) A tree projection algorithm for generation of frequent itemsets. J Parallel Distribut Comput 61:350–371

[14] Pei J, Han J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, Hsu M-C (2001) PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth. In: Proceeding of the 2001 international conference on data engineering (ICDE'01), Heidelberg, Germany, pp 215–224

[15] Grahne G, Zhu J (2003)Efficiently using prefix-trees in mining frequent itemsets. In: Proceeding of the ICDM'03 international workshop on frequent itemset mining implementations (FIMI'03), Melbourne, FL, pp 123–132

[16] ZakiMJ(2000) Scalable algorithms for association mining. IEEE Trans Knowl Data Eng 12:372–390

[17] Pramod S. and Vyas (2010), Survey Frequent Itemset Mining, International Journal of Computer Applications 1(15):pp 86–91

[18] Zaki MJ,Hsiao CJ (2002) CHARM: an efficient algorithm for closed itemset mining. In: Proceeding of the 2002SIAMinternational conference on data mining (SDM'02),Arlington,VA, pp 457–473

[19] Hong Cheng, Yan, X.; Jiawei Han (2007); Chih-Wei Hsu, Discriminative Frequent Pattern Analysis for Effective Classification, IEEE 23rd International Conference on Data Engineering pp 716 – 725

[20] Hong Cheng, Xifeng Yan, Jiawei Han, Philip S. Yu (2008), Direct Discriminative Pattern Mining for Effective Classification, IEEE 24th International Conference on Data Engineering, 169 – 178

[21] David Lo, Hong Chen, Han, Khoo,Chengnian Sun, Classification of software behaviors for failure detection: a discriminative pattern mining approach, 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Pages 557-566

[22] Nowozin, S., Tsuda, K., Uno, T., Kudo, T., Bakir, G.: Weighted substructure mining for image analysis. In: CVPR. (2007)

[23] Yuan, J., Wu, Y., Yang, M.: Discovery of collocation patterns: from visual words to visual phrases. In: CVPR. (2007)

[24] Yuan, J., Yang, M., Wu, Y.: Mining discriminative co-occurrence patterns for visual recognition. In: CVPR. (2011) 2777 -2784

[25] Kim, S., Jin, X., Han, J.: Disiclass: discriminative frequent pattern-based image classi cation. In: Tenth Int. Workshop on Multimedia Data Mining. (2010)

[26] Yao, B., Fei-Fei, L.: Grouplet: A structured image representation for recognizing human and object interactions. In: CVPR. (2010)

[27] Quack, T., Ferrari, V., Leibe, B., Van Gool, L.: E_cient mining of frequent and distinctive feature configurations. In: ICCV. (2007)

[28] Basura Fernando, Elisa Fromont, Tinne Tuytelaars, Effective Use of Frequent Itemset Mining for Image Classification, Lecture Notes in Computer Science Volume 7572, 2012, pp 214-227 in Computer Vision – ECCV 2012 (2012)

[29] Wim Pijls, Rob Potharst, classification and target group selection based on frequent patterns, ERS -2000-40-LS, October 2000.

[30] Monowar H. Bhuyan, Sarat Saharia, Dhruba Kr Bhattacharyya, An Effective Method for Fingerprint Classification, International Arab Journal of e-Technology, 2010, 1(3): 89 – 97.

[31] Jeroen De Knijf, FAT-CAT: Frequent Attributes Tree Based Classification, Comparative Evaluation of XML Information Retrieval Systems ,Lecture Notes in Computer Science Volume 4518, 2007, pp 485-496

[32] Dhaval Patel, Wynne Hsu, Mong Li Lee, "Integrating Frequent Pattern Mining from Multiple Data Domains for Classification," icde, pp.1001-1012, 2012 IEEE 28th International Conference on Data Engineering, 2012

[33] Batal, I. and Hauskrecht, M. (2010), Constructing Classification Features using Minimal Predictive Patterns. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM '10, pages 869–878, New York, NY, USA. ACM.
[34] Fokrul Alom Mazarbhuiya, Muhammad Abulaish, Clustering periodic frequent patterns using fuzzy Statistical parameters, International Journal of Innovative Computing, Information and Control ISSN 1349-4198 ,Volume 8, Number 3(B), March 2012.
[35] Beil F, EsterM, Xu X (2002) Frequent term-based text clustering. In: Proceeding of the 2002 ACM SIGKDD international conference on knowledge discovery in databases (KDD'02), Edmonton, Canada, pp 436–442
[36] Hui Xiong, Michael Steinbach, Arifin Ruslim, ,Vipin Kumar, "Characterizing pattern preserving clustering", Knowledge and Information Systems, June 2009, Volume 19, Issue 3, pp 311-336
[37] Jihong Guan, Yanglan Gan, Hao Wang, Discovering pattern-based subspace clusters by pattern tree, Knowledge-Based Systems, Volume 22, Issue 8, December 2009.
[38] Y.Z. Cheng, M. George, Biclustering of expression data, Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology, 2000, pp. 93–103.
[39] H.X. Wang, W. Wang, J. Yang, S. Yu. Philip, clustering by pattern similarity in large data sets, Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, 2002.
[40] Xiaoling Zhang; Moonjung Cho; Haixun Wang; Yu, P.S, MaPle: a fast algorithm for maximal pattern-based clustering, ICDM 2003. Third IEEE International Conference on Data Mining, 2003.
[41] Swarup Roy, Dhruba K.Bhattacharyya, Jugal K.Kalita, CoBi:Pattern Based Co-Regulated Biclustering of Gene Expression Data, Pattern Recognition Letters, March 2013
[42] Y.J. Li, S.M. Chung, J.D. Holt, Text document clustering based on frequent word meaning sequences, Data & Knowledge Engineering 64 (2008) 381–404.
[43] H. Edith, A.G. Rene, J.A. Carrasco-Ochoa, J.F. Martinez-Trinidad, Document clustering based on maximal frequent sequences, in: Proceedings of the FinTAL 2006, LNAI, vol. 4139, 2006, pp. 257–267.
[44] B. Fung, K. Wang, M. Ester, Hierarchical document clustering using frequent itemsets, in: Proceedings of the 3rd SIAM International Conference on Data Mining, 2003.
[45] Wen Zhang , Taketoshi Yoshida, Xijin Tang , Qing Wanga, Text clustering using frequent itemsets, Knowledge-Based Systems 23 (2010) 379–388.
[46] Chun-Ling Chen , Frank S.C. Tseng b, Tyne Liang, Mining fuzzy frequent itemsets for hierarchical document clustering, Information Processing and Management 46 (2010) 193–211
[47] Dong Xin Jiawei Han Xifeng Yan Hong Cheng, Mining Compressed Frequent-Pattern Sets, Proceedings of the 31st international conference on Very large data bases, Pages 709 – 720, 2005.
[48] James Cheng Yiping Ke Wilfred Ng, A survey on algorithms for mining frequent itemsets over data streams, Knowledge and Information Systems, Volume 16 Issue 1, July 2008.
[49] Mohamed Medhat Gaber, Arkady Zaslavsky and Shonali Krishnaswamy, Mining Data Streams: A Review, ACM SIGMOD Record Volume 34 Issue 2, June 2005 ,Pages 18 - 26
[50] Jiawei Han. Mining Useful Patterns: My Evolutionary View. Keynote talk at the Mining Useful Patterns , workshop KDD 2010
[51] Hoang Thanh Lam, Fabian Moerchen, Dimitry Fradkin, and Toon Calders. Mining compressing sequential patterns. In: SIAM Data Mining Conference 2012
[52] Qiaozhu Mei, Dong Xin, Hong Cheng, Jiawei Han, ChengXiang Zhai, Generating Semantic Annotations for Frequent Patterns with Context Analysis, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining Pages 337-346 , 2006.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 13, Issue 2 (Jul. - Aug. 2013), PP 14-24*
*www.iosrjournals.org*

# Persuasive Cued Click Based Graphical Password with Scrambling For Knowledge Based Authentication Technique with Image Scrambling.

BINITHA . V .M.
*Computer Science and Information Systems. Department Of Computer Science, Federal Institute Of Science And Technology, Mahatma Gandhi University, Kerala, India*

***Abstract:*** *Adequate user authentication is a persistent problem, particularly with hand- held devices such as Personal Digital Assistants (PDAs), which tend to be highly personal and at the fringes of an organization's influence. Yet, these devices are being used increasingly in corporate settings where they pose a security risk, not only by containing sensitive information, but also by providing the means to access such information over wireless network interfaces. User authentication is the first line of defense for a lost or stolen PDA. How- ever, motivating users to enable simple PIN or password mechanisms and periodically update their authentication information is a constant struggle. This paper describes a general-purpose mechanism for authenticating a user to a PDA using a visual login technique called Picture Password. The underlying rationale is that image recall is an easy and natural way for users to authenticate, removing a serious barrier to compliance with organizational policy. Features of Picture Password include style dependent image selection, password reuse, and embedded salting, which overcome a number of problems with knowledge-based authentication for handheld devices. Though designed specifically for handheld devices, Picture Password is also suitable for note-books, workstations, and other computational devices. Scrambling technique is applied to make image recognition more complex during the login process and thus protecting from the common attacks in the graphical password system.*

***Keywords:*** *Graphical Passwords, Security, Image Scrambling,KBRP.*

## I. Introduction

Computer security depends largely on passwords to authenticate human users. One of the key areas in security[1] research and practice is authentication, the determination of whether a user should be allowed access to a given system or resource. However, users have difficulty to remembering pass-words over time if they choose a secure password, i.e. a password that is long and random. Therefore, they tend to choose short and insecure pass-words. The continued domination of passwords over all other methods of end-user authentication is a major embarrassment to security researchers. As web technology moves ahead by leaps and bounds in other areas, passwords stubbornly survive and reproduce with every new web site. Extensive discussions of alternative authentication schemes have produced no definitive answers. A password authentication system should encourage strong passwords while maintaining memorability. We propose that authentication schemes allow user choice while influencing users toward stronger passwords. In our sys-tem, the task of selecting weak passwords (which are easy for attackers to predict) is more tedious, discouraging users from making such choices. In effect, this approach makes choosing a more secure password the path of least resistance. Rather than increasing the burden on users, it is easier to follow the systems suggestions for a secure passwords feature lacking in most schemes replace text passwords for general-purpose user authentication on the web using a broad set of twenty five usability, deployability and security benefits that an ideal scheme might provide. To validate the end user for authentication we usually prefer to adopt the knowledge-based authentication, which involves text based passwords. The text based passwords are vulnerable to be hacked. The attackers can easily guess the text passwords with other details of the system. If we want to avoid this, the system can assign a strong password, which the attacker cannot guess. But the system assigned passwords are very di cult to memorize and remembered by the user. The study on the graphical passwords states that the click point passwords are hard to guess by the attacker and easy to remember for the users. So the password authentication system should encourage the strong password selection while maintaining the memorability of the user. This paper proposes the idea of persuasive cued click point authentication[2,3] with the technique of scrambling. This scheme influence the user to set a number of clicks from a picture and size of passwords needed. The user can also change his passwords during a week oe everyday with altered images. This scheme fully depended on the memorability of the user about his selected images. Once he could not remember which portion of the image he selected for the click, the user will not authenticate even though he is a genuine user. To overcome this kind of

problem the system should keep some policies to retain the passwords.

**2.1 BACKGROUND**

The community of security researchers and practitioners has evolved rapidly in response to threats, on the one hand increasing vigilance in practice and, on the other hand, driving research innovation. Until recently the security problem has been formulated as a technical problem. Even though text passwords are the most popular user authentication method, they have security and us-ability problems. The alternatives for text based passwords such as biometric systems and tokens have their own drawbacks .Graphical passwords, which consist of clicking on images rather than typing alphanumeric strings, may help to overcome the problem of creating secure and memorable passwords. A graphical password scheme using click point offers the best alternative for the text password, cued click points are used to exploit the memeorablity of the user that it is fully a knowledge based authentication and is discussed in this paper the security and usability problems associated with alphanumeric pass-words as the password problem. The problem arises because passwords are expected to comply with two conflicting requirements, namely

(1) Passwords should be easy to remember, and the user authentication protocol should be executable quickly and easily by humans

(2) Passwords should be secure, i.e. they should look random and should be hard to guess; they should be changed frequently, and should be different on different accounts of the same user; they should not be written down or stored in plain text.

(3)The password problem arises primarily from fundamental limitations of human long-term memory (LTM).

Once a password has been chosen and learned the user must be able to recall it to log in. However, people regularly forget their passwords.

## II. Introduction To PassPoints

PassPoints[4], a new and more secure graphical password system.This work proposed a password scheme in which the user is presented with a predetermined image on a visual display and required to select one or more predetermined positions (tap regions) on the displayed image in a particular order to indicate his or her authorization to access the resource. Beyond this .This system was developed early in the evaluation of graphical passwords, and in this, the user is given with an image. The click points on the image are used as the password for user authentication. The user has to remember the order and position of the click points. The click points are not stored as such, but as a hashed value. For correct validation, discretization square is used which is the tolerance area around the original click point. The user should click on the discretization area. Here, the system does not have any influence over the selection of the click points. The user is free to set the password which the user can easily remember. Since it is being very simple, it can easily be attacked. In PassPoints, passwords consist of a sequence of click-points on a given image. Users may select any pixels in the image as click-points for their password.

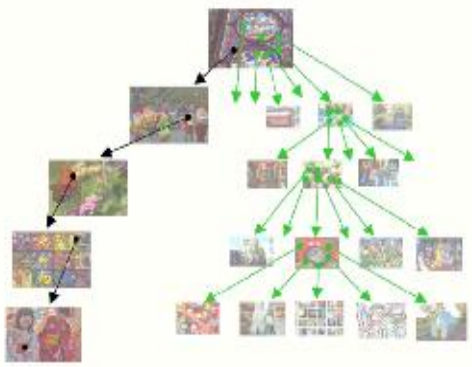


Figure 1 User Navigation through Clicks

To log in, they repeat the sequence of clicks in the correct order, within a system-defined tolerance square of the original click-points. The hypothesize that users will choose clickpoints based on their preference for certain points in the image, and that their preference for certain points will be influenced by how much they are naturally attracted to those points. Attention is the cognitive process of selectively fo cusing on one aspect of the environment while ignoring others, a mechanism that helps us prioritize sensory information. There are two different categories of visual attention models: bottom-up and top-down. Bottom-up visual attention captures

how attention is drawn to the parts of a scene or image that are salient or conspicuous. It is what naturally draws us to lo ok at the unexpected or different parts of a scene, prioritizing them from the other consistent parts. For example, if an image contains a large number of objects that are blue, and only one is yellow, human attention will instinctively focus on the yellow object. Top-down visual attention is task-dependent, based on cognitive, volitional control. With a priori knowledge about what object(s) to look for, our attention is brought to the parts of the scene containing


Figure 2 Passpoints

Those object(s). For example, if a user decides that people with dark hair are of interest for some reason, the users attention would shift between objects with features that might indicate a dark-haired person. In the PassPoints graphical password scheme a password consists of a sequence of click points (say 5 to 8) that the user chooses in an image. The image is displayed on the screen by the system. The image is not secret and has no role other than helping the user remember the click points. Any pixel in the image is a candidate for a click point.

To log in, the user has to click again closely to the chosen points, in the chosen sequence. Since it is almost impossible for human users to click repeatedly on exactly the same point, the system allows for an error tolerance r in the click locations (e.g., a disk with radius r = 10 or 15 pixels). This is done by quantizing (discretizing) the click locations, using three di erent square grids, as described in [3]. Each grid has width 6r between grid lines (horizontal or vertical). Each one of the three grids is staggered with respect to the previous grid by a distance 2r vertically and a distance 2r horizontally. If there were only one quantization grid then a

Figure 3(a)Actual click

Figure 3(b) Predicted Click in Passpoints

Selected click point could be close to a grid line and small variations in the user's clicking could lead to a click in a different grid square, thus leading to the wrong password. On the other hand, one can prove that with the three staggered grids every point in a two dimensional image is at distance at least r from the grid lines of at least one of the three grids; we say that the point is safe in that grid. We pursue heuristic-based strategies for purely automated dictionary generation (e.g., based on click-order patterns), and strategies to prioritize these dictionaries using image processing methods to identify points that users are more likely to choose.

**3.1 Cued Click Points (CCP)**

Cued Click Points [2] [3] [5] was designed to reduce patterns and to reduce the usefulness of hotspots for attackers. Instead of five click-points on one image, CCP uses one click-point on five different images. The next image displayed is based on the location of the previously entered click-point; it creates a path through an image set. Creating a new password with different click-points results in a different image sequence. One best feature of Cued Click Point is that the explicit indication of authentication failure is only provided after the final click-point, to protect against incremental guessing attacks. The cued click point method uses a series of images for click point password creation. The position of the click point on the previous image decides the next image to appear. It offers cued-recall and introduces visual cues that instantly alert valid users if they have made a mistake when entering their latest click-point (at which point they can cancel their attempt and retry from the beginning). The image used has the size 451x331 pixels and a tolerance square of 19x19 pixels. The candidate image or image, thus have approximately 400 squares. To have better discretization, 3 overlapping squares are assigned. So, in a candidate grid there could be 1200 squares. If a click on the first image is correct (by considering the tolerance squares), the user gets the next correct image.

Once the user practiced with the usage of click point password, user can readily understands when he/she clicks the wrong point, by looking at the next image. In this scheme also user is free to select the graphical password without systems intervention. So the attackers can easily guess the hot spot, which is the area where most of the users will tend to click. If the hacker[2] is succeeded in guessing the hot spots in the images then the hacker can log in to the system easily.

**3.1.1 Persuasive Technology**

Persuasive Technology [2] used to motivate and influence people to behave in a desired manner. An authentication system which applies Persuasive Technology should guide and encourage users to select

Stronger passwords, not the system-generated passwords[6]. Even though the users are guided, the resulting passwords must be memorable. This persuasion makes the password stronger by avoiding the hot spots in almost all the cases. The click points are more randomly scattered to avoid the correct guess by the attackers. The users must not ignore the persuasive elements and the resulting passwords must be memorable. As detailed below, PCCP accomplishes this by making the task of selecting a weak password more tedious and time consuming. The path of least resistance for users is to select a stronger password (not comprised entirely of known hotspots or following a predictable pattern). The formation of hotspots across users is
minimized since click-points are more randomly distributed.

**3.2 Persuasive Cued Click Points (PCCP)**

Using a skewed password distribution the attackers can guess the password in the previous graphical password schemes. Without the system guidance most of the users clicks on the hotspot in each image. In this method the system influence the user to select more random clicks, and also maintains the user memorability. In this scheme when the image is displayed the randomly selected block called the view port only clearly seen out. All the other parts of the image are shaded, so that the user can click only inside the view port. This is how the PCCP influence the user to select the position of the click point. The view ports are selected by the system randomly for each image to create a graphical password. It will be very hard for the attackers to guess the click point in all the images.


Figure 4 User interface of PCCP

Users are allowed to click anywhere in the view port. There is an option for changing the viewport position also. This option is called the Shuffle. There is a limit on the number of times the shuffle option to be used. While users may shuffle as often as desired, this significantly slows password creation. The viewport[1] and shuffle button appear only during password creation. Figure 4 User interface of PCCP During later password entry, the images are displayed normally, without shading or the viewport, and users may click anywhere on the images. Like PassPoints and CCP, login click-points must be within the defined tolerance squares of the original points. The theoretical password space for a password system is the total number of unique passwords that could be generated according to the system specifications. Ideally, a larger theoretical password space lowers the likelihood that any particular guess is correct for a given password. Whereas text passwords have very skewed distributions resulting in an effective password space much smaller than the theoretical space, PCCP is specifically designed to significantly reduce such skews. The recall studies of the PCCP approach proved that remembrance of the graphical password is much better than the text-based passwords.

## III. Scrambling

With the current development of ubiquitous wireless network technology and digital multimedia devices, wireless image/video data transmission is becoming more prevalent. As a result, information security becomes a key problem for consumers, companies and governments. Security of image and video data is very important in many areas, such as video-on-demand, confidential remote video conferencing, security communication, and also in military applications. Image scrambling (i.e., encryption) technologies are very useful tools to ensure image security by transforming the image into an unintelligible image[7]. Scrambling makes the image unrecognizable to prevent eavesdroppers from decoding the true form or meaning of the image using the human visual system or a computer system. Image scrambling [8] is a useful approach to secure the image data by scrambling the image into an unintelligible format. This paper introduces a new parameter based M-sequence which can be produced by a series shift registers. There are currently several techniques to perform the image scrambling. In addition, a new image scrambling algorithm based on the M-sequence is presented. Image scrambling is used to make images visually unrecognizable such that unauthorized users have difficulty decoding the scrambled image to access the original image. This article presents two new image scrambling algorithms based on Fibonacci p-code, a parametric sequence. The first algorithm works in spatial domain and the second in frequency domain (including JPEG domain). A parameter, p, is used as a security-key and has many possible choices to guarantee the high security of the scrambled images. The presented algorithms can be implemented for encoding/decoding both in full and partial image scrambling, and can be used in real-time applications, such as image data hiding and encryption. Examples of image scrambling are provided. Computer simulations are

This has shown to demonstrate that the presented methods also have good performance in common image attacks such as cutting (data loss), compression and noise. The new scrambling methods can be implemented on grey level images and 3-color components in color images. A new Lucas p-code is also introduced. The scrambling images based on Fibonacci p-code are also compared to the scrambling results of classic Fibonacci number and Lucas p-code.

Two new image scrambling algorithms based on Fibonacci p-code. One is working in spatial domain, the other is for frequency domain (including JPEG domain). The security keys of our image scrambling algorithms are parameters p and i, and the size of original image. There are many possible choices for security keys so that the scrambled image is di cult to decrypt by unauthorized users, and thus, greater security is

guaranteed. A new Lucas p-code is also introduced. The scrambling images obtained from Fibonacci p-co de are compared to the scrambling results of classic Fibonacci number and Lucas p-code. This will demonstrate that the classical Fibonacci number is a special sequence of Fibonacci p-code when
p=1. Additionally, this will show the di erence of scrambling results by using the Fibonacci p-code and Lucas p-code.

**4.1 P-Fibonacci And P-Lucas Transform**

Fibonacci p-code [9] and a new Lucas p-code are introduced in this section. A new 1-D transform and a new 2-D transform are generated for both Fibonacci p-code and Lucas p-co de. The inverse 2-D transform used for recovering the original image is also presented

Definition : The Fibonacci p-code[10,11] is a sequence defined by,

$$F_p(n) \begin{cases} 0 & n < 1 \\ 1 & n = 1 \\ F(n \quad 1) \; F(n \quad p \quad 1) & n > 1 \end{cases}$$

where p is a nonnegative integer. From the definition above, Fibonacci p-code sequences will di er based on the p value. Specially,
(1) Binary sequence: p=0, the sequence is powers of two, 1, 2, 4, 8, 16……etc
(2) Classical Fibonacci sequence: p=1, the sequence is 1, 1, 2, 3, 5, 8, 13, 21……etc
(3) For the large values of p the sequence starts with consecutive 1?s and immediately after that 1, 2, 3, 4 ...p Sample sequences are shown in Table

| p \ n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | … |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1024 | … |
| 1 | 1 | 1 | 2 | 3 | 5 | 8 | 13 | 21 | 34 | 55 | 89 | … |
| 2 | 1 | 1 | 1 | 2 | 3 | 4 | 6 | 9 | 13 | 19 | 28 | … |
| 3 | 1 | 1 | 1 | 1 | 2 | 3 | 4 | 5 | 7 | 10 | 14 | … |
| 4 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 4 | 5 | 6 | 8 | … |
| … | | | | | | | | | | | | |
| ∞ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | … |

Table 1 Fibnacci p-code sequence with different p value

**4.2 Image Scrambling Algorithm In The Spatial Domain**

The presented image scrambling algorithm in the spatial domain (shown in Figure 4.2) is designed to change the image pixel position using the 2-D P-Fibonacci Transform. Color images have three color components and the scrambling algorithm is applied to each color component individually. Grayscale images are treated as color images with one component. The presented algorithm is a lossless image scrambling method The Detailed description of the algorithm explained below for scrambling and unscrambling of images.

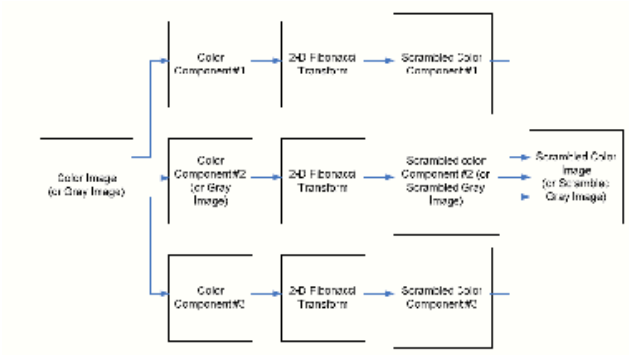



Figure 5 Block diagram of spatial domain scrambling

## Algorithm for image scrambling

*Step1:Choose the key parameter p, Calculate the row and column coefficient matrices of 2D p-Fibnaacci Transform*

*Step2:separate the 2D color image to three color component .Each component is a 2D matrix.*

*Step3: Apply 2D P-Fibnaacci transform to each color component to set the scrambled color component.*

*Step4 Recombine the three scrambled components to get the scrambled Image for password selection*

The above algorithm says how to scramble the given digital images in spatial domain.

### 4.3 KEY BASED RANDOM PERMUTATION (KBRP)

A permutation, also called an "arrangement number" or "order," is a rearrangement of the elements of an ordered list ***S*** into a one-to-one correspondence with *S* itself. The number of permutations on a set of *n* elements is given by ***n!*** (***n*** factorial) A random permutation is a permutation containing a fixed number ***n*** of a random selection from a given set of elements. There are two main algorithms for constructing random permutations. The first constructs a vector of random real numbers and uses them as keys to records containing the integers 1 to ***n***. The second starts with an arbitrary permutation and then exchanges the *i*th element with a randomly selected one from the first *i* elements for *i* = 1, ..., ***n.*** Key Based Random Permutation (**KBRP**) is a method that can generate one permutation of size n out of ***n!*** permutations. This permutation is generated from certain key (alphanumeric string) by considering all the elements of this given key in the generation process. The permutation is stored in one-dimensional array of size equal to the permutation size (***N***).

This technique is used to generate the row and column coefficient matrices of each image components.

### 4.4 IMPLEMENTATION ASPECTS

Image based mutual authentication [12] has become now more reliable, when the scrambling technique applied [7].Users will me users should keep the click points to enter into the system, because the image get scrambled and it will be rearranged according to the scrambling algorithm discussed above. User have the provision to select his favorite areas according to his interest.

For any password authentication scheme,the prime task is to become a valid user of that system. For performing this each user have to provide the user id and password for creating the account just like in the conventional (textual) login system by specifying the username and password. This is for keeping an entry in the administrative level for further use for checking the intended user is authenticated or not. When a new user is intended to become a valid user ,the user have to select the new user and proceed. On the way to registration it will ask the userid and password,and the user should provide it through textual passwords. Now the user is entering to the PCCP System, here the textual password is replaced by the graphical password via clickpoints (cued). Hence the user have to select the decide how many click points needed to create the password and it will effect the strength of the password security. In order to improve the total security strength of the target system the number of click points used can also be increased while creating the graphical passwords. This can be achieved by setting the number of click point to be received from the user as a predefined value, say v. A number of view ports, which is equal to v are made visible on the image, for the user to click on it.

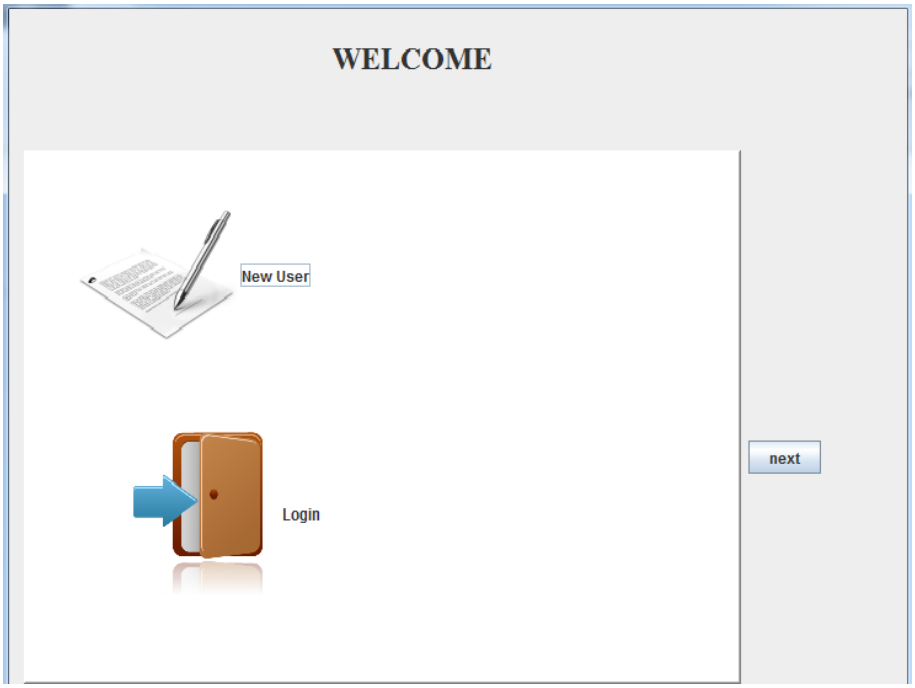



Fig 6 . New User SignUp : GUI

### 4.5 View port size

The effective password space is determined by the area of the view port of all images displayed for the password creation. The password strength is increased with the password space. So to create a strong graphical password, which cannot be guessed easily, the area of the view port should be higher. It can be done by deciding how many times the we can select the shuffle button which is directly proportional to the maximum number of viewports possible for an image. Then the number of click points also effecting as a predominant factor for ensuring the security. This idea may increase the strength of the password but this will decrease the user memorability of the password.


Fig 7 .Actual image for password Creation

### 4.5 Discretization of view port

In some occasions the user may accidentally click the point which is very near to the viewport, while logging in. If the user is genuine then he/she must be correctly logged in. Since we follow a very strict validation method, which requires the user to click on the view port, the genuine user cannot be allowed to use the application. To avoid this situation, we can compute the discretization are for the view port displayed on

each image. The user clicks are tolerated up to the discretization area. But this may reduce the robustness of the system.

**4.6 Authentication of a valid user**

The user after registration process have to memorise the click points what he selected to make the password.

The basis of PCCP starts from here. If he is a genuine user and he could not memorise the cues for click points, he cannot enter into the system. The system will treat him as an unauthorised user. This is the strength of PCCP. It is fully exploiting the memory and thus protecting your devices like PDA's from unauthorised access and other different kinds of attacks. Thus it termed as a knowledge based password authentication scheme in which the cues leads to the validating/invalidating session .Until the user selecting his last click points the system will not remind the user whether he given the right click or not even if he is a genuine user. This will help to protct the system from shoulder surfing attack and dictionary attack.

For login process the user have to enter the textual username and then he is entering to the PCCP system. System will allow only the valid username will enter into the PCCP system . There he starts accessing the images and starts clicking the clickpoints according to the order what he have received for password creation phase. Since the order is an essential property of PCCP the user have to ensure he is accessing the right images for password selection.

Here the scrambling is applied .While in the login session the user is receiving the scrambled images of the actual image what he selected for password creation.


Fig. 8 Scrambled Image for password selection in Login Phase

Here the cues are very important factor, because this will help the users to remember easily . The scrambling process is done by the algorithm shown above and the row and column coefficients are determined by the Key Based random Permutation (KBPR) explaned in section 4.3. There is a small comparison of alphanumeric password and graphical password is shown in the figure below with different parameters.

| | Image size | Grid square size (pixels) | Alphabet size/ No. squares | Length/No. click points | Password space size |
|---|---|---|---|---|---|
| Alphanumeric | N/A | N/A | 64 | 8 | $2.8 \times 10^{14}$ |
| Alphanumeric | N/A | N/A | 72 | 8 | $7.2 \times 10^{14}$ |
| Alphanumeric | N/A | N/A | 96 | 8 | $7.2 \times 10^{15}$ |
| Graphical | $451 \times 331$ | $20 \times 20$ | 373 | 5 | $7.2 \times 10^{12}$ |
| Graphical | $1024 \times 752$ | $20 \times 20$ | 1925 | 5 | $2.6 \times 10^{16}$ |
| Graphical | $1024 \times 752$ | $14 \times 14$ | 3928 | 5 | $9.3 \times 10^{17}$ |
| Graphical (1/2 screen used) | $1024 \times 752$ | $14 \times 14$ | 1964 | 5 | $2.9 \times 10^{16}$ |

Fig Comparison of textual and graphical password.

## V. Security Analysis

In this section a discussion on how the proposed system may behave for password guessing attack and capture attack.

### 5.1 Password guessing attack

The most basic guessing attack against PCCP is a brute force attack, with expected success after exploring half of the password space (i.e., with a theoretical password space of $2^{43}$, success after $2^{42}$ guesses). However, skewed password distributions could allow attackers to improve on this attack model. We now consider how these could be leveraged in guessing attacks. PassPoint system hotspots of small number of users can be collected and an attack dictionary can be formed, with the use of server-side information. Then this dictionary details can be used for the guessing of the click point in an image. But this does not work in PCCP with Image scrambling scheme, because the view port is entirely changing during the scrambling phase., and so it does not include the hot spot in almost all cases. If the attackers gain the access to hash table entry of the passwords, they cannot correctly predict the original password, which are kept in a different data base .which can be encrypted also using any of the strongest encryption scheme.

### 5.2 Capture attacks

Password capture attacks occur when attackers directly obtain passwords (or parts thereof) by intercepting user entered data, or by tricking users into revealing their passwords. For systems like PCCP, CCP, and PassPoints (and many other knowledge-based authentication schemes), capturing one login instance allows fraudulent access by a simple replay attack. All three security schemes (PP, CCP, PCCP) are vulnerable to shoulder surfing threat. Observing the approximate location of click points may reduce the number of guesses necessary to determine the user's password. User interface manipulations such as reducing the size of the mouse cursor or dimming the image may offer some protection, but have not been tested.

Malware is a major concern for text and graphical passwords, since key logger, mouse logger, and screen scraper malware could send captured data remotely or otherwise make it available to an attacker. For social engineering attacks against cued-recall graphical passwords, a frame of reference must be established between parties to convey the password in sufficient detail. One preliminary study suggests that password sharing through verbal description may be possible for PassPoints. For PCCP with image scrambling , more effort may be required to unscramble and get the actual picture during the login phase, each image and the exact location of each click-point. Graphical passwords may also potentially be shared by taking photos, capturing screen shots, or drawing, albeit requiring more effort than for text passwords.

## Acknowledgements

This paper was developed by having [2] as the base idea and lab studies done the implementation of [1] are taken as proof for this paper. Sonia Chiasson and her friends, Members of IEEE, are honorably well acknowledged here for their fruit full research work

## VII. Conclusion

The current graphical password techniques are still immature. Much more research and user studies are needed for graphical password techniques to achieve higher levels of maturity and usefulness. Two new image scrambling algorithms based on Fibonacci p-code are presented in this article: spatial domain and frequency domain algorithms (including JPEG domain).More Experimental results are needed on both color and grayscale images verify that this algorithms are lossless and show good performance in the presence of common image attacks. This algorithm introduces the technique to avoid the hotspot problem, reduces the shoulder attack. Also exploit the usability, memorability in graphical password scheme. Much more results are needed to show the effectiveness of the algorithm in 3D images.

## References

[1] The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes,Joseph Bonneau University of Cambridge Cambridge, UK jcb82@cl.cam.ac.uk Cormac Herley Microsoft Research Redmond, WA, USA cormac@microsoft.com Paul C. van Oorschot Carleton University Ottawa, ON, Canada paulv@scs.carleton.ca Frank Stajanoy University of Cambridge Cambridge, UK frank.stajano@cl.cam.ac.uk

[2] Persuasive Cued Click-Points: Design, Implementation, and Evaluation of a Knowledge-Based Authentication Mechanism Sonia Chiasson, Member, IEEE, Elizabeth Stobert, Student Member, IEEE, Alain Forget, Robert Biddle, Member, IEEE, and Paul C. van Oorschot, Member, IEEE.IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 9, NO. 2, MARCH/APRIL 2012

[3] Persuasive Cued Click Points with Click Draw Based Graphical Password Scheme P. R. Devale Shrikala M. Deshmukh, Anil B. Pawar.

[4] Purely Automated Attacks on PassPoints-Style Graphical Passwords Paul C. van Oorschot, Amirali Salehi-Abari, and Julie Thorpe

IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 5, NO. 3, SEPTEMBER 2010

[5] S. Chiasson, P. van Oorschot, and R. Biddle, ?Graphical Password Authentication Using Cued Click Pointsm Pro c. European Symp. Research in Computer Security (ESORICS), pp. 359-374, Sept. 2007.

[6] The science of guessing: analyzing an anonymized corpus of 70 million passwords Joseph Bonneau Computer Laboratory University of Cambridge
jcb82@cl.cam.ac.uk. 2012 IEEE Symposium on Security and Privacy

[7] Dimitri Van De Ville, W.P., Rik Van de Walle, Ignace Lemahieu, Image Scrambling Without Bandwidth Expansion. IEEE Transactions on Cirsuits and Systems for Video Technology, 2004. 14

[8] AN IMAGE SCRAMBLING ALGORITHM USING PARAMETER BASED M-SEQUENCES YICONG ZHOU1, KAREN PANETTA1, FELLOW, IEEE, SOS AGAIAN2, SENIOR MEMBER, IEEE.

[9] Guosheng Gu, g.H. The application of chaos and DWT in image scrambling. in Proceeding of the Fifith Interational Conference on Machine Learning and Cybernetics. 2006. Dalian.

[10] S. Agaian, J.A., K. Egiazarian, P. Kuosmanen, Decompositional methods for stack filtering using Fibonacci pcodes. Signal Processing, 1995. 41: p. 101-110.

[11] David Z. Gevorkian, K.O.E., Sos S. Agaian, Parallel Algorithms and VLSI Architectures for Stack Filtering Using Fibonacci p-Codes. IEEE Transactions on Signal Processing, 1995. 43(1): p. 286-295.

[12] Mutual Image-Based Authentication Framework with JPEG2000 inWireless Environment G. Ginesu, D. D. Giusto, and T. Onali MCLab, Department of Electronic Engineering, University of Cagliari, Cagliari 09123, Italy

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 13, Issue 2 (Jul. - Aug. 2013), PP 25-32*
*www.iosrjournals.org*

# Design of algorithm for detection of hidden objects from Tera hertz images

P.Vijayalakshmi[1], M.Sumathi[2]
[1] *(Department of Information Technology, Pandian Saraswathi Yadav Engineering College, Sivagangai TN,India)*
[2] *(Department of Computer Science, Sri Meenakshi Govt. Arts College for Women, Madurai, TN, India)*

***Abstract:*** *An algorithm for detection of hidden objects from tera hertz images is presented. Presently, terahertz imaging employs object's radiometric temperatures in human to acquire images of concealed objects. But, it presents problem in temperature sensitive areas like oil and coal mines, factories etc. The aim of the paper is to detect and extract hidden objects underneath person's clothing. Here, a three stage approach is presented: In the first stage, edge based segmentation is applied after smoothing the image using bilateral filter. In the next stage, transform invariant shape descriptors, Gabor and gray level co-occurrence (GLCM) texture features of interested object regions are computed. Finally, a Euclidean distance criterion is used for classification. To appraise the technique, detection error and detection rate are calculated. Test results are compared with ground truth data obtained from the original image. Experiment results are found to be promising with 1.04% detection error and 91.9% as detection rate. Potential applications in security include detection of weapons and explosive in public places like airports, stations etc.*

***Key words:*** *Edge detection, GLCM features, Gobor features, Shape descriptors, THZ images.*

## I. INTRODUCTION

In the recent years, increased threats of criminal action have led to the development of many techniques for the detection of concealed weapons, explosives etc. They include metal detectors, X-ray scanners, and detection of explosive and are based on energetic radiation. Electromagnetic waves at Tera hertz frequencies are safe, penetrate barriers, have short wave lengths to allow discrimination between objects. They have unique reflection and absorption properties. Also, many explosives have characteristic signatures at terahertz wavelengths. A survey of object recognition/classification methods based on image moments was presented. They had reviewed various types of moments and moment-based invariants. They had studied the role for various image degradations and distortions that affects the shape descriptors for classification. They had reviewed numerical algorithms for moment computation in real applications [1]. Another paper had explained millimeter wave (MMW) sensors designed for detecting both metallic and nonmetallic objects placed on a human body and hidden under clothes. The sensor was based on the synchronized detection principle. It had estimated the power of back-scattered signal from hidden objects. Time-gating algorithm combined with threshold level was implemented to detect hidden objects at the distance [2].Detection and segmentation of concealed objects in Terahertz Image was presented. Standard segmentation algorithms were unable to segment or detect concealed objects. A two stage approach was presented. First, the noise from the image was removed using the anisotropic diffusion algorithm and then detected the boundaries of the concealed objects. A mixture of Gaussian model was used to study the distribution of the temperature inside the image to identify the concealed objects [3].A inter ferro metric imaging and sensing for the detection of explosives, weapons and drugs was explored. It worked with three objectives: (a) THz radiation to detect concealed weapons (b) to detect compounds such as explosives and illicit drugs that have characteristic THz spectra [4].Detection and identification of explosive RDX by Diffuse Reflection Spectroscopy was presented. The reflection spectrum of the explosive RDX was acquired from a diffuse reflection measurement using a THz time-domain spectroscopy system. Kramers-Kronig transform was applied over the reflection spectrum to obtain the absorption spectrum. The investigation demonstrated that THz technique could be capable of detecting and identifying hidden RDX-related explosives [5].

The article was presented to provide a tutorial overview of developments in Tera hertz imaging. It was to detect concealed weapons from a standoff distance, especially where the flow of people could not be controlled. It' was a technological challenge that required innovative solutions in sensor technologies and image processing. A number of sensors based on different phenomenon as well as image processing techniques were to be developed to observe objects underneath people are clothing [6]. .A technology with a warning system was presented. It educated the customer about the use of technology and its applicability. It strongly recommended that potential customers to trial the technology in their own unique environments to determine the utility of this technology and for its adaptability to environmental pressures. The false alarms and missed detections that

occurred were easily accommodated. [7]. Object detection is a critical part in many applications like image search, image understanding and scene analysis. However, still it is an open problem due to complexity of object classes and images. Current approaches used for object detection are top down and bottom up. Top down approaches include training stage and to define object configurations. The later approach includes low level image features to high level image features. Here, we have included both the approaches for detection of concealed objects [8],[9].Recent developments in the area of concealed weapon detection using electromagnetic methods including metal detection, magnetic field distortion, electromagnetic resonance, acoustic and ultrasonic inspection, millimeter waves, Terahertz imaging, Infrared, X-ray were reviewed. The advantages and disadvantages of various approaches were discussed. Research challenges were presented. Future research perspectives were in the above areas was analyzed [10].

A 35GHz imager was presented that was based on conical scan technology from low cost materials. In conjunction with an illumination chamber it was used to collect indoor images of people with weapons and illegal imports hidden under their clothing. That imager had a spot size of 20mm and covered a field of view of 20 x 10 degrees that partially covered the body of knees to shoulders. A variant imager was designed and constructed. It had a field of view of 36 x 18 degrees and was capable of covering the whole body of an adult. That was achieved by increasing the number of direct detection receivers by implementing an improved optical design. The optics system consisted of a front grid, a polarization device which converted linear to circular polarization and a rotating scanner [11].THz time domain spectroscopy (THz-TDS) was presented. It was used to investigate explosives and establish a spectra database of explosive materials in the THz frequency range. Signatures of selected explosives and related compounds (ERCs) were identified in the THz band and maintained. THz spectra of ERCs were calculated based on Density Functional Theory [12]. The progress that had been made in Tera hertz technology over the years was reviewed. It identified the achievements, challenges and prospects in millimeter wave and terahertz technology for specialized screening tasks. It explored that many solids including explosives have characteristic spectroscopic signatures at terahertz wavelengths that were used to identify them [13].Detection of biological hazards using electronic Terahertz systems was presented. The discriminating sense of vulnerability to concealed threats by weapons or biological agents such as anthrax had led the scientific community to address the problem of detecting such concealed threats [14]. A survey of object recognition/classification methods based on image moments was presented. They had reviewed various types of moments and moment-based invariants. They had studied the role for various image degradations and distortions that affects the shape descriptors for classification. They had reviewed numerical algorithms for moment computation in real applications [15].

## II. PROPOSED METHOD

Initially, images are processed to compensate for losses due to noise and other variations. In general, objects are featured in images by its color, edge, and shape and texture information. In this paper, three main attributes of object are considered at various spatial resolutions to detect any hidden object. First, the search image is divided into number of overlapping grids to reduce the missing probability. Both region-based and contour based shape descriptors are computed to distinguish shapes of different objects. Gobor, GLCM features are extracted as texture information to characterize objects. For robust detection, a combined feature vector of edge, shape and texture is employed. Test feature vector is calculated and compared with feature vectors of the search image using Euclidean distance classifier. To appraise the algorithm, detection rate and detection error are measured against ground truth data. Experiments are repeated both with combined and individual feature vector for gun, knife and needle images.

### 2.1. Shape Features

Shape descriptors are mathematical functions that are applied on images to produce numerical values. These numerical values are processed to provide information about objects. Two dimensional shape descriptors are of two categories, region-based descriptors and contour-based descriptors. Region-based descriptors characterize spatial distribution of pixel intensities and they include pixels in boundary and interior. Region-based shape descriptors describe complex objects having multiple disconnected regions, simple objects with or without holes. Contour based shape descriptors are transform invariant and robust to noise. Moments invariants are more useful in shape analysis and they are used in distinguishing shape between different objects. A closed boundary is characterized by an ordered sequence $z(i)$ that represents the Euclidean distance between centroid and all $N$ bounding pixels of the digitized shape. Here, translation, rotation and scale invariant normalized contour sequence moment $\overline{m_r}$ and contour sequence central moment $\overline{\mu_r}$ are calculated for shape representation. Also, computation time is much less for spiral and concave shapes. Classification by using contour sequence moments comparatively good over area based moments. Shape descriptors $F_1, F_2, F_3$ are calculated using (2.1), (2.2) and (2.3).

$$F_1 = \frac{(\mu_2)^{\frac{1}{2}}}{m_1} \quad \ldots\ldots \ \ldots \ldots \ (2.1)$$

$$F_2 = \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}} \quad \ldots\ldots\ldots\ldots \ldots \ (2.2)$$

$$F_3 = \frac{\mu_4}{(\mu_2)^2} \quad \ldots\ldots\ldots\ldots\ldots \ (2.3)$$

$F_1, F_2, F_3$ - Shape descriptors of regions

$m_1$ - First order moment

$\mu_2, \mu_3, \mu_4$ -- Second, third, fourth order central moments

## 2.2. Texture Features

A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G, in a image. The matrix element P(i, j | d, θ) is the relative frequency with which two pixels, separated by distance d, and in direction specified by the particular angle (θ). Here, gray level co-occurrence based texture features are computed in two steps. In the first step, pair wise spatial co-occurrences of pixels separated by a particular distance are charted by a gray level co-occurrence matrix (GLCM). Then using the GLCM, a set of texture features of interested object regions is computed. Here, entropy, correlation, energy, contrast and homogeneity features are computed. Energy content is computed using: (2.4) and is the sum of squared elements in the GLCM. Entropy represents the randomness of intensity distribution and is computed using: (2.5). Contrast is computed using: ( 2.6) that presents the amount of local variation present in the image . Correlation reveals the linearity present in the image and is measured using : (2.7). Homogeneity measures the closeness of the distribution of elements in the GLCM and is evaluated using: (2.8).

$$Energy = \sum_{m=0}^{G-1}\sum_{n=0}^{G-1} p(m,n)^2 \quad \text{---------- (2.4)}$$

$$Entrophy = \sum_{m=0}^{G-1}\sum_{n=0}^{G-1} p(m,n)\log p(m,n) \quad \text{---------- (2.5)}$$

$$Contrast = \frac{1}{(G-1)^2}\sum_{m=0}^{G-1}\sum_{n=0}^{G-1}(m-n)^2 p(m,n) \quad \text{---------- (2.6)}$$

$$Correlation = \frac{\sum_{m=0}^{G-1}\sum_{n=0}^{G-1} mnp(m,n) - \mu_x\mu_y}{\sigma_x\sigma_y} \quad \text{---------- (2.7)}$$

$$\mu_x = \sum_{m=0}^{G-1} m \sum_{n=0}^{G-1} p(m,n)$$

$$\mu_y = \sum_{m=0}^{G-1} n \sum_{n=0}^{G-1} p(m,n)$$

$$\sigma_x = \sum_{m=0}^{G-1}(m-\mu_x)^2 \sum_{n=0}^{G-1} p(m,n)$$

$$\sigma_y = \sum_{m=0}^{G-1}(m-\mu_y)^2 \sum_{n=0}^{G-1} p(m,n)$$

$$Homogenity = \sum_{m=0}^{G-1}\sum_{n=0}^{G-1}\frac{p(m,n)}{(1+|m-n|} \quad \text{---------- (2.8)}$$

**Gobor Features**

Gabor filters have frequency and orientation representations similar to human visual system. They are are most appropriate for texture representation and for object discrimination. Results of symmetric and

asymmetric Gobor filter are combined in a single quantity, called Gabor-energy. The Gabor-energy is closely related to the local power spectrum and it is associated with a pixel in an image. It is calculated as the squared modulus of the Fourier transform of the product between the image and a window function. Here, Gaussian window is used as a neighborhood function. Transformation is applied on the image with different orientations and scales. The resultant magnitude represents the energy $E(m,n)$ present in the image at different scales and is given: (2.9). From the magnitude, homogenous texture $\sigma_m$ is calculated using: (2.10) and (2.11).

$$E(m,n) = \sum_x \sum_y \left| G_m(x,y) \right| \qquad \text{-------- (2.9)}$$

$$m = 0,1,2,\ldots\ldots\ldots M-1$$
$$n = 0,1,2,\ldots\ldots\ldots N-1$$

$$\mu_m = \frac{E(m,n)}{P \times Q} \qquad \text{------ (2.10)}$$

$$\sigma_m = \frac{\sqrt{\sum_x \sum_y (|G_m(x,y)| - \mu_m)^2}}{P \times Q} \qquad \text{- ------ (2.11)}$$

$$P \times Q - imagesize$$

**2.3. Algorithm**

Read search image;
img=imread('humen.jpg');/ h
Divide the image into overlapping sub images
For each sub image,
function[X1,map1]=edg(R,R1);
Compute edge features;
function [X,map]=shpe(RGB,RGB1);
Compute Shape descriptors;
function[X1,map1]=txtre1(R1,R2);
Compute GLCM feature
edg(img,img1);
shpe(img,img1);
txtre1(img,img1);
function gb=gabor_fn(sigma,theta,lambda,psi,gamma)
Compute Gabor features;
Compute entropy features
Compute a combined feature vector;
Train for gun images;
Train for knife images;
Train for needle images;
Create a feature database;
Read test image(gun/ knife/needle)
R=imread('gun.jpg');
R2=imread('1knife.jpg');
R3=imread('needle.jpg');
Create a test feature vector;
Classify using Euclidean criteria

## III. ANALYSIS OF EXPERIMENTAL RESULTS

This paper has been investigated a feature based technique for detection of hidden objects from tera hertz images. It is implemented on Intel Dual Core Processor in Windows XP platform using MAT Lab 6.5. It is tested with images that consisted of gun, knife, needle etc. Here, edge features, shape descriptors, Gabor, GLCM texture features are computed to form a strong feature vector. Sample images used for search and test images in the experimentation are shown in figure 3.1 and 3.2. Since edges are less prone to noises, edge features alone are not sufficient to detect complex objects. Transform invariant shape descriptors and texture features are computed. GLCMs provide a quantitative picture of a spatial pattern of pixels. Pixels of same object have a higher correlation value compared to adjacent objects. Also, they are highly correlated with neighboring pixels than with distant pixels. Correlation co efficient is highly dependent on image window size. GLCM values are normalized to remove the image dependencies. Detection rate and accuracy are performance indicators for any

detection algorithm. Experiments are conducted with individual feature vector and with combined feature vector. The following observations are made: Gradient energy is found to be larger near the edges and smaller in smooth areas of the image. Shape descriptors provide valuable information about intricate objects. Gobor features together with GLCM features provide robustness and stability to the system. Contour-based methods are sensitive to noise; region-based methods able to detect shape defections also. Pixel processing and connected component analysis are adopted to extract the object.

Experimental results are better for the combined feature compared to the individual feature vector. Detection error is found to be very less 0% for knife image. Similarly, detection error is 1.1% using combined feature than shape features for gun images. Detection errors for various images using combined feature vector are tabulated in Table 3.1., Table 3.2 and Table 3.3.Yet, overall performance is good for combined feature vector. An extensive search is carried out at various resolutions for multiple occurrences of same or different object. Due to limited availability of tera hertz images, we are able to conduct experiments only with a limited set of images that were taken from various scenes at different angles. For blurred and occluded images, detection error is poor. It is also found that detection rate is 94.4 % for knife images and 88.8% for needle images. Computation of detection rate for various images is shown in Table 3.4. Experimental outputs for gun image are shown in figure 3.3.Experimental outputs for knife image are shown in figure 3.4. Experimental outputs for needle image are shown in figure 3.5.

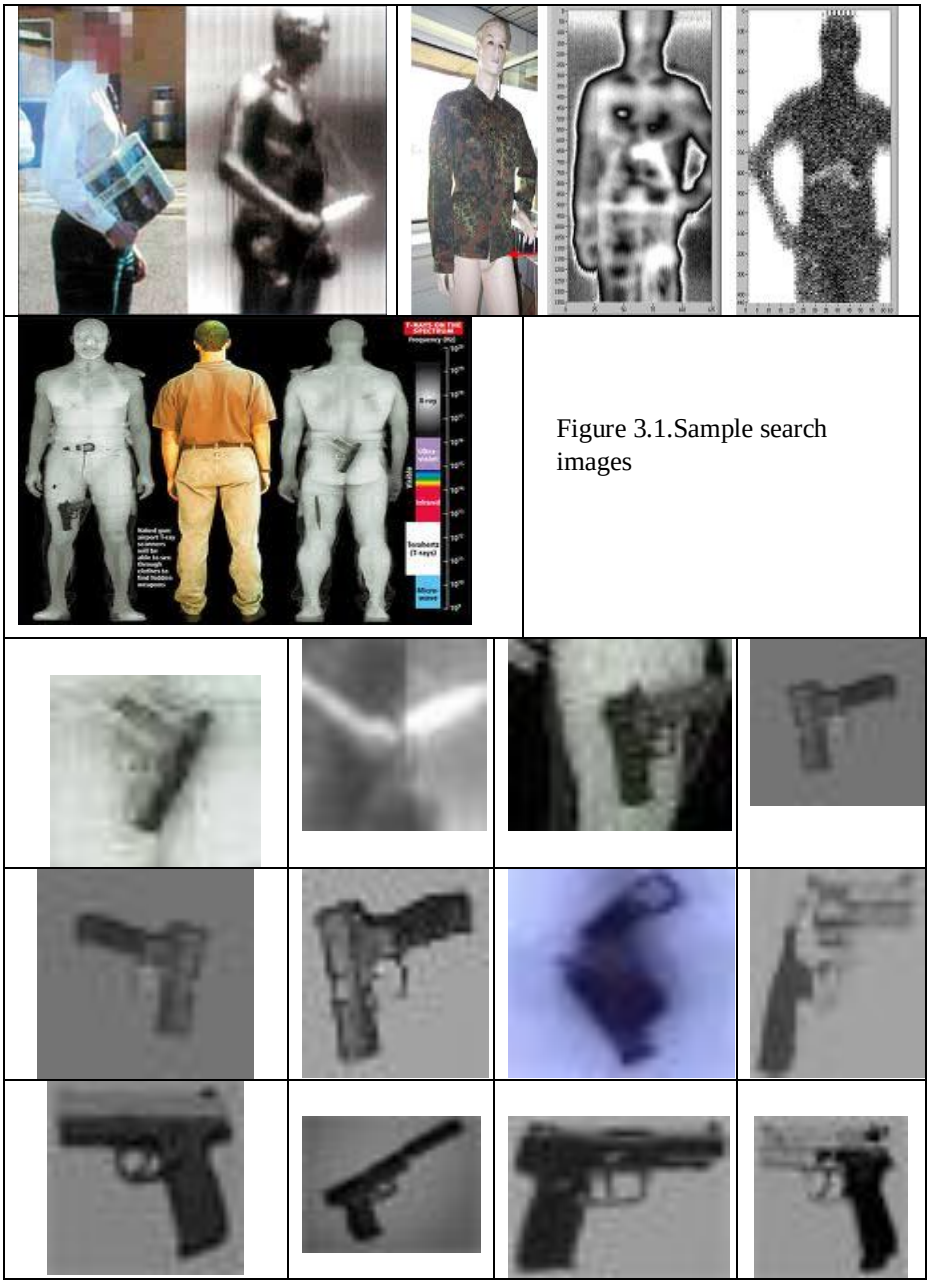


Figure 3.1.Sample search images

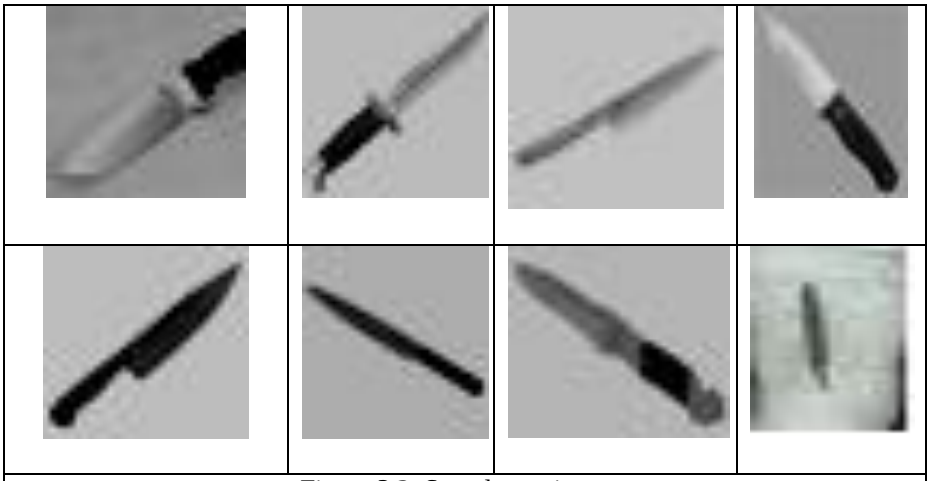
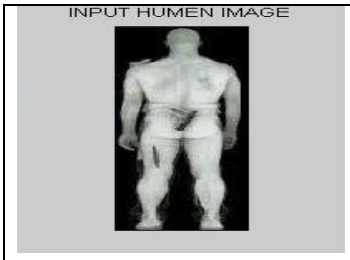


Figure 3.2 Sample test images




(a) (b) ©

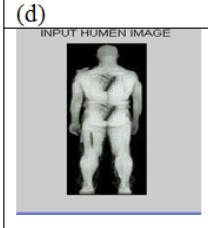
(d)

Figure 3.3 Detection outputs for gun
a) Search image (190x275)
b) Combined feature
c) Shape feature
d) Texture feature


(a) (b) ©

(d)

Figure .3.4 a) Detection outputs for knife
a) Search image (190x275)
b) Combined feature
c) Shape feature
d) Texture feature




a) b) c)

Figure.3.4 Detection outputs for needle a) Search image (190x275)
b)Texture feature c) Shape feature

TABLE 3.1
Computation of Detection Error (Combined Feature)

| No | Test Images | Test Images | Actual pixels | Estimated pixels (detected) | %Detection error E-A/Ax100 |
|---|---|---|---|---|---|
| 1 | Gun | 32x32 | 91 | 90 | 1.1 |
| 2 | Knife | 32x32 | 32 | 32 | 0 |
| 4 | Needle | (24x55) | 99 | 97 | 2.02 |
| | | | | Average error | 1.04 |

TABLE 3.2
Computation of Detection Error
For shape feature

| S.No | Test Images | Image Size | Shape features | | %Detection error |
|---|---|---|---|---|---|
| | | | Actual pixels | Estimated pixels (detected) | E-A/Ax100 |
| 1 | Gun | 32x32 | 91 | 89 | 2.1 |
| 2 | Knife | 32x32 | 32 | 32 | 0 |
| 4 | Needle | (24x55) | 99 | 94 | 5.05 |
| Average | | | | | 2.38 |

TABLE 3.3
Computation of Detection error for Texture Features

| Sno | Test Images | Test Images | Texture features | | %Detection error |
|---|---|---|---|---|---|
| | | | Actual pixels | Estimated pixels (detected) | E-A/Tx100 |
| 1 | Gun | 32x32 | 315 | 310 | 1.58 |
| 2 | Knife | 32x32 | 91 | 90 | 1.09 |
| 4 | Needle | (24x55) | 220 | 211 | 4.09 |
| | | | | Average | 2.25 |

TABLE 3.4
Computation of Detection rate

| S,No | Test images | Number of instances tested | Number of Successful detection | Number of Failures | Detection rate in % |
|---|---|---|---|---|---|
| **1** | Gun | 27 | 25 | 2 | 92.5 |
| **2** | Knife | 18 | 17 | 1 | 94.4 |
| **3** | Needle | 18 | 16 | 2 | 88.8 |
| | | | | Average | 91.9 |

## IV. CONCLUSION

This paper has explored the use of feature based technique for distinguishing interested regions in the sample tera hertz images. From the detection error, we have observed that combined feature vector makes the classification more stable and robust. Also, we have achieved 1.04% detection error for combined feature and 91.9% detection rate from the experiments. Future work may consist of using additional metrics to compare the experimental results and testing the algorithm with different image sets. This new brand of detection system may enable more applications in security and safety.

## REFERENCE


### Journal papers:

[1]. Jan Flusser , Moment Invariants in Image Analysis, *World Academy of Science, Engineering and Technology 11 , 2005,pp:376-381.*

[2] KapilevichB.,Detecting Hidden Objects on Human Body Using Active Millimeter Wave Sensor, *Sensors Journal, IEEE, Volume: 10, Issue: 11, On Page(s): 1746 – 1752.*

[3]. Xilin,Shencharles R Dietlein, Erich Grossman, Zoya Popvic and Francois G Meyer, Detection & Segmentation of concealed objects in Terahertz Images, *IEEE Transactions on Image Processing,* 17 (12), pp 2465-2475, 2008.

[4]. Federici, John F.; Schulkin, Brian; Huang, Feng; Gary, Dale; Barat, Robert; Oliveira, Filipe; Zimdars, David , THz imaging and sensing for security applications—explosives, weapons and drugs, *,Semiconductor Science and Technology, Volume 20, Issue 7, pp. S266-S280 (2005).*

[5]. Hai-Bo Liu, Yunqing Chen, Glenn J. Bastiaans, and X.-C. Zhang, Detection and identification of explosive RDX by THz diffuse reflection spectroscopy, *Optics Express, Vol. 14, Issue 1, pp. 415-423 (2006)*

### Magazine:

[6]. H. Chen, S. Lee, R. Rao, M.-A. Slamani, and P. Varshney, Imaging for concealed weapon detection*," IEEE Signal Processing Magazine, pp. 52–61, March 2005.*

[7]. Dr Alan J. Lipton, Craig H. Heartwell, Dr Niels Haering, and Donald Madden, Critical Asset Protection, Perimeter Monitoring, and Threat Detection Using automated Video Surveillance , *IEEE AESS systems magazine , 2003*

**Books:**

[8]. Milan Sonaka, Vaclav Hlavac, Roger Boyle, *Image processing, Analysis, and Machine Vision* (Thomson learning Inc. ISBN : 981-240-061-3, 1998).

[9]. Rafael C.Gonzalez, Richard E.Woods, *Digital Image processing* ( PHI Learning Private Ltd. New Delhi, ISBN: 978-81-203-3640-7. 2008).

**Proceedings:**

[10]. A. Agurto, Y. Li, G. Tian, N. Bowring, and S. Lockwood, "A review of Concealed weapon detection and research in perspective," in Proceedings of the 2007 IEEE International Conference on MonE02 Networking, Sensing and Control, 2007, pp. 443– 448 .

[11]. Appleby, Roger; Anderton, Rupert N.; Price, Sean; Sinclair, Gordon N.; Coward, Peter R. , Whole-body 35-GHz security scanner, Radar Sensor Technology VIII and Passive Millimeter-Wave Imaging Technology VII.Proceedings of the SPIE, Volume 5410, pp. 244-251 (2004).

[12]. Yunqing Chen, Haibo Liu, X.-C. Zhang, etc, "Spectroscopic characterization of explosives in the far infrared region", Proceedings of SPIE, 5411, pp1-7, 2004

[13]. Michael C Kemp , Millimetre Wave and Terahertz Technology for the Detection of Concealed Threats – A Review, , Iconal Technology Ltd, St John's Innovation Centre, Cambridge, United Kingdom, Optics and Photonics for Counter-Terrorism and Crime Fighting II Proc. of SPIE Vol. 6402, 64020D, · 0277-786X/06/$15 · doi: 10.1117/12.692612, 2006

**Reports:**

[14]. Peter Baines, Hazardous Materials: Chemical Biological Radiological and Nuclear – A Review: 2004 to 2007, Detective Report –, National Institute of Forensic Science, Australia.

[15]. Wai Lam Chan, Jason Deibel and Daniel M Mittleman, Imaging with terahertz radiation, IOP Publishing, Reports On Progress In Physics, Rep. Prog. Phys. 70 (2007) 1325–1379, doi:10.1088/0034-4885/70/8/R02.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 33-37*
*www.iosrjournals.org*

# Discovering Thematic Object in a Video

[1]Shalini N, [2]Sharada K A

***Abstract:*** *In this paper we are identifying the frequently appearing object in a video. This frequently appearing object is called thematic object. If an object appears many times in a video such objects are called thematic object. Identifying this frequently appearing object in a video is helpful for object search and tagging of that object. To identify thematic pattern in the video we must give an object as an input and then we try to find corner points of that object by Harris corner detection algorithm. Later we can find the similarity between the reference image and test frame by extracting the descriptors around the corner point. We are mining the video to identify the common patterns that appears in that video. The proposed approach will help to identify the object even when there is partial occlusion and variation in the viewpoint.*

***Index Terms:*** *Common Patterns, Corner points, Key object, Video data mining.*

## I. Introduction

We are given with a collection of videos. In that we need to identify the object which appears frequently. This frequently appearing object is called thematic object [1] or key object. Identifying this key object is helpful for visual object search and detection [2]. By keeping this key object we can be able to retrieve all the video frames which contain this key object.

Discovering objects in video is a challenging task [3]. By discovering, we mean that the object can be of any kind. Without having any prior knowledge about the object type or its position, we would like to identify an object from a video that occurs over a period of time. This is particularly challenging when the image sequence has low resolution and consists of highly cluttered background.

Object identification can be defined as [4] the process of segmenting an object of interest from a video scene and keeping track of its motion, orientation, occlusion etc. in order to extract useful information.

For identifying this key object, we characterize each video as a sequence of frames. We need to check whether this key object appears in the rest of the frames or not. It is not easy to

Identify whether that object is present in each frame or not. Because the shape of the object may vary from one frame to another frame. The view point of the object may vary i.e. an object can be viewed from front view or from side view. Sometimes there can be partial occlusion of the object. In all such cases we need to identify the key object in each and every frame of the video.

The video is nothing but the collection of frames. Each frames of the video contain some images. All the frames of the video may or may not contain the thematic object. So we need to

Identify which frames of the video contain this thematic object. To identify the thematic object, local feature points are calculated for every image. Local feature point is nothing but the corner points. The corner points can be identified by Harris corner detection method. After identifying the corner point, a window is generated around each corner point. A window is nothing but a descriptor which is collection of pixels around the corner point. Window is also called a patch around the corner point. Like this descriptors are extracted for the given image and the frames of the video. These patch descriptors of the given image and frame of video are compared each other to find similarity. If any similar object is present in the frame of the video, such frames are identified and saved so that it can be review in the later stage. The similarity between the given image and the frame of the video is done to identify the common object. Like this thematic object is identified in the given video. Identifying thematic object in a video is helpful for object search, tagging and video summarization

To identify this frequently appearing object we first convert the image into set of features. This feature vector is used for matching to find whether the object of interest [7] is present in the rest of the video.

## II. Literature Survey

To identify common visual patterns in the image, some previous work identify an image as a graph consisting of visual primitives such as corners, interest points.

Shalini N, Computer Science and Engineering, East West Institute of Technology, Bangalore, India, +91-8971810862., (e-mail: shalinitsn@gmail.com).
Sharada K.A, Assistant Professor, Computer Science and Engineering, East West Institute of Technology, Bangalore, India, +91-9448855473., (e-mail: sharadaa1234@gmail.com)

Yang and Cohen [5] use affine transformation for object recognition. An object is recognized by affine invariants to establish the correspondence between the vertices of a test image with a reference image. The algorithm used in this will recognize an object that must be represented as polygonal outliners and also as a set of scattered feature points. It uses point matching approach for recognition. However, if the objects have different shape then they have identical convex hulls.

Tan and Ngo [10] propose an approach to discover common patterns in a set of images by region matching. Here image is segmented into regions. Histogram is drawn for each segmented region. Histogram is the graphical representation of the given image. By the histograms of different segments, the presence of the common object is identified. If histograms are similar then there is common object is present in it. This is called region matching

Ying Shan and Harpreet [6] use a histogram to solve sequence to sequence matching problem. It will identify an object in the presence of large appearance and pose variations and also background clutter. However this approach uses intensity profile feature. This method is not invariant to illumination changes

Liu and Chen [2] propose an approach for video object discovery, which extracts unknown object form video. This approach uses video data mining and object oriented nature for video content analysis. But this provides a rough position for the object of interest.

## III. SYSTEM MODEL

To find the object of interest in a given video and to find its reoccurrences in the video we use different modules. First is the Feature extractionwhich converts the image into a set of features or feature vectors which is used for further analysis. Next is the Feature matching, where the features extracted from the input image are matched with the stored template or reference model and a recognition decision is made.

All thematic recognition systems have to serve two distinguished phases. The first one is referred to as enrolment or training phase and the second one is referred to as the operational or testing phase.

To identify thematic object in a video, we characterize video as a collection of video frames, each frame as an image sequence. Each image is characterized by collection of local features. A local feature is a property of an image located on a single point or small region. It is a single piece of information describing the distinctive property of the object. Examples for the local features of an object are color or gray value of a pixel. For object recognition task, the local feature must be invariant to illumination changes, noise, scale changes and changes in view direction. These local features are also called visual primitives. We match the visual primitives to identify the presence thematic object in a video. In the initialization step, we discard uncommon visual primitives that find few matches among the rest of the image because they will not belong to thematic object. For the remained visual primitive, we make a larger visual group and check the commonness for the extent of thematic object in a video.

Various videos and images are collected and stored as a dataset. First we need to select an image. We need to select a video to identify whether the same query image is present in it or not. The feature points are identified by Harris corner point detection. A patch window of 21*21 is identified for every feature point by extract descriptor module. The match function is applied to the patch descriptor to find there is any match between the given image and the frames of the video. Like this every frame of the video is compared to find whether the thematic object is present in a video or not.

## IV. BLOCK DIAGRAM

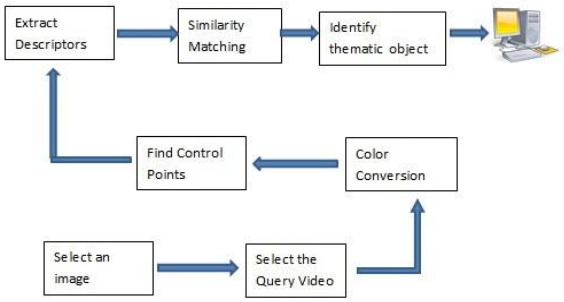



Figure 1

Above shows the block diagram for identifying thematic object in a video. First we are selecting an image and we are selecting a video, we are doing color conversion for both the reference image and the test video. For this we are finding the control points by using harris corner detection method. We are extracting the descriptors

around the corner point to do similarity matching. If similar object is found then it is displayed stating it as a thematic object.

## V. SYSTEM MODULES

Here we are showing the different modules with their description.

### A. Video acquisition module:

We can read the input video by uigetfile command which allow us to browse the video in which we are going to given as the input to the system. So we can get the video by selecting it from database.

### B. Video preprocessing module:

The size of the video is verified and it is resized to 320x240 of frame size. These videos are sampled at 2 frames per second.

So, in video preprocessing step frames are extracted from the video and they are resized if they are not of proper size.

### C. ROI Selection:

ROI is nothing but region of interest. Careful selection of video is essential for ROI analysis. For this we select an image and it is kept as a reference frame. This is used to compare the frames of the video to identify the thematic object is present in the video.

### D. Color conversion:

Color is the brains reaction to a specific visual stimulus. We can precisely describe color by measuring its spectral power distribution which leads to a large degree of redundancy.

There are three basic quantities they are

Radiance is the energy that flows from the light source and it is measured in watts.

Luminance is a measure of energy i.e. what an observer perceives from a light source and it is measured in lumens. Brightness is a subjective descriptor which difficult to measure.

### E. Edge Detection:

The purpose of edge detection [8] in general is to significantly reduce the amount of data in an image while preserving the structural properties to be used for further image processing.

The Canny Edge Detection algorithm has five steps:

1. Smoothing:

Blurring of image to remove noise.

2. Finding Gradients:

The edges should be marked where the gradients of the image has large magnitudes.

3. Non-maximum suppression:

Only local maxima should be marked as edges.

4. Double Thresholding:

Potential edges are determined by Thresholding.

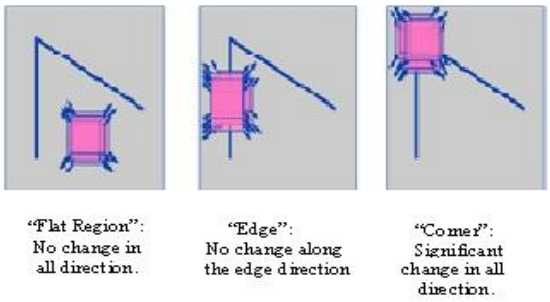
5. Edge tracking by hysteresis:

Final edges are determined by suppressing all the edges that are not connected to a strong edge.

### F. Harris Corner Point Identification:

Harris Corner point detector [9] is a popular interest point detector. It has strong invariance to rotation, scale, illumination variation and noise. The Harris Corner detector is based on the local auto-correlation function. The local auto-correlation function measures the local changes with the patches shifted by a small amount in different directions.

Harris Corner Detector Basic Idea:




**G. Similarity matching:**

The match function will read two images. The local feature points and descriptors are found out for the two images. Distance ratio value is initialized. It is the value where the similarity between the two images has to be computed. A match is found out if the ratio of vector angles from the nearest to second nearest neighbor is less than distance ratio. For each descriptor in the first image, its match to second image is selected. For this, first it is required to transpose the descriptor matrix. Dot product of the matrix and its transpose is done. Inverse cosine is applied for the dot product matrix and it is sorted. When sorted, if any of the value is less than the distance ratio value, there is a match between the two images. If value is more than the distance ratio, there will be no match between the two images. The similarity between the two images is identified.

The pseudo code for finding the corner points is shown below. The different steps such as reading an image as an input, the preprocessing steps are shown in below. The detection of edge is done by using built in cranny's edge detection algorithm. The number of corner points required can be specified by user. Finally the corner points can be detected by Harris corner detection algorithm.

```
%im = im2double(rgb2gray(imrgb));
g1 = fspecial('gaussian', 9,1);
g2 = fspecial('gaussian', 11,1.5);
G1 = g1;
G2 = g2;
img1 = conv2(im, g1, 'same');
Ix = conv2(img1, [-1 0 1], 'same');
Iy = conv2(img1, [-1;0;1], 'same');
Ix2 = conv2(Ix.*Ix, g2, 'same');
Iy2 = conv2(Iy.*Iy, g2, 'same');
IxIy = conv2(Ix.*Iy, g2, 'same');
R = (Ix2.*Iy2 - IxIy.*IxIy)
./ (Ix2 + Iy2 + eps);
R([1:20, end-21:end], :) = 0;
R(:,[1:20,end-21:end]) = 0;s
nonmax = inline('max(x)');
Rmax = colfilt(R, [3 3], 'sliding', nonmax);
Rnm = R .*(R == Rmax);
[y, x, strength] = find(Rnm);
[yy ii]  = sort(-strength);
nn       = min(topn, length(ii));
y        = y(ii(1:nn));
x        = x(ii(1:nn));
```

The below figures shows the sample results to identify corner points. Fig 3.1 is the selection of image for identification of corner points. Fig 3.2 shows the identification of corner points by applying Harris method.

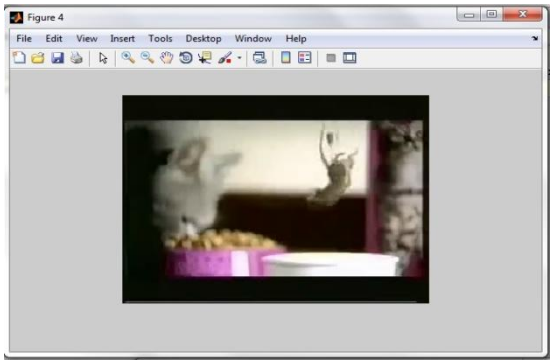



Fig 3.1

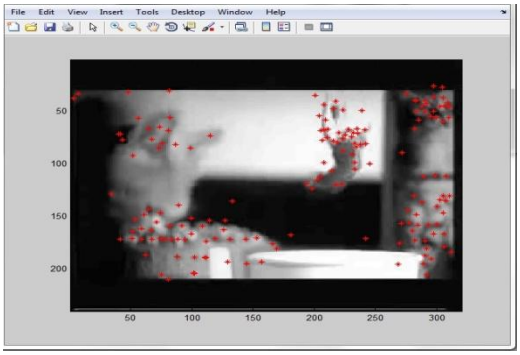



Fig 3.2

## VI. CONCLUSION AND FUTURE WORK

In this paper, we are identifying the frequently appearing object in a video. This frequently appearing object is also called thematic object. Identifying this thematic object is helpful for object search and their tagging. It is also helpful for video summarization. Our approach will identify this frequently appearing object even in the presence of background clutters, partial occlusion. We are identifying the thematic object by considering the region of interest in the selected object. Then we are applying the Harris algorithm to select the corner points in the image. Here we are identifying only one thematic object at a time. Our future work is to identify multiple thematic object in a single frame.

## ACKNOWLEDGMENT

Special thanks to East West Institute of Technology for helping us in our work and supporting the research. Visveshvaraya Technological research centre are also helping us in this work.

## REFERENCES

[1] Junsong Yuan, Gangqiang Zhao, Yun Fu, Zhu Li, Aggelos K. Katsaggelos and Ying Wu, "Discovering Thematic Objects in Image Collections and Videos" IEEE Transactions on Image Processing vol.21, No. 4, April 2012.
[2] Davide Liu, Tshun Chen: "DISCOV :A Framework for Discovering Objects in Video", IEEE Transactions on Multimedia, Vol. 10,No.2, February 2008.
[3] David Liu and Tshun Chen: "A Topic-Motion Model for Unsupervised Video Object Discovery".
[4] Alok K. Watve: "Object tracking in video scenes".
[5] Zhengwei Yang and Fernand S. Cohen, "Image Registration and Object Recognition Using Affine Invariants and Convex Hulls", IEEE Transactions on Image Processing, Vol 8, No. 7, July 1999.
[6] Ying Shan, Harpreet S. Sawhney, Art Pope: "Measuring the Similarity Of Two Image Sequences".
[7] David Liu, Gang Hua, Tsuhan Chen: " A Hierarchical Visual Model for Video Object Summerization".IEEE Transactions on Pattern analysis and Machine Intelligence vol 32 December 2010.
[8] John Canny: "A Computational approach to Edge Detection". Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI Nov 1986.
[9] Harris, C., Stephens : "A Combined Corner and Edge Detector", Proceedings of Alvey Vision Conference, 1988.
[10] Hung Khoon Tan and chongwah Ngo " Localized matching using earth mover's distance towards discovery of common patterns from small image samples:

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 38-43*
*www.iosrjournals.org*

# Exponential software reliability using SPRT: MLE

S. Murali Mohan[1], Dr. R. Satya Prasad [2], G. Krishna Mohan[3]
[2](Controller of Examinations, Vikrama Simhapuri University, Nellore)
[1](Associate Professor Dept. of CS&Engg. Acharya Nagrjuna University, Nagarjuna Nagar)
[3](Reader, Dept. of Computer Science, P.B.Siddhartha College, Vijayawada)

***Abstract:*** *In Classical Hypothesis testing volumes of data is to be collected and then the conclusions are drawn, which may need more time. But, Sequential Analysis of Statistical science could be adopted in order to decide upon the reliability or unreliability of the developed software very quickly. The procedure adopted for this is, Sequential Probability Ratio Test (SPRT). It is designed for continuous monitoring. The likelihood based SPRT proposed by Wald is very general and it can be used for many different probability distributions. In the present paper we propose the performance of SPRT on 4 data sets of Time domain data using exponential model and analyzed the results. The parameters are estimated using Maximum Likelihood Estimation method.*
***Keywords:*** *SPRT, SRGM,HPP,NHPP, MLE, Goel-Okumoto model.*

## I. INTRODUCTION

Sequential analysis is a method of statistical inference whose main feature is that the number of observations required by the procedure is not determined in advance. The decision to end the observations depends, at each stage, on the results of the samples already taken. (SPRT), which is usually applied in situations, requires a decision between two simple hypothesis or a single decision point. Wald's (1947) SPRT procedure has been used to classify the software under test into one of two categories (e.g., reliable/unreliable, pass/fail, certified/noncertified) (Reckase, 1983). Wald's procedure is particularly relevant if the data is collected sequentially. Classical Hypothesis Testing is different from Sequential Analysis. In Classical Hypothesis testing, the number of cases tested or collected is fixed at the beginning of the experiment. In this method, the analysis is made and conclusions are drawn after collecting the complete data.

In the analysis of software failure data, either TBFs or failure count in a given time interval is dealt with. If it is further assumed that the average number of recorded failures in a given time interval is directly proportional to the length of the interval and the random number of failure occurrences in the interval is explained by a Poisson process. Then it is known that the probability equation of the stochastic process representing the failure occurrences is given by a Homogeneous Poisson Process with the expression

$$P\left[N(t)=n\right]=\frac{e^{-\lambda t}(\lambda t)^{n}}{n!} \qquad (1.1)$$

Stieber (1997) observes that, the application of SRGMs may be difficult and reliability predictions can be misleading, if classical testing strategies are used. However, he observes that statistical methods can be successfully applied to the failure data. He demonstrated his observation by applying the well-known sequential probability ratio test of Wald for a software failure data to detect unreliable software components and compare the reliability of different software versions. In this chapter the popular SRGM – Exponential model is considered and the principle of Stieber is adopted in detecting unreliable software in order to accept or reject the developed software. The theory proposed by Stieber is presented in Section 2 for a ready reference. Extension of this theory to the considered SRGM is presented in Section 3. Maximum Likelihood parameter estimation method is presented in Section 4. Application of the decision rule to detect unreliable software with reference to the SRGM is given in Section 5.

## II. SEQUENTIAL TEST FOR A POISSON PROCESS

A.Wald, developed the SPRT at Columbia University in 1943. A big advantage of sequential tests is that they require fewer observations (time) on the average than fixed sample size tests. SPRTs are widely used for statistical quality control in manufacturing processes. The SPRT for Homogeneous Poisson Processes is described below.

Let $\{N(t), t \geq 0\}$ be a homogeneous Poisson process with rate ' $\lambda$ '. In this case, $N(t)$ = number of failures up to time ' t' and ' $\lambda$ ' is the failure rate (failures per unit time). If the system is put on test (for example a software system, where testing is done according to a usage profile and no faults are corrected) and that if we want to estimate its failure rate ' $\lambda$ '. We can not expect to estimate ' $\lambda$ ' precisely. But we want to

reject the system with a high probability if the data suggest that the failure rate is larger than $\lambda_1$ and accept it with a high probability, if it is smaller than $\lambda_0$. As always with statistical tests, there is some risk to get the wrong answers. So we have to specify two (small) numbers '$\alpha$' and '$\beta$', where '$\alpha$' is the probability of falsely rejecting the system. That is rejecting the system even if $\lambda \le \lambda_0$. This is the "producer's" risk. '$\beta$' is the probability of falsely accepting the system .That is accepting the system even if $\lambda \le \lambda_1$ . This is the "consumer's" risk. Wald's classical SPRT is very sensitive to the choice of relative risk required in the specification of the alternative hypothesis. With the classical SPRT, tests are performed continuously at every time point $t > 0$ as additional data are collected. With specified choices of $\lambda_0$ and $\lambda_1$ such that $0 < \lambda_0 < \lambda_1$ , the probability of finding $N(t)$ failures in the time span $(0,t)$ with $\lambda_1$, $\lambda_0$ as the failure rates are respectively given by

$$P_1 = \frac{e^{-\lambda_1 t}\left[\lambda_1 t\right]^{N(t)}}{N(t)!} \tag{2.1}$$

$$P_0 = \frac{e^{-\lambda_0 t}\left[\lambda_0 t\right]^{N(t)}}{N(t)!} \tag{2.2}$$

The ratio $\frac{P_1}{P_0}$ at any time '$t$' is considered as a measure of deciding the truth towards $\lambda_0$ or $\lambda_1$ , given a sequence of time instants say $t_1 < t_2 < t_3 < \ldots\ldots < t_K$ and the corresponding realizations $N(t_1), N(t_2), \ldots\ldots N(t_K)$ of $N(t)$. Simplification of $\frac{P_1}{P_0}$ gives

$$\frac{P_1}{P_0} = \exp(\lambda_0 - \lambda_1)t + \left(\frac{\lambda_1}{\lambda_0}\right)^{N(t)}$$

The decision rule of SPRT is to decide in favour of $\lambda_1$, in favour of $\lambda_0$ or to continue by observing the number of failures at a later time than 't' according as $\frac{P_1}{P_0}$ is greater than or equal to a constant say A, less than or equal to a constant say B or in between the constants A and B. That is, we decide the given software product as unreliable, reliable or continue (Satyaprasad, 2007) the test process with one more observation in failure data, according to

$$\frac{P_1}{P_0} \ge A \tag{2.3}$$

$$\frac{P_1}{P_0} \le B \tag{2.4}$$

$$B < \frac{P_1}{P_0} < A \tag{2.5}$$

The approximate values of the constants A and B are taken as $A \cong \frac{1-\beta}{\alpha}$, $B \cong \frac{\beta}{1-\alpha}$

Where '$\alpha$' and '$\beta$' are the risk probabilities as defined earlier. A good test is one that makes the $\alpha$ and $\beta$ errors as small as possible. The common procedure is to fix the $\beta$ error and then choose a critical region to minimize the error or maximize the power i.e $1-\beta$ of the test. A simplified version of the above decision processes is to reject the system as unreliable if $N(t)$ falls for the first time above the line

$$N_U(t) = a.t + b_2 \tag{2.6}$$

To accept the system to be reliable if $N(t)$ falls for the first time below the line

$$N_L(t) = a.t - b_1 \tag{2.7}$$

To continue the test with one more observation on $\left(t, N\left(t\right)\right)$ as the random graph of $\left[t, N\left(t\right)\right]$ is between the two linear boundaries given by equations (2.6) and (2.7) where

$$a = \frac{\lambda_1 - \lambda_0}{\log\left(\frac{\lambda_1}{\lambda_0}\right)} \tag{2.8}$$

$$b_1 = \frac{\log\left[\frac{1-\alpha}{\beta}\right]}{\log\left(\frac{\lambda_1}{\lambda_0}\right)} \tag{2.9}$$

$$b_2 = \frac{\log\left[\frac{1-\beta}{\alpha}\right]}{\log\left(\frac{\lambda_1}{\lambda_0}\right)} \tag{2.10}$$

The parameters $\alpha, \beta, \lambda_0$ and $\lambda_1$ can be chosen in several ways. One way suggested by Stieber is

$$\lambda_0 = \frac{\lambda.\log(q)}{q-1}, \quad \lambda_1 = q\frac{\lambda.\log(q)}{q-1} \qquad where\ q = \frac{\lambda_1}{\lambda_0}$$

If $\lambda_0$ and $\lambda_1$ are chosen in this way, the slope of $N_U\left(t\right)$ and $N_L\left(t\right)$ equals $\lambda$. The other two ways of choosing $\lambda_0$ and $\lambda_1$ are from past projects (for a comparison of the projects) and from part of the data to compare the reliability of different functional areas.

## III. SEQUENTIAL TEST FOR SOFTWARE RELIABILITY GROWTH MODELS

In Section 2, for the Poisson process it is known that the expected value of $\mathrm{N}\left(\mathrm{t}\right) = \lambda \mathrm{t}$ called the average number of failures experienced in time '$t$' .This is also called the mean value function of the Poisson process. On the other hand if we consider a Poisson process with a general function (not necessarily linear) $m\left(t\right)$ as its mean value function the probability equation of a such a process is

$$P\left[N(t) = Y\right] = \frac{\left[m(t)\right]^y}{y!}.e^{-m(t)}, y = 0,1,2,----$$

Depending on the forms of $m\left(t\right)$ various Poisson processes called NHPP are obtained. For our two parameter Exponential model, the mean value function is given as $m\left(t\right) = a\left(1-e^{-(bt)}\right)$ where $a > 0, b > 0$

It may be written as

$$P_1 = \frac{e^{-m_1(t)}.\left[m_1(t)\right]^{N(t)}}{N(t)!}$$

$$P_0 = \frac{e^{-m_0(t)}.\left[m_0(t)\right]^{N(t)}}{N(t)!}$$

Where, $m_1(t)$, $m_0(t)$ are values of the mean value function at specified sets of its parameters indicating reliable software and unreliable software respectively. Let $P_0, P_1$ be values of the NHPP at two specifications of b say $b_0, b_1$, where $\left(b_0 < b_1\right)$. It can be shown that for our model $m\left(t\right)$ at $b_1$ is greater than that at $b_0$. Symbolically $m_0\left(t\right) < m_1\left(t\right)$. Then the SPRT procedure is as follows:

Accept the system to be reliable if, $\frac{P_1}{P_0} \le B$

i.e., $\frac{e^{-m_1(t)}.\left[m_1(t)\right]^{N(t)}}{e^{-m_0(t)}.\left[m_0(t)\right]^{N(t)}} \le B$

i.e., $N(t) \le \dfrac{\log\left(\dfrac{\beta}{1-\alpha}\right) + m_1(t) - m_0(t)}{\log m_1(t) - \log m_0(t)}$ (3.1)

Decide the system to be unreliable and reject if, $\dfrac{P_1}{P_0} \ge A$

i.e., $N(t) \ge \dfrac{\log\left(\dfrac{1-\beta}{\alpha}\right) + m_1(t) - m_0(t)}{\log m_1(t) - \log m_0(t)}$ (3.2)

Continue the test procedure as long as

$$\frac{\log\left(\dfrac{\beta}{1-\alpha}\right) + m_1(t) - m_0(t)}{\log m_1(t) - \log m_0(t)} < N(t) < \frac{\log\left(\dfrac{1-\beta}{\alpha}\right) + m_1(t) - m_0(t)}{\log m_1(t) - \log m_0(t)} \quad (3.3)$$

Substituting the appropriate expressions of the respective mean value function $-m(t)$ of Exponential we get the respective decision rules and are given in following lines

Acceptance region:

$$N(t) \le \frac{\log\left(\dfrac{\beta}{1-\alpha}\right) + a\left(e^{-(b_0 t)} - e^{-(b_1 t)}\right)}{\log\left(\dfrac{1-e^{-(b_1 t)}}{1-e^{-(b_0 t)}}\right)} \quad (3.4)$$

Rejection region:

$$N(t) \ge \frac{\log\left(\dfrac{1-\beta}{\alpha}\right) + a\left(e^{-(b_0 t)} - e^{-(b_1 t)}\right)}{\log\left(\dfrac{1-e^{-(b_1 t)}}{1-e^{-(b_0 t)}}\right)} \quad (3.5)$$

Continuation region:

$$\frac{\log\left(\dfrac{\beta}{1-\alpha}\right) + a\left(e^{-(b_0 t)} - e^{-(b_1 t)}\right)}{\log\left(\dfrac{1-e^{-(b_1 t)}}{1-e^{-(b_0 t)}}\right)} < N(t) < \frac{\log\left(\dfrac{1-\beta}{\alpha}\right) + a\left(e^{-(b_0 t)} - e^{-(b_1 t)}\right)}{\log\left(\dfrac{1-e^{-(b_1 t)}}{1-e^{-(b_0 t)}}\right)} \quad (3.6)$$

It may be noted that in the above mentioned model the decision rules are exclusively based on the strength of the sequential procedure $(\alpha, \beta)$ and the values of the respective mean value functions namely, $m_0(t)$, $m_1(t)$. If the mean value function is linear in '$t$' passing through origin, that is, $m(t) = \lambda t$ the decision rules become decision lines as described by Stieber. In that sense equations (3.1), (3.2), (3.3) can be regarded as generalizations to the decision procedure of Stieber. The applications of these results for live software failure data are presented with analysis in Section 5.

## IV. MAXIMUM LIKELIIHOOD PARAMETER ESTIMATION: G-O MODEL

The likelihood function of G-O model is given as, $L = \prod_{i=1}^{N} abe^{-(bt)}$

Taking the natural logarithm on both sides, The Log Likelihood function is given as:

$$\log L = \sum_{i=1}^{n} \log(abe^{-(bt_i)}) - a[1 - e^{-(bt_n)}] \quad (4.1)$$

Taking the Partial derivative of log L with respect to 'a' and equating to '0'.

$$a = \frac{n}{\left[1-e^{-(bt_n)}\right]} \tag{4.2}$$

Taking the Partial derivative of log L with respect to 'b' and equating to'0'.

$$g(b) = \sum_{i=1}^{n} t_i - \frac{n}{b} + nt_n \frac{e^{-(bt_n)}}{\left(1-e^{-(bt_n)}\right)} = 0 \tag{4.3}$$

Taking the partial derivative of log L again with respect to 'b' and equating to '0'.

$$g'(b) = \frac{n}{b^2} - nt_n^2 \left\{ \frac{1}{\left(1-e^{-(bt_n)}\right)} + \frac{e^{-(bt_n)}}{\left(1-e^{-(bt_n)}\right)^2} \right\} e^{-bt_n} \tag{4.4}$$

The parameters 'a' and 'b' are estimated as follows. The parameter 'b' is estimated by iterative Newton Raphson Method using $b_{n+1} = b_n - \frac{g(b_n)}{g'(b_n)}$, which is substituted in finding 'a'.

## V. SPRT ANALYSIS OF DATA SETS : TIME DOMAIN

In this section, the developed SPRT methodology is shown for a software failure data which is of time domain. The decision rules based on the considered mean value function for Four different data sets, borrowed from Pham (2006), Xie *et al.*, (2002) are evaluated. Based on the estimates of the parameter 'b' in each mean value function, we have chosen the specifications of $b_0 = b - \delta$, $b_1 = b + \delta$ equidistant on either side of estimate of b obtained through a data set to apply SPRT such that $b_0 < b < b_1$. Assuming the value of $\delta = 0.0025$, the choices are given in the following table.

**Table 5.1: Estimates of a, b & Specifications of $b_0$, $b_1$ for Time domain**

| Data Set | Estimate of 'a' | Estimate of 'b' | $b_0$ | $b_1$ |
|---|---|---|---|---|
| 1 (Xie) | 31.899246 | 0.003819 | 0.001319 | 0.006319 |
| 2 (AT&T) | 23.582254 | 0.003973 | 0.001473 | 0.006473 |
| 3 (NTDS) | 30.168926 | 0.007917 | 0.005417 | 0.010417 |
| 4 (IBM) | 17.608791 | 0.006451 | 0.003951 | 0.008951 |

Using the selected $b_0$, $b_1$ and subsequently the $m_0(t), m_1(t)$ for the model, we calculated the decision rules given by Equations 3.4 and 3.5, sequentially at each '$t$' of the data sets taking the strength $(\alpha, \beta)$ as (0.05, 0.2). These are presented for the model in Table 5.2.

**Table 5.2: SPRT analysis for 5 data sets of Time domain data**

| Data Set | T | N(t) | Acceptance region (≤) | Rejection Region (≥) | Decision |
|---|---|---|---|---|---|
| 1 | 1 | 30.02 | 1.818603 | 4.719158 | *Accept* |
| 2 | 1 | 5.5 | -0.629760 | 2.323070 | *Reject* |
| | 2 | 7.33 | -0.491494 | 2.470496 | |
| | 3 | 10.08 | -0.285944 | 2.689878 | |
| 3 | 1 | 9 | -0.465126 | 6.390944 | *Accept* |
| | 2 | 21 | 1.867158 | 9.050913 | |
| | 3 | 32 | 3.798211 | 11.300292 | |
| 4 | 1 | 10 | -0.923786 | 4.536966 | *Continue* |
| | 2 | 19 | -0.100773 | 5.514319 | |
| | 3 | 32 | 0.994143 | 6.842165 | |
| | 4 | 43 | 1.840126 | 7.894901 | |
| | 5 | 58 | 2.884670 | 9.236447 | |
| | 6 | 70 | 3.636896 | 10.239468 | |
| | 7 | 88 | 4.638963 | 11.641133 | |

| | | | | | |
|---|---|---|---|---|---|
| | **8** | 103 | 5.368646 | 12.726648 | |
| | **9** | 125 | 6.283826 | 14.204225 | |
| | **10** | 150 | 7.125308 | 15.748486 | |
| | **11** | 169 | 7.639797 | 16.846683 | |
| | **12** | 199 | 8.258289 | 18.483307 | |
| | **13** | 231 | 8.686670 | 20.143135 | |
| | **14** | 256 | 8.873289 | 21.409309 | |
| | **15** | 296 | 8.927876 | 23.435562 | |

From the above table it is observed that a decision of either to accept or reject the system is reached well in advance of the last time instant of the data.

## VI. CONCLUSION

The table 5.2 of Time domain data as exemplified for 4 Data Sets shows that Exponential model is performing well in arriving at a decision. Out of 4 Data Sets of Time domain the procedure applied on the model has given a decision of rejection for 1, acceptance for 2 and continue for 1 at various time instant of the data as follows. Data Set #1 and #3 are accepted at 1st and 3rd instant of time respectively. Data Set #2 is rejected at 3rd instant of time. Data Set #4 is continued. Therefore, by applying SPRT on data sets it can be concluded that we can come to an early conclusion of reliable or unreliable software.

## REFERENCES


[1] Goel, A.L and Okumoto, K. (1979). "A Time Dependent Error Detection Rate Model For Software Reliability And Other Performance Measures", IEEE Transactions on Reliability, vol.R-28, pp.206-211, 1979.

[2] Pham. H., (2006). "System software reliability", Springer.

[3] Reckase, M. (1983). A procedure for Decision Making Using Tailored Testing, 238-257, New Horizons In Testing – Latent Trait Test Theory and Computerized Adaptive Testing, Ed. David J. Weiss, Academic Press, New York.

[4] Satya Prasad (2007)."Half logistic Software reliability growth model "Ph.D Thesis of ANU, India.

[5] Stieber, H.A. (1997). "Statistical Quality Control: How To Detect Unreliable Software Components", Proceedings the 8th International Symposium on Software Reliability Engineering, 8-12.

[6] Wald. A., 1947. "Sequential Analysis", John Wiley and Son, Inc, New York.

[7] Xie, M., Goh. T.N., Ranjan.P., "Some effective control chart procedures for reliability monitoring" -Reliability engineering and System Safety 77 143 -150, 2002.


**Authors**

**MURALI MOHAN.** S, working as controller of examinations in Vikrama Simhapuri University. He worked as a controller of examinations in Dravidian University for three and half years. He worked as a Principal of an affiliated college of Andhra University in Visakhapatnam (A.P). He is presently a research scholar in the Department of Computer Science of Acharya Nagarjuna University.

**Dr. R Satya** Prasad received Ph.D. degree in Computer Science in the faculty of Engineering in 2007 from Acharya Nagarjuna University, Guntur, Andhra Pradesh, India. He have a satisfactory consistent academic track of record and received gold medal from Acharya Nagarjuna University for his outstanding performance in a first rank in Masters Degree. He is currently working as Associate Professor in the Department of Computer Science & Engineering, Acharya Nagarjuna University. He has occupied various academic responsibilities like practical examiner, project adjudicator, external member of board of examiners for various Universities and Colleges in and around in Andhra Pradesh. His current research is focused on Software Engineering, Image Processing & Database Management System. He has published several papers in National & International Journals.

**Mr. G. Krishna Mohan,** working as a Reader in the Department of Computer Science, P.B.Siddhartha College, Vijayawada. He obtained his M.C.A degree from Acharya Nagarjuna University, M.Tech from JNTU, Kakinada, M.Phil from Madurai Kamaraj University and pursuing Ph.D from Acharya Nagarjuna University. He qualified AP State Level Eligibility Test. His research interests lies in Data Mining and Software Engineering. He published 16 research papers in various National and International journals.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 44-51*
*www.iosrjournals.org*

# Survey of Real Time Scheduling Algorithms

Swati Pandit [1], Rajashree Shedge [2]
[1](*Computer Department, Ramrao Adik Institute of Technology/ Mumbai University,India )*
[2](*Computer Department, Ramrao Adik Institute of Technology/ Mumbai University,India)*

***Abstract:*** *Real-Time systems are becoming pervasive. In a Real-Time System the correctness of the system behavior depends not only on the logical results of the computations, but also on the physical instant at which these results are produced. A missed deadline in hard real-time systems is catastrophic and in soft real-time systems it can lead to a significant loss .This work talks about static and dynamic scheduling algorithms for real time task. The problem of real-time scheduling spans a broad spectrum of algorithms from simple uniprocessor to highly sophisticated multiprocessor scheduling algorithms which are priority driven and divided into three classes fixed priority, dynamic priority and hybrid priority. Finally conclusion shows that Instantaneous utilization factor scheduling Algorithm gives better result in uniprocessor scheduling algorithms and Modified Instantaneous utilization factor scheduling Algorithm gives better context switching, response time and CPU utilization as compared to previous scheduling algorithms.*

***Keywords-*** *Deadline, Laxity, Utilization, Precedence, context switching.*

## I. Introduction

A Real Time Scheduling System is composed of the scheduler, clock and the processing hardware elements. In a Real-Time system, a process or task has schedulability; tasks are accepted by a real-time system and completed as specified by the task deadline depending on the characteristic of the scheduling algorithm.. A deadline is defined as the time required for a task to be processed. In a critical operation the task must be processed in the time specified by the deadline [1]. A system is said to be unschedulable when tasks cannot met the specified deadlines. Goals of the real time Scheduling algorithms are Meeting the timing constraints of the system, preventing simultaneous access to shared resources and devices, attaining a high degree of utilization while satisfying the timing constraints of the systems however this is not a primary driver, Reducing the cost of context switches caused by preemption, reducing the communication cost in real-time distributed systems; we should find the optimal way to decompose the real-time application into smaller portions in order to have the minimum communication cost between mutual portions each portion is assigned to a computer [2].

Several studies have developed optimal scheduling policies for implicit deadline task systems. So far however, studies have failed to develop effective scheduling strategies for more general task systems such as constrained deadline tasks. We argue that a narrow focus on deadline satisfaction (urgency) is the primary reason for this lack of success [3]. Different priority driven algorithms for uniprocessor such as Earliest Deadline First, Least Laxity First, and Maximum Urgency First, Instantaneous utilization factor scheduling Algorithm and their corresponding enhanced or modified version in the multiprocessor that is Earliest Deadline until Zero Laxity, Improved Least Laxity First, Modified Maximum Urgency First, Modified Instantaneous utilization factor scheduling Algorithm are explained in detail with their performance analysis.

The report proceeds with what other related work we have done before in chapter 2. Chapter 3 explains study of uniprocessor priority driven real time scheduling strategies precedes study of multiprocessor priority driven real time scheduling strategies in chapter 4. After that in chapter 5 Analysis of the all the algorithms is done with showing that how Modified Instantaneous utilization factor scheduling Algorithm is efficient, finally conclusion is produced in chapter 6 and the report ends with references that we have taken for whole work.

## II. Literature Survey

The literature survey starts with the introduction of RTS and focuses on some basic concepts of real time system and scheduling.

### 1. Real Time Systems

A real Time System is a computer-controlled mechanism in which there are strict timing constraints on the computer's actions. In another words, the correctness of the system depends not only on the logical result of computation, but also on the time at which the results are produced. Such systems can be as simple as an alarm clock or can be the most complex systems built (or considered), such as the once-planned strategic defense initiative missile defense system.

RTS can be divided further into two categories:

*Hard-Real-Timesystem* which must meet its deadline, otherwise it will cause undesirable damage or a fatal error to the system. *Soft-Real-Time System*, which has an associated deadline that is desirable but not mandatory, it still makes sense to schedule and complete the task even if it has passed its deadline. We will focus on the Hard-Real-Time system.

A RTS consists of four parts: Physical process (which is controlled by the computer for some productive end); Sensors (Converts state of physical process into information *analog or digital*); Computer (Based on information from sensors, deduces state of physical process and issues commands to control the process) and Actuators (In response to commands issued computer, modifies the physical process)sa shown in figure 2.1

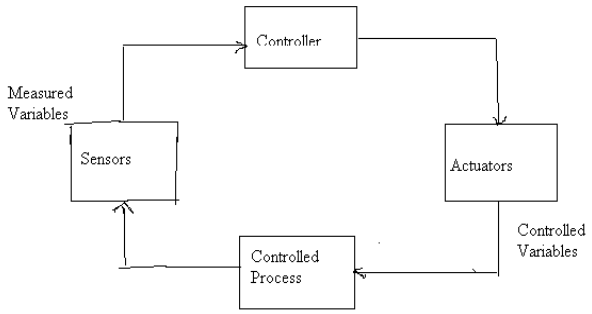



**Fig.1. Real Time System Architecture**

**Characteristics of Real Time Systems:**

- Determinism: An operating system is deterministic to the extent that it performs operations at fixed, predetermined times or within predetermined time intervals.
- Responsiveness: Responsiveness is concerned with how long, after acknowledgment, it takes an operating system to service the interrupt.
- User control: A real-time system may also allow the user to specify such characteristics as the use of paging or process swapping, what processes must always be resident in main memory, what disk transfer algorithms are to be used, what rights the processes in various priority bands have and so on.
- Reliability: Reliability in real-time system is responding to and controlling events in real time, loss or degradation of performance may have catastrophic consequences, ranging from financial loss to major equipment damage and even loss of life.
- Fail-soft operation: the ability of a system to fail in such a way as to preserve as much capability and data as possible.

## 2 Real Time Task Model

### 2.1 Task Description

A task model has been defined with the main timing parameters. A task is defined by chronological parameters denoting delays and by chronometric parameters denoting times. The model includes primary and dynamic parameters as shown in figure 2.2. Primary parameters are:

• *r*, task release time, i.e. the triggering time of the task execution request.
• *C*, task worst-case computation time, when the processor is fully allocated to it.
• *D*, task relative deadline, i.e. the maximum acceptable delay for its processing.
• *T*, task period (valid only for periodic tasks).

When the task has hard real-time constraints, the relative deadline allows computation of the absolute deadline $d = r + D$.

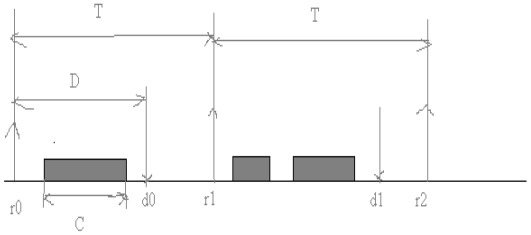



**Fig.2. Task Model (Timing Diagram)**

The parameter T is absent for an aperiodic task. A periodic task is modeled by the four previous parameters. Each time a task is ready, it releases a periodic request. The successive release times (also called request times, arrival times or ready times) are request release times at rk = r0 + kT, where r0 is the first release and rk the k + 1th release; the successive absolute deadlines are dk = rk + D. If D = T , the periodic task has a relative deadline equal to period. A task is well formed if $0 < C \leq D \leq T$ [4].

**2.2 Task Characteristics**

- **Arrival Patterns**

Based on their arrival patterns over time, tasks can be divided into two categories: periodic and aperiodic. A task is said to periodic when it is available for execution, regardless of processor availability and inter-task dependencies. Aperiodic tasks have arbitrary arrival patterns.

- **Inter-Task Dependencies**

Depending on the system's definition of the task, there may be dependencies among tasks. A task j is defined to be dependent on a set of tasks S (j) if j cannot start executing until all tasks in S (j) have completed their execution.

- **Deadline Laxity**

All real-time tasks define a deadline by which they have to complete executing. However, the problem of finding an optimal schedule for real-time tasks has been shown to be NP-complete.

## III. Uniprocessor Scheduling Algorithms

**1 Earliest-Deadline-First Scheduling (EDF)**

EDF Scheduling is proposed by Liu and Leyland [6]. This Scheduling Algorithm is based on uniprocessor dynamic-priority preemptive scheme and optimal among all dynamic priority scheduling algorithm. This Algorithm schedules the task with the earliest deadline first.

The algorithm of EDF scheduling is following as;

Step 1: Set up all tasks' start time, end time, remaining time and deadline.

Step 2: If system is idle, then add task to schedule and go to step 4. Otherwise go to step 3.

Step 3: If new task's deadline is earlier than processing task's deadline, then update processing task's remaining time and exchange tasks. Otherwise update new task's start time.

Step 4: If all the tasks have not been scheduled, then go to step 2. Otherwise stop.

**2 Least Laxity First (LLF)**

David B. Stewart and Pradeep K. Khosla proposed this algorithm [7]. The minimum-laxity- first algorithm assigns a laxity to each task in a system, and then selects the task with the minimum laxity to execute next. This algorithm is also called as Least Laxity First (LLF).Laxity is defined as follows:

laxity=deadline_time- current_time - CPU_time_still_needed

Laxity is a measure of the flexibility available for scheduling a task. A laxity of *tl* means that even if the task is delayed by *tl* time units, it will still meet its deadline. A laxity of zero means that the task must begin to execute now or it will risk failing to meet its deadline. The laxity of a process is defined as the deadline minus remaining computation time. The algorithm gives the highest priority to the active job with the smallest laxity [5]. Then the job with the highest priority is executed. While a process is executing, it can be preempted by another whose laxity has decreased to below that of the running process. A problem arises with this scheme when two processes have similar laxities. One process will run for a short while and then get preempted by the other and vice versa. Thus, many context switches occur in the lifetime of the processes. The least laxity first algorithm is an optimal scheduling algorithm for systems with periodic real-time tasks. If each time a new ready task arrives, it is inserted into a queue of ready tasks, sorted by their laxities.

**3 Maximum-Urgency-First scheduling algorithm (MUF)**

David B. Stewart and Pradeep K. Khosla proposed this algorithm [7]. The maximum-urgency-first scheduling algorithm which we have developed is a combination of fixed and dynamic priority scheduling, also called **mixed priority** scheduling. With this algorithm, each task is given urgency. The urgency of a task is defined as a combination of two fixed priorities, and a dynamic priority. One of the fixed priorities, called the **criticality,** has higher precedence over the dynamic priority. The other fixed priority, which we call **user priority,** has lower precedence than the dynamic priority. The dynamic priority is inversely proportional to the laxity of a task.

The MUF algorithm consists of two parts. The first part is the assignment of the criticality and user priority, which is done apriori. The second part involves the actions of the MUF scheduler during run-time.

The steps in assigning the criticality and user priority are the following:

- As with RM, order the tasks from shortest period to longest period.

- Define the critical set as the first N tasks such that the total worst-case CPU utilization does not exceed 100%. Assign **high** criticality to all tasks in the critical set, and **low** criticality to all other tasks.
- Optionally assign a unique user priority to every task in the system.

The static priorities are defined once, and do not change during execution. The dynamic priority of each task is assigned at run-time, inversely proportional to the laxity of the task. Before its cycle, each task must specify its desired start time, deadline time, and worst-case execution time. Whenever a task is added to the ready queue, a reschedule operation is performed.

MUF scheduler is used to determine which task is to be selected for execution, using the following algorithm:

1). Select the task with the highest criticalness.

2). If two or more tasks share highest criticalness, then select the task with the highest dynamic priority (i.e. minimum laxity). Only tasks with pending deadlines have a non- zero dynamic priority. Tasks with no deadlines have a dynamic priority of zero.

3). If two or more tasks share highest criticalness, and have equal dynamic priority, then the task among them with the highest user priority is selected.

4). If there are still two or more tasks that share highest criticalness, dynamic priority, and highest user priority, then they are serviced in a *first-come-first-serve* manner.

**4 Instantaneous utilization factor scheduling Algorithm**

Radhakrishna Naik proposed this algorithm targets the soft real time systems. It schedules the periodic tasks. This algorithm is based on instantaneous Utilization Factor [IUF]. Instantaneous utilization Factor [IUF] is the processor utilization of a task at any time instant. Here for every task after one time quantum the instantaneous utilization of task is computed. The task having highest instantaneous utilization is given the highest priority and task is given to the processor. The quantum for which task is applied to CPU is Qi. The total sum of quantum of all task (for that quantum iteration only) is $\Sigma Qi=Q$. Here, the Periodic task cycle (PTC) is computed for a given task set which is the LCM of period of invocation of all tasks. After computation of initially the CPU utilization is computed. CPU Utilization is the ratio of initial computation time and its initial period. Depending on this initial utilization the task having highest initial utilization is mapped to the CPU. In a given PTC one task has executed for one time quantum. Now for calculating new instantaneous utilization factor gain the value of computation time and period is computed. For computing new computation time, the total sum of quantum of all tasks is subtracted from the previous computation time. In a similar way the new period is calculated. After this, the new instantaneous utilization is computed. Likewise the process will be repeated [6].

Algorithm for IUF:

7. Initially for given task set, calculate CPU utilization of each task using formula

$U^i_0 = C^i_0 / P^i_0$ (1)

$U^i_0$ =Initial utilization of ith task.

$C^i_0$ =Initial computation time.

$P^i_0$ =Initial period of invocation.

Based on utilization [ $U^i_0$] the task which is having higher value of Utilization is mapped for the CPU.

**2)**: Now in a given PTC, one task has executed for one quantum of time. Again calculate value of $C^i_1$ , $P^i_1$ by using following formula-

$C^i_1=C^i_0-Q_i$ (2)

$P^i_1=P^i_0-Q$ (3)

Where $\Sigma qi=Q$.

Then calculate new Instantaneous Utilization factor for ith task for using formula 1 and 2

$U^i_1=C^i_1/P^i_1$. (4)

Where,

$C^i_1$ =Instantaneous computation time for ith task

$P^i_1$= Instantaneous period of execution of ith task.

$U^i_1$=Instantaneous Utilization of ith task.

For second iteration of time derive the table (Ti, $C^i_1$, $P^i_1$, $U^i_1$)

Again the task which is having highest instantaneous utilization will be having highest priority of execution for second iteration quantum. Likewise, calculate

$C^i_j=C^i_{j-1}-Q_i$ (5)

$P^i_j=P^i_{j-1}-Q$ (6)

Calculate U using equation 4 and 5

$U^i_j=C^i_j/P^i_j$. (7)

Where,

$U^i_j$ =Instantaneous utilization of ith task for the jth iteration of quantum

j = PTC end point.

**3)**: Hence task sequence in first PTC is derived. It is observed that at every step we can check whether the instantaneous utilization is less than initial utilization $U^i_0$. If at any given instant of time, it is observed that it is greater than $U^i_0$, it means that task is going to miss its deadline.

## IV. Multiprocessor Scheduling Algorithms

**1. Earliest Deadline first until Zero Laxity (EDZL)**

Yi-Hsiung Chao proposed this algorithm [10]. EDF cannot guarantee all periodic tasks to meet their deadlines if the total utilization is greater than (m+1)/2 when the system has *m* processors. On the other hand, LLF has a good schedulability condition but has a high context switching overhead and causes a large number of preemptions.

Earliest Deadline first until Zero Laxity (EDZL) algorithm which combines the EDF and MLF algorithms. EDZL schedules jobs based on their deadlines and laxities. When all jobs have positive laxities, EDZL schedules jobs according to EDF. Whenever the laxity of a job becomes zero, EDZL schedules the job with the highest priority. Algorithm for EDZL

1). While true do
2). If there is any job with zero laxity
3). If there is an idle processor
4). Execute the zero-laxity job on the idle processor
5). Else if there is any task with positive laxity
6). Preempt the task with largest deadline
7). Else Output "Fail to Schedule"
8). Else Perform EDF
9). End if
10). End While

**2. Improved Least Laxity First Scheduling Algorithm (ILLF)**

H.S Behera, Satyajit Khuntia & Soumyashree Nayak proposes the Improved Least Laxity First Scheduling Algorithm [11]. This algorithm with intelligence time slice finds the time quantum by taking the greatest common divisor (GCD) of all the execution time of the processes. After every unit of time slice the laxity of each remaining process (present in the ready queue) is calculate. The loop is continued until all the processes are being executed by the CPU. Here as the GCD of execution the execution time is always greater than equal to 1 so loop will be continue for lesser no of time or same no of time Our proposed ILLF algorithm shows less context switching to least laxity First algorithm.

The algorithm is as follows:

1). Initialize all the processes in ready queue.
2). Initialize ready queue.
3). Calculation of the Time Quantum using GCD of the execution time of all the processes.
4). If execution time of process Pi is 0 then remove the process form ready queue and goto Step 5 otherwise calculate laxity time of the Pi and goto step 6.
5). If n==0 then stop and Exit otherwise goto step 4.
6). If i<n then increment i and goto 4 otherwise goto 7.
7). Sort P1 to Pn according to laxity in ascending order goto 8.
8). If (Lj==Lj+1) then goto 9; otherwise Execute Pj for
TQ time;
for (k=0;k<n; k++)
{Dk=Dk – TQ ;}
Ej=Ej-TQ; goto 4;
9). If (Ej>=Ej+1) then goto 9;
Otherwise {Execute Pj+1 till completion
for (k=0;k<n; k++)
{Dk=Dk-Ej+1; Ej+1=0;
goto 4 ;}
10). Execute Pj till completion
for (k=0; k<n; k++)
{Dk=Dk-Ej; Ej=0; goto 4 ;}

**3. Modified Maximum Urgency First Scheduling Algorithm (MMUF)**

V. Salmani proposes this algorithm [8]. The MMUF algorithm consists of two phases with the following details:

The algorithm is as follows:

Phase-1:- This phase defines the fixed priorities. These priorities should not be changed any further.
The steps are same as we discussed in MUF algorithm.
Phase-2:-This phase defines the dynamic priority. Dynamic priority is calculated at each clock cycle. Dynamic priorities are set according to EDZL algorithm which is explained in the following steps.
1*)* If there is only 1 critical task, schedule it at any free processor without pre-emption.
2) If more than 1 critical task is present in the ready queue, schedule the tasks with earliest deadline first (EDF) scheduling algorithm until there is a task with Zero laxity.
3) Laxity at each clock cycle is computed for all the remaining processes in the ready queue
4) If processes with zero laxity are available then these processes are assigned with higher priority over other process having non-zero laxity.
5) The process with zero laxity is scheduled at any available free processor without preemption.
6) If no free processor is available, and zero laxity process is present in the ready queue then preemption occurs. The process with the longest deadline is preempted and the zero laxity process is assigned to that processor.
7) After the execution of all zero laxity processes the remaining processes in the ready queue are assigned to the processor as per EDF scheduling policy.
The above steps are performed again for the non-critical task set.

**4. Modified Instantaneous utilization factor scheduling Algorithm (MIUF)**
Algorithm for MIUF is as follows [9]:
1)**:** Take input of tasks containing period, mandatory execution time, and optional execution time.
2)**:** Calculate the mandatory utilization and optional utilization for each task.
3): Check the schedulability for tasks according to their utilization.
4)**:** Generate CPU mapping for tasks.
5)**:** Execute the mandatory portion of tasks according to the highest instantaneous utilization. Meanwhile if there is interrupt due to corruption of task while executing the mandatory portion of task then a backup image is maintained so that the task can be restored from the image.
6): After executing mandatory portion of all tasks execute optional portion according the Shortest job first policy.
7)**:** If interrupt occurs due to corruption of task in the optional portion of task then optional portion does not run to its completion and gets aborted.

## V. Comparison of Real Time scheduling algorithms

Before doing comparison let us see the definitions of all Performance Metrics:

**1. Performance parameters of the Scheduling algorithms.**

- **CPU Utilization**
  In the Multiprocessor Real time Operating System, CPU utilization is the parameter which is very important. All the processors in the system must be effectively utilized.
- **Deadline Miss Chances**
  Number of times the task may fail to execute to given time or to reach the deadline.
- **Number of context switching**
  It measures the number of preemptions of a task by a Higher Priority task.
- **Response Time**
  Response time of a task or thread is defined as the time elapsed between the dispatch (time when task is ready to execute) to the time when it finishes its job (One dispatch).
- **Effectiveness**
  Effectiveness of a Scheduler can be measured on any factor for example, Energy.
- **Predictability**
  Predictability for these systems should mean that we are able to satisfy the timing requirements of critical tasks with 100% guarantee over the life of the system, be able to assess overall system performance over various time frames (a stochastic evaluation), and be able to assess individual task and task group performance at different times and as a function f the current system state [14]. If all these assessments meet the timing requirements, then the system is predictable with respect to its timing requirements.
- **Optimality**

Scheduling algorithm is referred to as optimal if it can schedule all of the task sets that can be scheduled by any other algorithm, that is, all of the feasible task sets.

**Table 1. Comparative Analysis of scheduling algorithms**

| Algorithms / Performance Metric | Uniprocessor Algorithms | | | | Multiprocessor Algorithms | | | |
|---|---|---|---|---|---|---|---|---|
| Priority | EDF D | LLF D | MUF Hybrid | IUF D | EDZL D | ILLF D | MMUF Hybrid | MIUF D |
| CPU Utilization | High | High | High | High | High | High | High | High |
| No of Context Switching | Less | High | High | High | Very less | Less | Less | Less |
| Optimal | Yes | Yes | for critical tasks | Yes | No | Yes | Yes | Yes |
| Deadline miss chances | Average | Average | Less | Less | Less | Less | Less | Very Less |
| Response Time | High | Average | Low | High | Low | High | Average | Low |
| Predictability | Not Predictable | Not Predictable | Predictable under transient load | Dynamic predictability | More Predictable than EDF | More Predictable | Predictable under transient load | Dynamic predictability |
| Effectiveness | Optimal, Easy to implement | Takes execution time into cosideration | Work in transient overload | maximize utilization bound of schedule | Context switching overhead is low | Less context switching | Optimal for noncritical tasks | Improves context switching, response time and CPU utilization |
| Limitations | Not Work in overload, not optimal for pro.>1 | In laxity tie, more context switches occurs | Non critical task may miss deadline | Context switching is very high | chances of deadline miss of the critical tasks | Excut-ion time is more | Only consider static utilization of task set | |

## VI. Conclusion

As we have seen, all uniprocessor and their extended versions for multiprocessor scheduling algorithms are priority based. The comparative analysis shows that In EDF context switching is high in overload condition whereas EDZL removes that drawback, in LLF When laxity occurs between two tasks context switching taken place is more that disadvantage is overcome in ILLF, whereas the way of assigning static priorities to the tasks is improved in MMUF, the Instantaneous Utilization Factor scheduling algorithm gives better result in uniprocessor scheduling and Modified Instantaneous Utilization Factor scheduling algorithm gives better context switching, response time and CPU utilization as compared to previous multiprocessor scheduling algorithms.

## Acknowledgements

We wish to give our heartiest gratitude to Dr. Leena Ragha, Head of the Department of Computer Engineering, Ramrao Adik Institute of Technology and Prof. Vanita Mane, PG coordinator Department of Computer Engineering for her constant motivation, knowledge sharing and support behind this paper.

## References

[1] Jashweeni Nandanwar, Urmila Shrawankar, "An Adaptive Real Time Task Scheduler", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, November 2012.

[2] H. S. Behera, Naziya Raffat, Minarva Mallik, **"** A Modified Maximum Urgency First Scheduling Algorithm with EDZL for Multiprocessors in Real Time Applications", Volume 2, Issue 4, April 2012.

[3] Jinkyu Leea, Arvind Easwaranb, Insik Shina,∗ , Insup Lee, "Zero-laxity based real- time multiprocessor scheduling", The Journal of Systems and Software, 2011.

[4] Francis Cottet,"Scheduling in Real Time Syatem", John W. Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2002.

[5] Arezou Mohammadi and Selim G. Akl, "Scheduling Algorithms for Real-Time Systems", Technical Report No. 2005-499, July 15, 2005.

[6] Yoo, Myungryun and M. Gen, "Study on Scheduling for Real- time Task by Hybrid Multiobjective Genetic Algorithm", Thesis, 2006.

[7] David B. Stewart, Pradeep Khosla, "Real-Time Scheduling of Sensor-Based Control Systems", 1991.

[8] Komal S. Bhalotiya, "Customised Multiprocesssor Scheduling Algorithms For Real time Systems", Proceedings published by International Journal of Computer Applications® (IJCA) ISSN: 0975 – 8887, April, 2012.

[9] Radhakrishna Naik, "Instantaneous Utilization Based Scheduling Algorithms for Real Time Systems", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2), 654-662, 2011.

[10] Yi-Hsiung Chao, Shun-Shii Lin, Kwei-Jay Lin, "Schedulability issues for EDZL scheduling on real-Time multiprocessor systems", 2008.

[11] H.S Behera, Satyajit Khuntia & Soumyashree Nayak, "An Improved Least-Laxity-First Scheduling Algorithm For Real-Time Tasks", International Journal of Engineering Science and Technology (IJEST), Vol. 4 No.04 April 2012.

[12] G. Umarani Srikanth, A. P. Shanthi, V. Uma Maheswari, "A Survey on Real Time Task Scheduling", Europeans Journal of Scientific Research ISSN 1450-216X Vol.69 No.1 (2012), pp.33-41 © EuroJournals Publishing, Inc. 2012.
[13] Farooq Muhammad, "Ordonnancement de T^aches E_cace _aComplexit_e Ma^_tris_ee pouSyst_emesTempsR_eel' Pd.D. thesis, 2010.
[14] Robert I. Davis and Alan Burns, "A Survey of Hard Real- Time Scheduling for Multiprocessor Systems", ACM Computing Surveys, Vol. 43, No. 4, Article 35, Publication date October 2011.
[15] S.Behera, Naziya Raffat, Minarva Mallik, "Enhanced Maximum Urgency First Algorihm with Intelligent Laxity Real Time Systems", International Journal of Computer Applications (0975 – 8887), Volume 44– No.8, April 2012.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 52-56*
*www.iosrjournals.org*

# A Study of Image Compression Methods

ANJU
*M.Tech(CSE), Hindu College Of Engineering*

***Abstract:*** *Image compression is now essential for applications such as transmission and storage in data bases. In this paper we review and discuss about the image compression, need of compression, its principles, and types of compression and various algorithm of image compression. This paper attempts to give a recipe for selecting one of the popular image compression algorithms based on Wavelet, DCT, and VQ. We review and discuss the implementation of these algorithms.*
*Initially, the colored image of any resolution is selected by the user which gets converted into grayscale image to be taken as an input image for both the techniques. Finally the Performance analysis of the image compression using self organized feature maps and JPEG 2000 Algorithm.*
***Keywords: -*** *Image Compression, Types, SOM, JPEG2000, Algorithms.*

## I. Introduction

Early evidence of image compression suggests that this technique was, in the beginning, most commonly used in the printing, data storage and telecommunications industries. Now-a-days, the digital form of image compression is also being put to work in industries such as fax transmission, satellite remote sensing and high definition television, to name but a few. In certain industries, the archiving of large numbers of images is required. A good example is the health industry, where the constant scanning and/or storage of medical images and documents take place. Image compression offers many benefits here, as information can be stored without placing large loads on system servers [1]. Depending on the type of compression applied, images can be compressed to save storage space, or to send to multiple physicians for examination and these images can uncompress when they are ready to be viewed, retaining the original high quality and detail that medical imagery demands.

Image compression is also useful to any organization which requires the viewing and storing of images to be standardized, such as a chain of retail stores or a federal government agency. In the retail store example, the introduction and placement of new products or the removal of discontinued items can be much more easily completed when all employees receive, view and process images in the same way. Federal government agencies that standardize their image viewing, storage and transmitting processes can eliminate large amounts of time spent in explanation and problem solving. The time they save can then be applied to issues within the organization, such as the improvement of government and employee programs. In the security industry, image compression can greatly increase the efficiency of recording, processing and storage. However, in this application it is imperative to determine whether one compression standard will benefit all areas. For example, in a video networking or closed-circuit television application, several images at different frame rates may be required. Time is also a consideration, as different areas may need to be recorded for various lengths of time. Image resolution and quality also become considerations, as does network bandwidth and the overall security of the system. Museums and galleries consider the quality of reproductions to be of the extreme importance. Image compression, therefore, can be very effectively applied in cases where accurate representations of museum or gallery items are required, such as on a web site. Detailed images which offer short download times and easy viewing benefit all types of visitors, from the student to the discriminating collector. Compressed images can also be used in museum or gallery kiosks for the education of that establishment's visitors. In a library scenario, students and enthusiasts from around the world can view and enjoy a multitude of documents and texts without having to incur traveling or lodging costs to do so. Regardless of industry, image compression has virtually endless benefits wherever improved storage, viewing and transmission of images are required.

The basic idea behind the method of compression is to treat a digital image as an array of numbers i.e., a matrix. Each image consists of a fairly large number of little squares called **pixels** (picture elements). The matrix corresponding to a digital image assigns a whole number to each pixel. For example, in the case of a 256x256 pixel gray scale image, the image is stored as a 256x256 matrix, with each element of the matrix being a whole number ranging from 0 (for black) to 225 (for white). The JPEG compression technique divides an image into 8x8 blocks and assigns a matrix to each block [9].
There are two types of image compression algorithm which can be used to compress the image: Lossless Compression Algorithm & Lossy Compression Algorithm.

**1.1 Lossless Compression Technique**

In lossless compression techniques, the original image can be perfectly recovered from the compressed (encoded) image. These are also called noiseless since they do not add noise to the signal (image). It is also known as entropy coding since it use statistics/decomposition techniques to eliminate/minimize redundancy. Following techniques are included in lossless compression:

- Run length encoding
- Huffman encoding

**1.1.1Run Length Encoding**

Run Length encoding performs lossless data compression which is a very simple compression method used for sequential data. It is very useful in case of repetitive data. It replaces sequences of identical symbols (pixels), called runs by shorter symbols. The run length code for a gray scale image is represented by a sequence {Vi , Ri} where Vi is the intensity of pixel and Ri refers to the number of consecutive pixels with the intensity Vi as shown in the figure. If both Vi and Ri are represented by one byte, this span of 12 pixels is coded using eight bytes yielding a compression ratio n of 1: 5

| 8 | 8 | 8 | 8 | 8 | 7 | 7 | 7 | 9 | 9 |
|---|---|---|---|---|---|---|---|---|---|

| {8,5} | {7,3} | {90,2} |
|---|---|---|

Fig.1: Run –Length Encoding

**1.1.2 Huffman Encoding**

Huffman encoding based on their statistical occurrence frequencies (probabilities). The pixels in the image are treated as symbols. The symbols that occur more frequently are assigned a smaller number of bits, while the symbols that occur less frequently are assigned a relatively larger number of bits. Huffman code is a prefix code. This means that the (binary) code of any symbol is not the prefix of the code of any other symbol. Most image coding standards use lossy techniques in the earlier stages of compression and use Huffman coding as the final step.

**1.2 Lossy compression technique**

Lossy schemes provide much higher compression ratios than lossless schemes. Lossy schemes are widely used since the quality of the reconstructed images is adequate for most applications.
Lossy compression techniques includes following schemes:

- Transformation coding
- Vector quantization
- Fractal coding
- Block Truncation Coding
- Subband coding

**1.2.1 Transformation Coding**

In this coding scheme, transforms such as DFT (Discrete Fourier Transform) and DCT (Discrete Cosine Transform) are used to change the pixels in the original image into frequency domain coefficients (called transform coefficients).These coefficients have several desirable properties. One is the energy compaction property that results in most of the energy of the original data being concentrated in only a few of the significant transform coefficients. This is the basis of achieving the compression. Only those few significant coefficients are selected and the remaining is discarded. The selected coefficients are considered for further quantization and entropy encoding. DCT coding has been the most common approach to transform coding.

**1.2.2 Vector Quantization**

The basic idea in this technique is to develop a dictionary of fixed-size vectors, called code vectors. This technique is a lossy compression technique. A vector is usually a block of pixel values. A image is then partitioned into non-overlapping blocks called image vectors. Thus, each image is represented by a sequence of indices that can be further entropy coded. After the quantization technique, the Zigzag reordering is done.

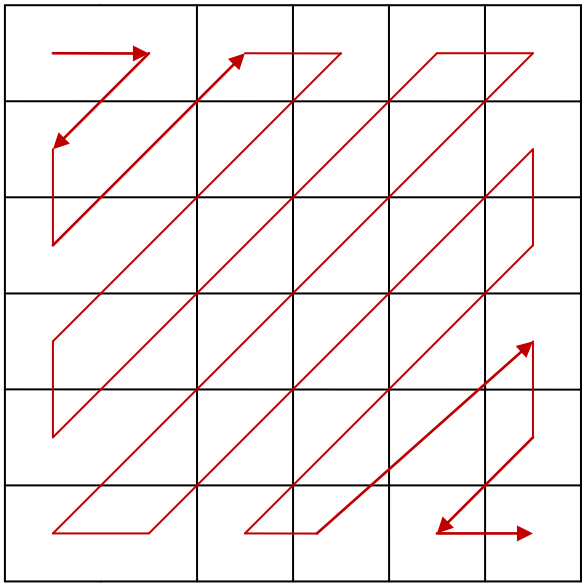


Fig.2: Zigzag ordering of JPEG image components

**1.2.3 Fractal Coding**

The essential of Fractal Coding is to decompose the image into segments by using standard image processing techniques such as color separation, edge detection, and spectrum and texture analysis. Then each segment is looked up in a library of fractals. The library actually contains codes called iterated function system (IFS) codes, which are compact sets of numbers. Using a systematic procedure, a set of codes for a given image are determined, such that when the IFS codes are applied to a suitable set of image blocks yield an image that is a very close approximation of the original. This scheme is highly effective for compressing images that have good regularity and self-similarity.

**1.2.4 Block truncation coding**

In this, the image is divided into non overlapping blocks of pixels. For each block, threshold and reconstruction values are determined. The threshold is usually the mean of the pixel values in the block. Then a bitmap of the block is derived by replacing all pixels whose values are greater than or equal (less than) to the threshold by a 1 (0). Then for each segment (group of 1s and 0s) in the bitmap, the reconstruction value is determined. This is the average of the values of the corresponding pixels in the original block.

**1.2.5 Sub band coding**

In this scheme, the image is analyzed to produce the components containing frequencies in well-defined bands, the sub bands. Subsequently, quantization and coding is applied to each of the bands. The advantage is that the quantization and coding well suited for each of the sub bands can be designed separately [4].

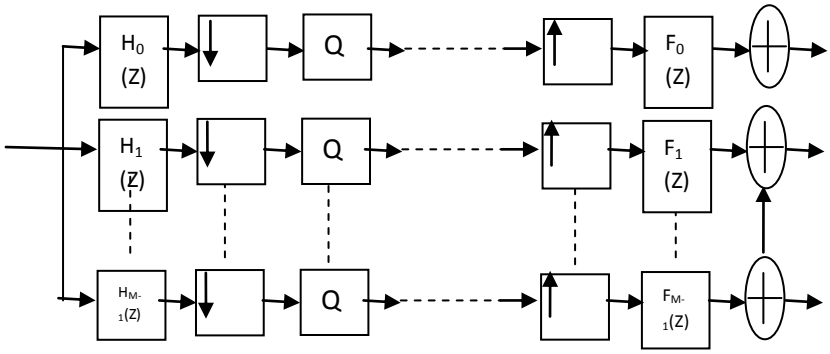



Fig.3: Block Diagram of Sub Band Coding

## II. Algorithm for Image Compression using JPEG 2000 Standard

The Process of JPEG 2000 Standard is:

1. The Image is broken into 8x8 Blocks of Pixels.
2. Working from Left to Right, Top to Bottom, the DCT is applied to each Block.
3. Each Block is compressed through Quantization.
4. The array of compressed blocks that constitute the image is stored in a drastically reduced amount of space.
5. When desired, the image is reconstructed through decompression, a process that uses the Inverse discrete Cosine Transformation (IDCT).

## III. Applications of JPEG 2000

1. Consumer applications such as multimedia devices (e.g., digital cameras, personal digital assistants, 3G mobile phones, color facsimile, printers, scanners, etc.)
2. Client/server communication (e.g., the Internet, Image database, video streaming, video server, etc.)
3. Military/surveillance (e.g., HD satellite images, Motion detection, network distribution and storage, etc.)
4. Medical Imaging.
5. Remote sensing
6. High-quality frame-based video recording, editing and storage [7].

## IV. SOM

A **self-organizing map** (**SOM**) or **self-organizing feature map** (**SOFM**) is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional, discredited representation of the input space of the training samples, called a **map**. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space.

SOFM learn to classify input vectors according to how they are grouped in the input space. They differ from competitive layers in that neighboring neurons in the SOM learn to recognize neighboring sections of the input space. Thus, self-organizing maps learn both the distribution and topology of the input vectors they are trained on [2]. It consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a regular spacing in a hexagonal or rectangular grid. It describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to first find the node with the closest weight vector to the vector taken from data space. Once the closest node is located it is assigned the values from the vector taken from the data space.

While it is typical to consider this type of network structure as related to feed forward networks where the nodes are visualized as being attached, this type of architecture is fundamentally different in arrangement and motivation. Useful extensions include using steroidal grids where opposite edges are connected and using large numbers of nodes [5]. It has been shown that while SOM with a small number of nodes behave in a way that is similar to K-means, larger SOM rearrange data in a way that is fundamentally topological in character. It is also common to use the U-Matrix. The U-Matrix value of a particular node is the average distance between the node and its closest neighbors [6]. In a square grid for instance, we might consider the closest 4 or 8, or six nodes in a hexagonal grid.

The principal goal of SOM is to transform an incoming signal pattern of arbitrary dimension into a one- or two- dimensional discrete map, and to perform this transformation adaptively in a topologically ordered fashion. Fig.4 shows the schematic diagram of a two-dimensional lattice of neurons commonly used as a discrete map. Each neuron in the lattice is fully connected to all source nodes in the input layer [6]. This network represents a feed forward structure with a single computational layer consisting of neurons arranged in row and columns.

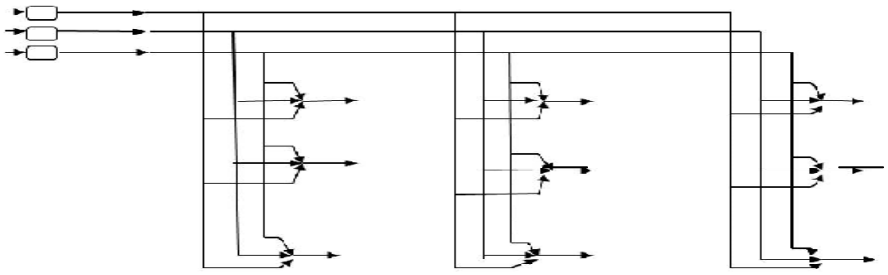


Fig.4: Two- Dimensional lattice of neurons

**4.2. Algorithm for SOM**

The algorithm responsible for the formation of the SOM proceeds first by initializing the synaptic weights in the network. This can be done by assigning them small values picked from a random number generator. Once the network has been properly initialized, there are three essential processes involved in the formation of the SOM, as summarized as:

1. Competition: For each input pattern, the neurons in the network compute their respective values of a discriminate function. This function provides the basis for competition among neurons. The particular neuron with the largest value of discriminate function is declared winner of the competition.
2. Cooperation: The winning neuron determines the spatial location of a topological neighborhood of excited neurons, providing the basis for cooperation among such neighboring neurons.
3. Synaptic Adaptation: This last mechanism enables the excited neurons to increase their individual values of the discriminated function in relation to the input pattern through suitable arrangements applied to their synaptic weights. The essential parameters of the algorithm are:
    i. A continuous input space of activation patterns that are generated in accordance with a certain probability distribution.
    ii. A topology of the network in the form of a lattice of neurons, which defines a discrete output space.
    iii. A time varying neighborhood function hj,i(x)(n) that is defined around a winning neuron i(x).
    iv. A learning rate parameter ŋ(n) that starts at an initial value ŋ0 and then decreases gradually with time, n, but never goes to zero.

**4.3Learning Vector Quantization (VQ)**

A type of neural network consisting of a set of vectors, the position of which are optimized with respect to a given data set. The network consists of an input and output layer, [8] with vectors storing the *connection weights* leading from input to output neurons. The learning method of learning vector quantization is also called "competition learning." For each training pattern cycle an input neuron finds the closest vector, and selects the corresponding output neuron as the "winner neuron." The weights of the connection to the winner are then adapted - either closer to or farther away from the training pattern, based on the class of the neuron. This movement is controlled by a *learning rate* parameter. It states how far the reference vector is moved. Usually the learning rate is decreased in the course of time, so that initial changes are larger than changes made in later epochs of the training process. Learning may be terminated when the positions of the reference vectors hardly change any more.

## References


[1.] Banerjee and A. Halder,"An Efficient Dynamic Image Compression Algorithm based on Block Optimization, Byte Compression and Run-Length Encoding along Y-axis", IEEE Transaction, 2010.

[2.] Y. H. Dandawate and M.A. Joshi, "Performance analysis of Image Compression using Enhanced Vector Quantizer designed with Self Organizing Feature Maps: The Quality perspective" IEEE transaction, pg. 128-132, 2007.

[3.] S. Immanuel Alex Pandian and J.Anitha,"A Neural Network Approach for Color Image Compression in Transform Domain", International Journal of Recent Trends in Engineering, Vol 2, No. 2, November 2009, 152.

[4.] H. G. Lalgudi, A. Bilgin, M. W. Marcellin, and M. S. Nadar, "Compression of Multidimensional Images Using JPEG2000", IEEE Signal processing, vol. 15, pg. 393-396, 2008.

[5.] Cheng-Fa Tsai, Chen-An Jhuang, Chih-Wei Liu, "Gray Image Compression Using New Hierarchical Self-Organizing Map Technique, IEEE conference, 2008.

[6.] D. Kumar, C.S. Rai and S. Kumar, "Face Recognition using Self-Organizing Map and Principal Component Analysis", IEEE Transaction, 2005.

[7.] Gregory K. Wallace, "The JPEG Still Picture Compression Standard", IEEE Transactions on Consumer Electronics, December 1991.

[8.] S. Anna Durai & E. Anna Saro, "An Improved Image Compression approach with Self-Organizing Feature Maps", GVIP Journal, Volume 6, Issue 2, September, 2006.

[9.] Sonal, Dinesh Kumar, "A Study of various Image Compression Techniques", IEEE Transaction, 2005.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 57-59*
*www.iosrjournals.org*

# Jamming Anticipation and Convolution through Immaculate Hiding Process of Packets

T. Sandeep[1], Ms.P.Subhadra[2]
[1](*Master of Technology, Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad, India,*
[2](*Associate Professor Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad, India)*

***Abstract:*** Cached data not only replies local access, but also replies data request issued from other nodes. Wireless Mesh Networks (WMNs) have emerged as an important technology in building next generation fixed wireless broadband networks that provide low cost Internet access for fixed and mobile users. Reduce the number of hops that request/data need to travel in the network. In these attacks, the adversary selectively targets specific packets of "high" importance by exploiting his knowledge on the implementation details of network protocols at various layers of the protocol stack. We illustrate the impact of selective jamming on the network performance by illustrating various selective attacks against the TCP protocol. We show that such attacks can be launched by performing real-time packet classification at the physical layer. We study the idealized case of perfect knowledge by both the jammer and the network about the strategy of one another, and the case where the jammer or the networks lack this knowledge. The latter is captured by formulating and solving optimization problems, the solutions of which constitute best responses of the attacker or the network to the worst-case strategy of each other.

***Keywords:*** *Denial-of-service, jamming, Wireless network, packet classification*

## I. Introduction

Wireless networks are built upon a shared medium that makes it easy for adversaries to launch jamming-style attacks. These attacks can be easily accomplished by an adversary emitting radio frequency signals that do not follow an underlying MAC protocol. Jamming attacks can severely interfere with the normal operation of wireless networks and, consequently, mechanisms are needed that can cope with jamming attacks. As these networks gain popularity, providing security and trustworthiness will become an issue of critical importance. Many wireless security threats may be addressed through appropriately designed network security architectures which are essentially modifications of traditional security services, such as confidentiality, authentication, and integrity to the wireless domain. Wireless networks, however, are susceptible to threats that are not able to be adequately addressed via cryptographic methods. One serious class of such threats are attacks of radio interference. This exposes them to passive and active attacks, which are different in their nature and objectives. In the former, a malicious entity does not take any action except passively observing ongoing communication, e.g. eavesdropping so as to intervene with the privacy of network entities involved in the transaction. On the other hand, an active attacker is involved in transmission as well. Depending on attacker objectives, different terminology is used. If the attacker abuses a protocol with the goal to obtain performance benefit itself, the attack is referred to as misbehavior. If the attacker does not directly manipulate protocol parameters but exploits protocol semantics and aims at indirect benefit by unconditionally disrupting network operation, the attack is termed jamming or Denial-of-Service (DoS), depending on whether one looks at its cause or its consequences.

Jamming can disrupt wireless transmission and can occur either unintentionally in the form of interference, noise or collision at the receiver side or in the context of an attack. A jamming attack is particularly effective since (i) no special hardware is needed in order to be launched,
(ii) it can be implemented by simply listening to the open medium and broadcasting in the same frequency band as the network and
(iii) if launched wisely, it can lead to significant benefit with small incurred cost for the attacker

For an adversary agnostic to the implementation details of the network, a typical jamming strategy is the continuous emission of high-power interference signals such as continuous wave tones, or FM modulated noise. However, adopting an "always-on" jamming strategy has several disadvantages. First, the adversary has to expend a significant amount of energy to jam frequency bands of interest. Second, the continuous presence of high interference levels makes this type of jamming easy to detect. Third, these attacks are easy to mitigate either by spread spectrum communications, spatial retreats, or localization and removal of the jamming nodes.

## II. Background Work And Literature Survey

Intelligent attacks which target the transmission of specific packets were presented. Thuente considered an attacker who infers eminent packet transmissions based on timing information at the MAC layer Channel-selective jamming attacks were considered. It was shown that targeting the control channel reduces the required power for performing a DoS attack by several orders of magnitude. To protect control channel traffic, control information was replicated in multiple channels. The "locations" of the channels where control traffic was broadcasted at any given time, was cryptographically protected. We proposed a randomized frequency hopping algorithm, to protect the control channel inside jammers. The jammer controls probability of jamming and transmission range in order to cause maximal damage to the network in terms of corrupted communication links. The jammer action ceases when it is detected by the network, namely by a monitoring node, and a notification message is transferred out of the jamming region. The fundamental tradeoff faced by the attacker is the following: a more aggressive attack in terms of higher jamming probability or larger transmission range increases the instantaneously derived payoff but exposes the attacker to the network and facilitates its detection and later on its isolation. In an effort to withstand the attack and alleviate the attacker benefit, the network adapts channel access probability. In Strong Hiding Commitment Scheme we use DES algorithm to encrypt packets where single secret key is used between sender and receiver. The major disadvantage is, the attacker can easily retrieve the packets based on brute force attacks so we need to provide more security for the packets. In Cryptographic puzzle Hiding Scheme, where each packet is attached with the puzzle and encrypted. We specify the time limit for the solution of the puzzle. If puzzle is not solved within the time limit there may be dropping of packets and also there is a delay in receiving the packets. Selective jamming attacks have been experimentally implemented using software defined radio engines. USRP2-based jamming platform called RF React was implemented by Wilhelm that enables selective and reactive jamming. We develop three schemes that prevent jamming attacks; they are Strong Hiding Commitment Scheme, Cryptographic Puzzle Hiding Scheme and All or Nothing Transformation

## III. Problem Statement And Model Assumptions

Consider the scenario depicted in Figure 1(a). Nodes A and B communicate over the wireless medium and a jamming node J is within communication range of both A and B. Node A transmits a packet m to B which is eavesdropped by node J. Node J is able to classify m by receiving only its first few bytes. J then corrupts m by interfering with its reception at B. We address the problems of

(a) evaluating the ability of the adversary in classifying transmitted messages in real-time, and

(b) developing resource-efficient mechanisms for preventing real-time packet classification.

Network model–Our network consists of a collection of nodes connected via wireless links. Nodes may communicate directly, or over multiple hops. The nodes of the network can establish globally shared keys, either by manual preload, or via an online key distribution center. Communication Model–Communication can be either broadcast or unicast. Packets are transmitted at a rate of R bauds. Each symbol corresponds to q bits according to the underlying digital modulation scheme. Here the transmission bit rate is equal to qR bps. To generalize our analysis, we do not consider any spreading of the data. However, our results hold even if data is spread to a wider spectrum according to any technique such as DSSS or FHSS. Transmitted packets have the generic frame format depicted in Figure 1(b). The preamble is used for synchronizing the sampling process at the receiver. The PHY header contains information regarding the length of the frame and the transmission rate.

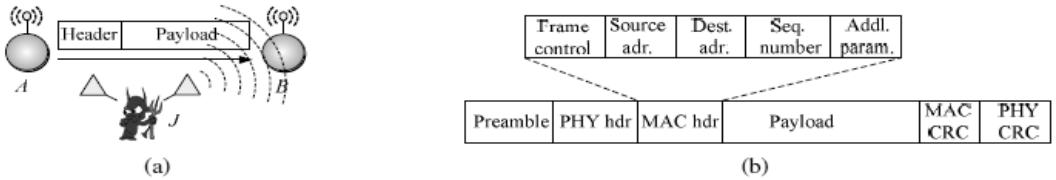



Fig. 1. (a) Realization of a selective jamming attack, (b) a generic frame format for a wireless network

The MAC header contains information relevant to the MAC layer. In particular, the MAC header determines the MAC protocol version, the type of packet (management, control, or data) and its subtype (e.g. association request/response, RTS, CTS, ACK, etc.), the source and destination addresses plus some additional fields regarding power management, security parameters, and information for future transmissions. The MAC header is followed by the frame body that contains higher layer information. Finally, the MAC frame is protected by a CRC code attached in the CRC field. Adversary Model–We assume the adversary is in control of the communication medium and can jam messages at any part of the network of his choosing. The adversary can

operate in full-duplex mode, thus being able to receive and transmit concurrently. This can be achieved, for example, with the use of multiple radios. In addition, the adversary is equipped with directional antennas that enable the reception of a signal from one node and jamming of the same signal at another. The adversary is assumed to be computationally bounded, although he can be significantly more powerful than the network devices. Solving well-known hard cryptographic problems is assumed to be time-consuming.

The network employs a monitoring mechanism for detecting potential malicious activity by a jammer. The monitoring mechanism consists of the following: (i) determination of a subset of nodes M that will act as network monitors, and (ii)

Employment of a detection algorithm at each monitor node. The assignment of the role of monitor to a node can be affected by energy limitations and detection performance specifications. In this work, we fix M and formulate optimization problems for one or more monitor nodes. We now fix attention to detection at one monitor node. First, we define the quantity to be observed at each monitor node. In

Our case, the readily available metric is probability of collision that a monitor node experiences, namely the percentage of packets that are erroneously received. During normal network operation, and in the absence of a jammer, we consider a large enough training period in which the monitor node "learns" the percentage of collisions it experiences as the long-term average of the ratio of number of slots in which there was a collision over total number of slots of the training period. Assume now the network operates in the open after the training period and fix attention to a time window much smaller than the training period. An increased percentage of collisions over this time window compared to the learned long-term average may be an indication of an ongoing jamming attack or only a temporary increase of percentage of collisions compared to the average during normal network operation.

## IV. Conclusion

We illustrated the effectiveness of selective jamming attacks by implementing such attacks against the TCP protocol. We showed that an adversary can exploit its knowledge of the protocol implementation to increase the impact of his attack at a significantly lower energy cost. We illustrated the feasibility of selective jamming attacks by performing real time packet classification.

Therefore, to improve detection, we introduced the notion of consistency checking, where the packet delivery ratio is used to classify a radio link as having poor utility, and then a consistency check is performed to classify whether poor link quality is due to jamming. There exist several directions for future study. Interesting issues arise in multi-channel networks. In that case, the defense strategy space has an additional dimension, channel switching, while the jammer has higher energy costs when jamming more channels. Another interesting issue is to find alternatives for modeling lack of knowledge for the attacker and the network. An idea would be to average over all strategies of the opponent.

## References


[1] IEEE Std 802.11i/d3.0. Available at http://www.cs.umd.edu/ mhshin/doc/802.11/802.11i-D3.0.pdf.

[2] AusCERT. AA-2004.02 - denial of service vulnerability in IEEE 802.11 wireless devices. http://www.auscert.org.

[3] P. Bahl and V. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In Proceedings of IEEE Infocom 2003, pages 775 to784, 2000.

[4] J. Bellardo and S. Savage. 802.11 denial-of-service attacks: Real vulnerabilities and practical solutions. In Proceedings of the USENIX Security Symposium, pages 15 to 28, 2003.

[5] S. Capkun and J. Hubaux. Secure positioning in sensor networks. Technical report EPFL/IC/200444, May 2004

[6] I. Damgard. Commitment schemes and zero-knowledge protocols. Lecture notes in computer science, 1561:63–86, 1999.

[7] A. Juels and J. Brainard. Client puzzles: A cryptographic countermeasure against connection depletion attacks. In Proceedings of the Network and Distributed System Security Symposium, pages 151–165, 1999.

[8] Y. W. Law, M. Palaniswami, L. V. Hoesel, J. Doumen, P. Hartel, and P. Havinga. Energy-efficient link-layer jamming attacks against WSN MAC protocols. ACM Transactions on Sensor Networks, 5(1):1–38, 2009.

[9] L. Lazos, S. Liu, and M. Krunz. Mitigating control-channel jamming attacks in multi-channel ad hoc networks. In Proceedings of the second ACM conference on wireless network security, pages 169– 180, 2009.

[10] R. C. Merkle. Secure communications over insecure channels. Communications of the ACM, 21(4):294–299, 1978.

[11] W. Xu, W. Trappe, Y. Zhang and T. Wood, The feasibility of launching and detecting jamming attacks in wireless networks,Proc. ACM Mobi Hoc, 2005.

[12] W. Xu, T. Wood, W. Trappe and Y. Zhang, Channel surfing and spatial retreats: defenses against wireless denial of service, Proc. Workshop on Wireless Security (WiSe), 2004.

[13] J. M. McCune, E. Shi, A. Perrig and M. K. Reiter, Detection of denial-of-message attacks on sensor network broadcasts ,Proc. IEEE Symposium on Security and Privacy, 2005.

[14] A. Wald, Sequential Analysis, Wiley 1947.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 60-65*
*www.iosrjournals.org*

# Simple Load Rebalancing For Distributed Hash Tables In Cloud

Ch. Mounika[1], L. RamaDevi[2], P.Nikhila[3]
[1]*M.Tech (S.E), VCE, Hyderabad, India,*
[2]*M.Tech (S.E), VCE, Hyderabad, India,*
[3]*M.Tech (S.E), VCE, Hyderabad, India,*

***Abstract:*** *Distributed file systems are key building blocks for cloud computing applications based on the Map Reduce programming paradigm. In such file systems, nodes simultaneously serve computing and storage functions; a file is partitioned into a number of chunks allocated in distinct nodes so that Map Reduce tasks can be performed in parallel over the nodes. However, in a cloud computing environment, failure is the norm, and nodes may be upgraded, replaced, and added in the system. This dependence is clearly inadequate in a large-scale, failure-prone environment because the central load balancer is put under considerable workload that is linearly scaled with the system size, and may thus become the performance bottleneck and the single point of failure. In this paper, a fully distributed load rebalancing algorithm is presented to cope with the load imbalance problem.*

*Our algorithm is compared against a centralized approach in a production system and a competing distributed solution presented in the literature. The simulation results indicate that our proposal is comparable with the existing centralized approach and considerably outperforms the prior distributed algorithm in terms of load imbalance factor, movement cost, and algorithmic overhead.*

***Keywords:*** *DHT, CentraliseSystem, LoadImBalancing, Distributed System*

## I. Introduction

The Distributed file systems an important issue in DHTs is load-balance the even distribution of items to nodes in the DHT. All DHTs make some effort to load balance; generally by randomizing the DHT address associated with each item with a "good enough" hash function and making each DHT node responsible for a balanced portion of the DHT address space. Chord is a prototypical example of this approach: its "random" hashing of nodes to a ring means that each node is responsible for only a small interval of the ring address space, while the random mapping of items means that only a limited number of items land in the (small) ring interval owned by any node. The cloud computing Current distributed hash tables do not evenly partition the address space into.

Which keys get mapped; some machines get a larger portion of it. Thus, even if keys are numerous and random, some machines receive more than their fair share, by as much as a factor of n times the average. To cope with this problem, many DHTs use virtual nodes each real machine pretends to be several distinct machines, each participating independently in the DHT protocol. The machine's load is thus determined by summing over several virtual nodes', creating a tight con cent ration of load near the average. As an example, the Chord DHT is based upon consistent hashing which requires virtual copies to be operated for every node.

The node will occasionally check its inactive virtual nodes, and may migrate to one of them if the distribution of load in the system has changed. Since only one virtual node is active, the real node need not pay the original Chord protocol's multiplicative increase in space and band width costs. Our solution to this problem therefore allows nodes to move to arbitrary addresses; with this freedom we show that we can load balance an arbitrary distribution of items, without expending much cost in maintaining the load balance. Our scheme works through a kind of "work stealing" in which under loaded nodes migrate to portions of the address space occupied by too many items. The protocol is simple and practical, with all the complexity in its performance analysis. In this paper, we are interested in studying the load rebalancing problem in distributed file systems specialized for large-scale, dynamic and data-intensive clouds. Lastly, permitting nodes to choose arbitrary addresses in our item balancing protocol makes it easier for malicious nodes to disrupt the operation of the P2P network. It would be interesting to find counter-measures for this problem.

The paper is organized as follows. Section II Related work, Section III. System Model Section IV Load balancing algorithm. Section V Distributed files system. Section VI. . Performance Evaluation VII concludes.

## II. Related Work

This attempt to load-balance can fail in two ways. First, the typical "random" partition of the address space among nodes is not completely balanced. Some nodes end up with a larger portion of the addresses and thus receive a larger portion of the randomly distributed items. Second, some applications may preclude the

randomization of data items' addresses. For example, to support range searching in a database application the items may need to be placed in a specific order, or even at specific addresses, on the ring. In such cases, we may find the items unevenly distributed in address space, meaning that balancing the address space among nodes is not adequate to balance the distribution of items among nodes. We give protocols to solve both of the load balancing challenges just described.

**Performance in a P2P System:**

Our online load balancing algorithms are motivated by a new application domain for range partitioning peer-to-peer systems. P2P systems store a relation over a large and dynamic set of nodes, and support queries over this relation. Many current systems, known as Distributed Hash Tables (DHTs) use hash partitioning to ensure storage balance, and support point queries over the relation. There has been considerable recent interest in developing P2P systems that can support efficient range queries. For example, a P2P multi-player game might query for all objects located in an area in a virtual 2-D space. In a P2P web cache, a node may request (pre-fetch) all pages with a specific URL prefix. It is well-known that hash partitioning is inefficient for answering such ad hoc range queries, motivating a search for new networks that allow range partitioning while still maintaining the storage balance offered by normal DHTs.

**Handling Dynamism in the Network:**

The network is a splits the range of Nh to take over half the load of Nh, using the NBRADJUST operation. After this split, there may be NBRBALANCE violations between two pairs of neighbors and In response, ADJUSTLOAD is executed, first at node Nh and then at node N. It is easy to show (as in Lemma 3) that the resulting sequence of NBRADJUST operations repair all NBRBALANCE violations.

**Node Departure:**

While in the network, each node manages data for a particular range. When the node departs, the data is stored becomes unavailable to the rest of the peers. P2P networks reconcile this data loss in two ways: (a) Do nothing and let the "owners" of the data deal with its availability. The owners will frequently poll the data to detect its loss and re-insert the data into the network.

Maintain replicas of each range across multiple nodes. A Skip Net DHT organizes peers and data objects according to their lexicographic addresses in the form of a variant of a probabilistic skip list. It supports logarithmic time range-based lookups and guarantees path locality. Mercury is more general than Skip Net since it supports range-based lookups on multiple-attributes. Our use of random sampling to estimate query selectivity constitutes a novel contribution towards implementing scalable multi-dimensional range queries. Load balancing is another important way in which Mercury from Skip Net. While Skip Net incorporates a constrained load-balancing mechanism, it is only useful when part of a data name is hashed, in which case the part is inaccessible for performing a range query. This implies that Skip Net supports load-balancing or range queries not both.

## III. System Model

### 3.1 Data Popularity:

Unfortunately, in many applications, a particular range of values may exhibit a much greater popularity in terms of database insertions or queries than other ranges. This would cause the node responsible for the popular range to become overloaded. One obvious solution is to determine some way to partition the ranges in proportion to their popularity. As load patterns change, the system should also move nodes around as needed.

We leverage our approximate histograms to help implement load-balancing in Mercury. First, each node can use histograms to determine the average load existing in the system, and, hence, can determine if it is relatively heavily or lightly loaded. Second, the histograms contain information.

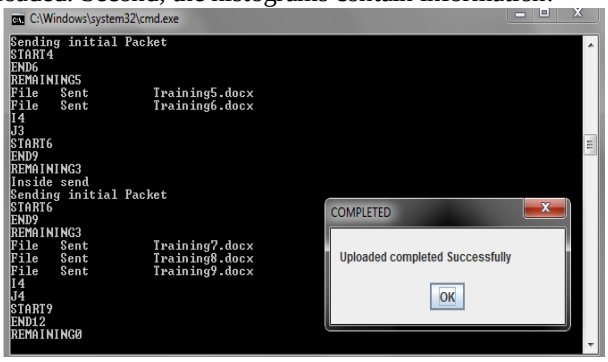



Fig.1 About which parts of the overlay are lightly loaded.

**3.2 Load Balancing:**

We have shown how to balance the address space, but sometimes this is not enough. Some applications, such as those aiming to support range-searching operations, need to specify a particular, non-random mapping of items into the address space. In this section, we consider a dynamic protocol that aims to balance load for arbitrary item distributions. To do so, we must sacrifice the previous protocol's restriction of each node to a small number of virtual node locations instead, each node is free to migrate anywhere. Our protocol is randomized, and relies on the underlying P2P routing framework only insofar as it has to be able to contact "random" nodes in the system (in the full paper we show that this can be done even when the node distribution is skewed by the load balancing protocol). The protocol is the following, to state the performance of the protocol, we need the concept of a half-life [LNBK02], which is the time it takes for half the nodes or half the items in the system to arrive or depart.

**3.3 DHT Implementation**

The storage nodes are structured as a network based on distributed hash tables (DHTs), e.g., discovering a file chunk can simply refer to rapid key lookup in DHTs, given that a unique handle (or identifier) is assigned to each file chunk. DHTs enable nodes to self-organize and Repair while constantly offering lookup functionality in node dynamism, simplifying the system provision and management. The chunk servers in our proposal are organized as a DHT network. Typical DHTs guarantee that if a node leaves, then its locally hosted chunks are reliably migrated to its successor; if a node joins, then it allocates the chunks whose IDs immediately precede the joining node from its successor to manage. Now we describe the application of this idea to DHTs. Let h0 be a universally agreed hash function that maps peers onto the ring. Similarly, let h1; h2; : : :hd be a series of universally agreed hash functions mapping items onto the ring. To insert an item x using d hash functions, a peer calculates h1(x); h2(x); : : : ; hd(x). Then, d lookups are executed in parallel to and the peers p1; p2; : : : ; pd responsible for these hash values, according to the mapping given by h0,

**3.4 Chunk creation:**

A file is partitioned into a number of chunks allocated in distinct nodes so that Map Reduce Tasks can be performed in parallel over the nodes. The load of a node is typically proportional to the number of file chunks the node possesses. Because the files in a cloud can be arbitrarily created, deleted, and appended, and nodes can be upgraded, replaced and added in the file system, the file chunks are not distributed as uniformly as possible among the nodes. Our objective is to allocate the chunks of files as uniformly as possible among the nodes such that no node manages an excessive number of chunks.


Fig.2 Chunk creation

**3.5 Replica Management:**

In distributed file systems (e.g., Google GFS and Hadoop HDFS), a constant number of replicas for each file chunk are maintained in distinct nodes to improve file availability with respect to node failures and departures. Our current load balancing algorithm does not treat replicas distinctly. It is unlikely that two or more replicas are placed in an identical node because of the random nature of our load rebalancing algorithm. More specifically, each under loaded node samples a number of nodes, each selected with a probability of 1/n, to share their loads (where n is the total number of storage nodes).

## IV. Load Balancing Algorithm

In our proposed algorithm, each chunk server node I first estimate whether it is under loaded (light) or overloaded (heavy) without global knowledge. A node is light if the number of chunks it hosts is smaller than the threshold. Load statuses of a sample of randomly selected nodes.

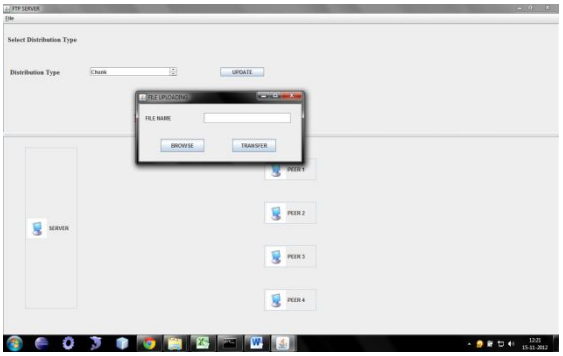



Fig.3 Load Balancing

Specifically, each node contacts a number of randomly selected nodes in the system and builds a vector denoted by V. A vector consists of entries, and each entry contains the ID, network address and load status of a randomly selected node. Fig. 3 shows the total number of messages generated by a load rebalancing algorithm.

**Load-balanced state:**
If each chunk server hosts no more than Am chunks. In our proposed algorithm, each chunk server node I first estimates whether it is under loaded (light) or overloaded (heavy) without global knowledge. Lj A from j to relieve j's load. Node j may still remain as the heaviest node in the system after it has migrated its load to node i. In this case, the current least-loaded node, say node i departs and then rejoins the system as j's successor. That is, I become node j+1, and j's original successor i thus becomes node j + 2. Such a process repeats iteratively until j is no longer the heaviest. Then, the same process is executed to release the extra load on the next heaviest node in the system. This process repeats until all the heavy nodes in the system become light nodes.

**Others:** We will offer a rigorous performance analysis for the effect of varying nV in Appendix E. Specifically, we discuss the tradeoff between the value of nV and the movement cost. A larger nV introduces more overhead for message exchanges, but results in a smaller movement cost.

**Procedure 1** ADJUSTLOAD (Node Ni)

On Tuple Insertg
1 : *Let* $L(Ni) = x\ 2\ (Tm;\ Tm+1)$
2 : *Let Nj be the lighter loaded of Ni* ?1 *and Ni* +1.
3 : **if***L* $(Nj)$ _ *Tm* ?1**then***fDo NBRADJUSTg*
4: Move tuples from Ni to Nj to equalize load.
5: ADJUSTLOAD(Nj)
6: ADJUSTLOAD(Ni)
7: **else**
8: Find the least-loaded node Nk.
9: **if***L* $(Nk)$ _ *Tm* ? 2**then***fDo REORDERg*
10: Transfer all data from Nk to N = Nk_1.
11: Transfer data from Ni to Nk, s.t. L(Ni) = dx=2e and L(Nk) = bx=2c.
12: ADJUSTLOAD (N)
13: fRename nodes appropriately after REORDER.g
14: **end if**
15: **end if**

**Example1:** In the setting above, the maximum load is at most log log n= log d+O with high probability. Our proof (not included for reasons of space) uses the layered induction technique from the seminal work of Because of the variance in the arc length associated with each peer; we must modify the proof to take this into account. The standard layered induction uses the fact that if there is k bins that have load at least k,

**Examole2:** long distance links are constructed using the harmonic distribution on node-link distance. Value Link denotes the overlay when the harmonic distribution on value distance. Given the capacities of nodes (denoted by {β1, β2, · · · , βn}), we enhance the basic algorithm in Section III-B2 as follows: each node i approximates the ideal number of file chunks that it needs to host in a load balanced state as follows:

$$A_i = \gamma\beta_i,$$

Note that the performance of the Value Link overlay is representative of the performance of a plain DHT under the absence of hashing and in the presence of load balancing algorithms which preserve value contiguity.

```
map(String key, String value):
// key: document name
// value: document contents
for each word w in value:
EmitIntermediate(w, "1");
reduce(String key, Iterator values):

// key: a word
// values: a list of counts
int result = 0;
for each v in values:
result += ParseInt(v);
Emit(AsString(result));
```

## V. Distributed File System

We have given several provably efficient loadbalancing for distributed file's protocols for distributed data storage in P2P systems. More details and analysis can be found in a thesis. Our algorithms are simple, and easy to implement in. distributed files so an obvious next research step should be a practical evaluation of these schemes. In addition, several concrete open problems follow from our work.

First, it might be possible to further improve the consistent hashing scheme as discussed at the end of our range search data structure. Distributed does not easily generalize to more than one order. For example (Fig.4) when storing music files, one might want to index them by both artist and song title, allowing lookups according to two orderings. Since our protocol rearranges the items according to the ordering, doing this for two orderings at the same time seems difficult. A simple, but inelegant, solution is to rearrange not the items themselves, but just store pointers to them on the nodes. This requires far less storage, and Network Setting.

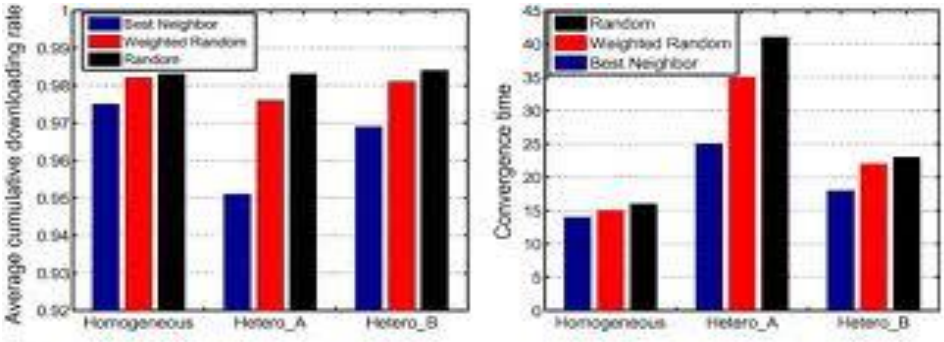



Fig.4 The average downloading rate and Convergence time

Makes it possible to maintain two or more orderings at once. Lastly, emitting nodes to choose arbitrary addresses in our item balancing protocol for distributed file's makes it easier for malicious nodes to disrupt the operation of the P2P network. It would be interesting to find counter-measures for this problem.

## VI. Performance Evaluation

We run a varying number of players. The players move through the world according to a random waypoint model, with a motion time chosen uniformly at random from seconds, a destination chosen uniformly at random, and a speed chosen uniformly at random from (0, 360) pixels per second. The size of the game world is scaled according to the number of players. The dimensions are 640n _ 480n, where n is the number of players. All results are based on the average of 3 Experiments, with each experiment lasting 60 seconds. The

experiments include the bent of log n sized LRU cache long pointers. The HDFS load balancer and our proposal. Our proposal clearly outperforms the HDFS load balancer. When the name node is heavily loaded (i.e., small M's), our proposal remarkably performs better than the

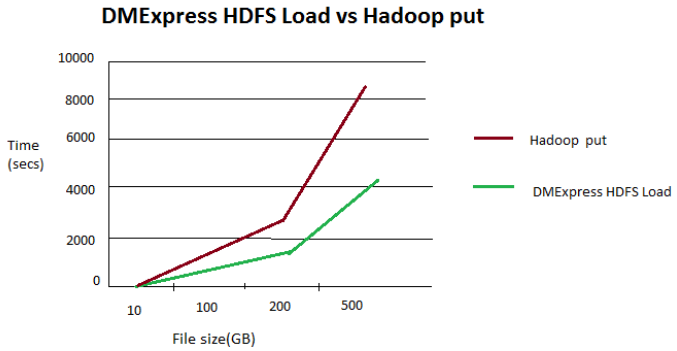



Fig.5 HDFS

HDFS load balancer. For example, if M = 1%, the HDFS load balancer takes approximately 60 minutes to balance the loads of data nodes. By contrast, our proposal takes nearly 20 minutes in the case ofM= 1%. Specifically, unlike the HDFS load balancer, our proposal is independent of the load of the name node. In particular, approximating the unlimited scenario is expensive, and the use of blog2 nc virtual peers as proposed in introduces a large amount of topology maintenance track but does not provide a very close approximation. Finally, we observe that while we are illustrating the most powerful instantiation of virtual peers, we are comparing it to the weakest choice model further improvements are available to us just by increasing d to 4.

## VII. Conclusions

A load balancing algorithm to deal with the load rebalancing problem in large-scale, dynamic, and distributed file systems in clouds has been presented in this paper. Our proposal work is to balance the loads of nodes and reduce the demanded movement cost as much as possible, while taking advantage of physical network locality and node heterogeneity. In the absence of representative real workloads (i.e., the distributions of file chunks in a large-scale storage system) in the public domain, we have investigated the performance of our proposal and compared it against competing algorithms through synthesized probabilistic distributions of file chunks. The synthesis workloads stress test the load balancing algorithms by creating a few storage nodes that are heavily loaded. The computer simulation results are encouraging, indicating that our proposed algorithm performs very well.

## Reference

[1] I. Stoica, R. Morris, D. Liben-Nowell, D. R. Karger, M. F. Kaashoek, F. Dabek, and H. Balakrishnan, "Chord: a Scalable Peer-to-Peer Lookup Protocol for Internet Applications," IEEE/ACM Trans. Netw., vol. 11, no. 1, (Feb. 2003), 17–21.

[2] A. Rowstron and P. Druschel, "Pastry: Scalable, Distributed Object Location and Routing for Large-Scale Peer-to-Peer Systems," LNCS 2218, (Nov. 2001), 161–172.

[3] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels, "Dynamo: Amazon's Highly Available Key-value Store," in Proc. 21st ACM Symp. Operating Systems Principles (SOSP'07), (Oct. 2007), 205–220.

[4] A. Rao, K. Lakshminarayanan, S. Surana, R. Karp, and I. Stoica, "Load Balancing in Structured P2P Systems," in Proc. 2nd Int'l Workshop Peerto-Peer Systems (IPTPS'02), (Feb. 2003), 68–79.

[5] D. Karger and M. Ruhl, "Simple Efficient Load Balancing Algorithms for Peer-to-Peer Systems," in Proc. 16th ACM Symp. Parallel Algorithms and Architectures (SPAA'04), (June 2004), 36–43.

[6] D. DeWitt, R. H. Gerber, G. Graefe, M. L. Heytens, K. B. Kumar, and M. Muralikrishna. Gamma -a high performance dataflow database. In Proc. VLDB, 1986.

[7] H. Feelifl, M. Kitsuregawa, and B. C. Ooi. A fast convergence technique for online heat-balancing of btree indexed database over shared-nothing parallel systems. In Proc. DEXA, 2000.

[8] P. Ganesan, M. Bawa, and H. Garcia-Molina. Online balancing of range-partitioned data with applications to p2p systems. Technical Report http://dbpubs.stanford.edu/pubs/2004-18, Stanford U., 2004.

[9] P. Ganesan, B. Yang, and H. Garcia-Molina. One torus to rule them all: Multi-dimensional queries in p2p systems. In WebDB, 2004.

[10] D. R. Karger and M. Ruhl. Simple efficient load-balancing algorithms for peer-to-peer systems. In Proc. IPTPS, 2004.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 66-73*
*www.iosrjournals.org*

# Detection of Session Hijacking and IP Spoofing Using Sensor Nodes and Cryptography

Abhishek Kumar Bharti [1], Manoj Chaudhary[2]
[1]*Research Scholar, Computer Science, YCoE , Punjabi University Patiala, Punjab, India*
[2]*Assistant Professor, Computer Science, YCoE , Punjabi University Patiala, Punjab, India*

***Abstract:*** *Many web applications available today make use of some way of session to be able to communicate between the server and client. Unfortunately, it is possible for an attacker to exploit session in order to impersonate another user at a web application. The session hijacking is the most common type of attack in the infrastructure type of network. The confidentially is not providing under this attack to user information. Session hijacking attack is launched by making fake access point. If we detect the fake access point then we can stop session hijacking, and various techniques had been proposed. In this paper, we are giving a new mechanism to detect the fake access point with the use of sensor nodes in the network. In this mechanism we are also giving the protection against IP Spoofing by the use of public private key cryptography key exchange algorithm. We also discuss the results through simulations in Network Simulator 2.*

***Keywords:*** *Session Hijacking, Fake Access Point, IP spoofing.*

## I. Introduction

Wireless process control has been a popular topic recently in the field of industrial control. Compared to traditional wired process control systems, their wireless counterparts have the potential to save costs and make installation easier [12]. Wireless technologies range from complex systems, such as Wireless Local Area Networks (WLAN) and cell phones to simple devices such as wireless headphones, microphones, and other devices that do not process or store information. They also include infrared (IR) devices these use the direct line of sight between transmitter and receiver. As the wireless technology is growing very fast. So security in wireless networks have some additional challenges compared to wired networks. This is due to the fact that the traffic is transferred as radio waves in the air and anyone close enough with an antenna can receive them. There are various types of attacks are possible in wireless networks. The most common attacks are man-in-middle attack, denial-of-service attack.The man-in-the-middle (or middleperson) attack is one in which legitimate parties communicate via a hostile adversary but without their knowledge or consent. This attack can be devastatingly effective because the adversary enjoys complete control of the communication link and can inspect, inject, delay, delete, modify and re-order traffic to suit their purpose. It may be used, for example, to bypass weak authentication protocols, hijack legitimate sessions, perform active traffic analysis and deny service[14]. Session Hijacking is one of the popular attack in man-in-middle attack. In this paper we give a mechanism to prevent session hijacking attack. Its one of the favorite attack for the attackers because of the nature of the attack. A user who is trying to login or already logged in to a server , the attacker takes control over a session, basically hijacks the session from the user and continues the connection to the server pretending to be the user. Session hijacking have a great advantage to the attackers they don't have to waste hours and hours to crack the password , since the user has already been authenticated and in a active session it makes is so much easier to just listen to the traffic on the network without the knowledge of the user. There are three different types of session hijack attacks:

- Active Session Hijacking
- Passive Session Hijacking
- Hybrid Session Hijacking

*Active Session Hijacking:* The active attack is when the attacker hijacks a session on the network. The attacker will silence one of the machines, usually the client computer, and take over the clients' position in the communication exchange between the workstation and the server. And drop the connection between the user and the server. There are various methods for dropping the connection to the server, one of the most common is to send the huge amount of traffic, and this type of attack is known as Denial of Service. By doing this attacker has full control over the session and it communicate with the server pretending that it is the authenticated user fig1 shows how a typical session hijacking is conducted between a client and a server by an attacker. actual situation of the active session hijacking.

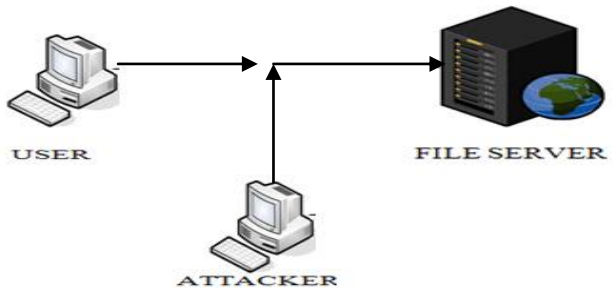



***Figure 1:*** *Active Session Hijacking*

*Passive Session Hijacking :* Passive session hijack attacks are similar to the active attack, but rather than removing the user from the communication session, the attacker monitors the traffic between the workstation and server . In a passive session the attacker listens to all the data and captures them for future attacks, in most cases to perform any type of a hijacking attack it is important that the attacker starts off with passive mode. Figure 2 shows a typical passive hijacking.

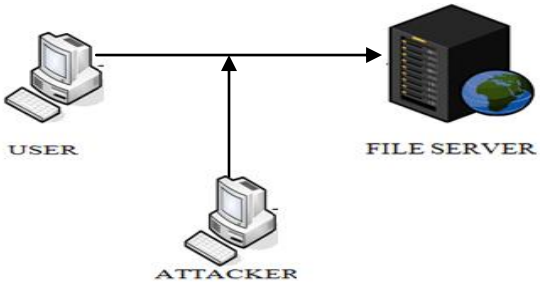



***Figure 2:*** *Passive Session Hijacking*

*Hybrid Session Hijacking:* This attack is a combination of the active and passive attacks, which allow the attacker to listen to network traffic until something of interest is found. The attacker can then modify the attack by removing the workstation computer from the session, and assuming their identity.

In the infrastructure-based network, communication typically takes place only between the wireless nodes and the access point, but not directly between the wireless nodes. The access point does not just control medium access, but also acts as a bridge to other to wireless or wired networks. One of the ways in which a WLAN can be attacked is by introducing one or more unauthorized fake Access Points (APs) in the network. A fake AP can be set up by a malicious attacker. This fake AP is used to fool a wireless node in the WLAN into accessing the network through the fake AP instead of the authorized one. The fake AP can then launch a variety of attacks thereby comprising the security of the wireless communication. Setting up fake APs is not hard. And these Access points that are installed without proper authorization and verification that overall security policy is not obeyed by an AP is called rogue APs. These AP'S are installed and used by invalid users. Such APs are configured poorly, and it will used by an attackers.[11]

After the attacker set up a fake AP , it uses the packet sniffer software to sniff the packets . A (packet) sniffer is a program that intercepts and decodes network traffic broadcast through a medium. Sniffing is the act by a machine S of making copies of a network packet sent by machine A intended to be received by machine B. Sniffer is a program that is used for monitors and analyzes network traffic, detecting bottlenecks and problems. A network manager can keep traffic flowing efficiently by using this information. A sniffer can also be used legitimately or illegitimately to capture data being transmitted on a network. Network router reads every packet of data that passed from it, to determining whether it is delivered for a destination within the router's own network or whether it should be passed further along the Internet. A route that uses a sniffer it may be able to read the data in the packet as well as the source and destination addresses. In file sharing applications, sniffers are used on academic networks to prevent traffic bottlenecks. There are many techniques are given to detect fake access point like using clock skew [13].

In this paper a novel approach is given to detect the fake access point by the use of the sensor nodes. Even if we are using the sensor nodes to detect the fake access point if the attacker spoof the IP and MAC addresses of the client and pretend to be the valid authenticated user but it does not exist or may be that values belong to others. So in this novel approach a protection against the IP spoofing by public private key cryptographic key exchange algorithm.

## II. Background And Related Work

Session hijacking is the stealing of the session of the user that is use to communicate with the server. In session hijacking a user who is already logged in (authenticated) to a web server and has a valid session existing between the user and the server, the attacker takes control over such a session, basically hijack the session from the user and continues the connection to the server pretending to be the user. And IP spoofing is the creation of IP packets using somebody else's IP source addresses. So in order to detect the session hijacking and IP spoofing various algorithms has been develop Extant defensive techniques and procedures are not completely effective against such attacks.

### A. Technique for preventing Session Hijacking :

**A.1** Rupinder Gill, and Smith, Jason and Clark, Andrew proposed an algorithm in 2006 that is based on IEEE 802.11 Network to detect the Session Hijacking distinct test scenarios. A correlation engine has also been introduced to maintain the false positives and false negatives at a manageable level. We also explore the process of selecting optimum thresholds for both detection techniques. This paper extends earlier work and explores usability, robustness and accuracy of these intrusion detection techniques by applying them to eight distinct test scenarios. A correlation engine has also been introduced to maintain the false positives and false negatives at a manageable level. We also explore the process of selecting optimum thresholds for both detection techniques. For the purposes of our experiments, Snort-Wireless open source wireless intrusion detection system was extended to implement these new techniques and the correlation engine. Absence of any false negatives and low number of false positives in all eight test scenarios successfully demonstrated the effectiveness of the correlation engine and the accuracy of the detection techniques[3].

**A.2** Thawatchai Chomsiri in year 2008 made a comparative study between the security of Hotmail, Gmail and Yahoo Mail by Using Session Hijacking Hacking Test These three Web Mails were hacked by means of Session Hijacking. The researcher conducted this experiment on the LAN system and used information capturing technique to gain Cookies and Session ID inside Cookies. Then, Hijacking was conducted by using two Hijacking methods. The first method, which was common and easy to conduct, used only one Cookie. The second method, which was not very popular but offered high penetrating power, used all Cookies (Cookies cloned by SideJacking tools). The results show that the Web Mail with the highest security level is Yahoo Mail; the second one is Hotmail; and the Web Mail with the lowest security level is Gmail[1].

**A.3** Bhavna C.K. Nathani Erwin Adi in year 2012 gives a procedure to identify the Website Vulnerability to Session Fixation Attacks Session fixation is a vulnerability of web applications where a malicious attacker gains full control of a victim's web account without having to use the victim's credentials such as username and password. The authors found that some 48% of Indonesian websites are vulnerable to such attacks because, contrary to best software engineering practices, many use default session management IDs generated by their development platforms. And tells that each website with a URL and/or cookie containing a string of characters that can represent a session ID was classified as vulnerable[4].

**A.4** In year 2012 Srikanth Kamuni, S.ShreehaTejaswini, Bhaskar.J proposed a technique for detecting the session hijacking based on wavelet based real time session hijack detection based on Bluetooth signal analysis in which it tells A session hijacking is predominantly a man in the middle attack for wireless network where the attacker places itself in the route between the source and the destination node such that all the traffic is routed through the malicious node. A session hijacking is the result of accidental association attack or MAC spoofing attack. In this case the attacker places itself close to either the source or the destination or any other router node in such a way that it is considered as a legitimate router node. Therefore detection rate in most of the existing technique is low and unreliable. In this work we propose a real time mechanism for detecting the session hijacking attack by analyzing the signals received from the nodes through a monitoring station in the wavelet domain[2].

### *B. IP spoofing:*

**B.1** Al-Sammarraie Hosam ,Adli Mustafa and Shakeel Ahmad in year 2009 proposed an algorithm for IP spoofing over online environment IP and email spoofing gained much importance for security concerns due to the current changes in manipulating the system performance in different online environments. Intrusion Detection System (IDS) has been used to secure these environments for sharing their data over network and host based IDS approaches[5].

**B.2** Vimal Upadhyay , Rajeev kumar proposed a technique to prevent IP spoofing using hashing encryption in year 2011 this paper gives a better technique for the protection against IP spoofing and this paper tells that the attack on any node or system can be done from any direction[7].

**B.3** Noureldien A. Noureldien, Mashair O. Hussein in year 2012 proposed A method for Detecting and Preventing All Types of Spoofed Source IP Packets and SYN Flooding Packets at Source this method is

based on a network authentication server (AS), which performs an authentication process on SYN packets. The authentication process verifies the legitimacy of SYN packet's source IP address that initiate a connection request from a network subnet host to an external host. During the authentication process of SYN packets, AS identifies and blocks SYN packets with legal source IP address that chip in a TCP/SYN flooding attack. AS preserves network performance by exchanging authentication messages in plain text, and acts as a stateful inspection firewall and only SYN packets are subject for inspection. Our method which is capable to detect and prevent all types of spoofing packets including subnet spoofing contributes to standard ingress/egress methods in eliminating bogus traffic on the Internet[8].

**B.4** In year 2012 an ant-based traceback is proposed to detect the IP spoofing. The proposed traceback approach uses flow level information to identify the spoofing request. To validate the detection method further, this paper considers the number of hop needs to reach the destination end. Using a mapping between IP addresses and their flow level with hop-counts, the server can distinguish spoofed IP packets from legitimate ones. The simulations results show that this approach discards almost 90% of spoofed IP request[9].

### *C. Fake access point:*

**C.1** Kiruthiga.S and Yuvarani.G in year 2012 proposed a technique for the fast and accurate detection of fake access point using non-cryptomethod in which they calculate the clock skew of an AP from the IEEE 802.11 Time Synchronization Function (TSF) time stamps sent out in the beacon/probe response frames. They uses two different methods for this purpose—one based on linear programming and the other based on least-square fit. They supplement these methods with a heuristic for differentiating original packets from those sent by the fake APs. They collect TSF time stamp data from several APs in three different residential settings. Using their measurement data as well as data obtained from a large Setting, they find that clock skews remain consistent over time for the same AP but vary significantly across APs..

**C.2** Hemanshu Kamboj, Gurpreet Singh in year 2012 they also present the detection of fake access point using the sensors in the network but the use the beacon frames received in fix time . Fake access point is the honey. In the session hijacking attack we attract legitimate user to connect with the unencrypted access point .When the legitimate user connect with the access point, we hack the cookies, sessions of the legitimate user. In this paper, they are proposed a hybrid technique to detect fake access point. Their proposed technique is based on the number beacon frames received in fixed time according to the climate conditions[10].

To summarizing the background work and more work, technique and results we are presenting in table.

| Author(s) | Year | Paper Name | Technique | Result |
|---|---|---|---|---|
| *A. Session Hijacking* | | | | |
| Rupinder Gill, and Smith, Jason and Clark, Andrew[3] | 2006 | Experiences in Passively Detecting Session Hijacking Attacks in IEEE 802.11 Networks | Using sensor | Sensors cant do work in co-operation |
| Thawatchai Chomsiri[1] | 2008 | A Comparative Study of Security Level of Hotmail, Gmail and Yahoo Mail by Using Session Hijacking Hacking Test | | Yahoo has maximum security and gmail has the lowest security |
| Bhavna C.K. Nathani Erwin Adi[4] | 2012 | Website Vulnerability to Session Fixation Attacks | Checks the url for checking the vnlnerability | |
| Srikanth Kamuni, S.ShreehaTejaswini, Bhaskar.J, Dr. G.Manjunath[2] | 2012 | Wavelet Based Real Time Session Hijack Detection Based On Bluetooth Signal Analysis | Monitering System using bluetooth | Efficiency increased to 90 % And its real implemented and can be applied to wifi and other networks. |
| *B. IP spoofing* | | | | |
| Al-Sammarraie Hosam and Adli Mustafa[5] | 2009 | Exception Agent Detection System for IP Spoofing Over Online Environments | Create an intrusion detection system for ip spoofing | More efficient but costly. |
| Vimal Upadhyay , Rajeev kumar[7] | 2011 | Detection and preventing IP spoofing attack by Hashed Encryption | Hashing | Attacks in wireless node can be from any direction |
| Noureldien A. Noureldien, Mashair O. Hussein[8] | 2012 | Block Spoofed Packets at Source (BSPS): A method for Detecting and Preventing All Types of Spoofed Source IP Packets and SYN Flooding Packets at Source | Network authentication server | By network ad-ministrators and ISP's to alleviate bogus traffic in the Internet. |

| | | | | |
|---|---|---|---|---|
| N.Arumugam, Dr.C.Venkatesh[9] | 2012 | A Dynamic Method to Detect IP Spoofing on Data Network Using Ant Algorithm | Ant-based traceback | Ensure good filtering of packets |
| Mrs. Mridu Sahu and Rainey C. Lal [6] | 2012 | Controlling IP spoofing through packet filtering | Packet filtering | Preventing spoofing through Packet filtering |
| *C. Fake Access Point Detection Techniques* | | | | |
| Hemanshu Kamboj, Gurpreet Singh[10] | 2012 | Detection of Fake Access Point to Prevent Session Hijacking | Sensor nodes | Can detect the fake access point but not after IP spoofing |
| Kiruthiga.S and Yuvarani.G[11] | 2012 | Fast and Accurate Detection of Fake Access Points Using Non-crypto Method in WLAN | Clock Skew of Access Point | Can not find the MAC spoofing |

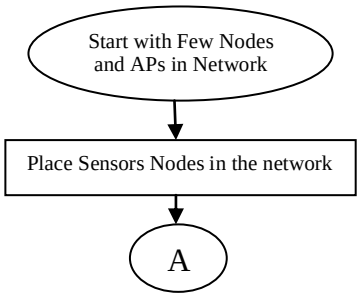
## III. Problem Formulation

A typical session hijacking is a well-known man-in-the-middle attack in the world of network security and its one of the favorite attack for the attackers because of the nature of the attack. The session hijacking attack is generally implemented with the honey pot. The rogue access point acts like a honey pot. Like legitimate AP, fake AP broadcast the beacons signals. A user checks the beacon signals and make probe request to fake AP. Fake access point respond back with probe response. When client and access point are successfully mutually authenticated, connection between access point and client will be established. When client and access point are associated, client can able to access service provided by the access point. A problem arises when legitimate client connects to the fake access point. All the information send by the client can be spoofed by the attacker which operated the fake access point. This leads to the session hijacking. So there have to need some security mechanisms that detect the fake AP and prevent the session hijacking. If detect the fake AP then can prevent the session hijacking. Even if fake access point detection is possible then by IP spoofing the fake access point can't be detected so we need a mechanism that detect the fake access point as well as give the protection against IP spoofing.

## IV. Proposed Technique

To prevent session hijacking, a novel technique is proposed under which, sensor nodes are placed in the network. Sensor nodes sense all the communication between AP and wireless devices i.e. probe request and probe responses. All sensor nodes stores all the information about AP in their database, such as MAC address, coverage area, attached devices, etc. When client wants to communicate with fake access point, client send probe request to access point and access point reply with probe response message. Sensor nodes sensing this communication and they check detail of AP in their database. If the detail of access point doesn't exist in the database of Sensor node, it sends an alarm message to the wireless device about fake AP. In this way a device is protected against fake AP.

In this technique the MAC address spoofing is possible so to prevent against the MAC spoofing new enhancement in the proposed mythology has been proposed. In this technique the public and private key cryptography is used to prevent MAC spoofing. At the start when the sensor nodes storing the knowledge about each node like MAC address, coverage area, attached devices, etc in the network at that time it assign a unique key to each access point in the network to prove its identity. And stores the unique key in their databases corresponding to each AP configuration that have saved. When a access point starts communication it send the unique key encrypt with the sensor node's public key to the sensor node. Sensor node decrypt with its private key and verifies from their database that the key is assign to this MAC address if yes positive acknowledgement will send to the access point if not matched it will send the alarm message to the wireless device about the fake access point. This message exchange is encrypted and decrypted which the asymmetric encryption algorithm. Below flowchart of the proposed technique is given



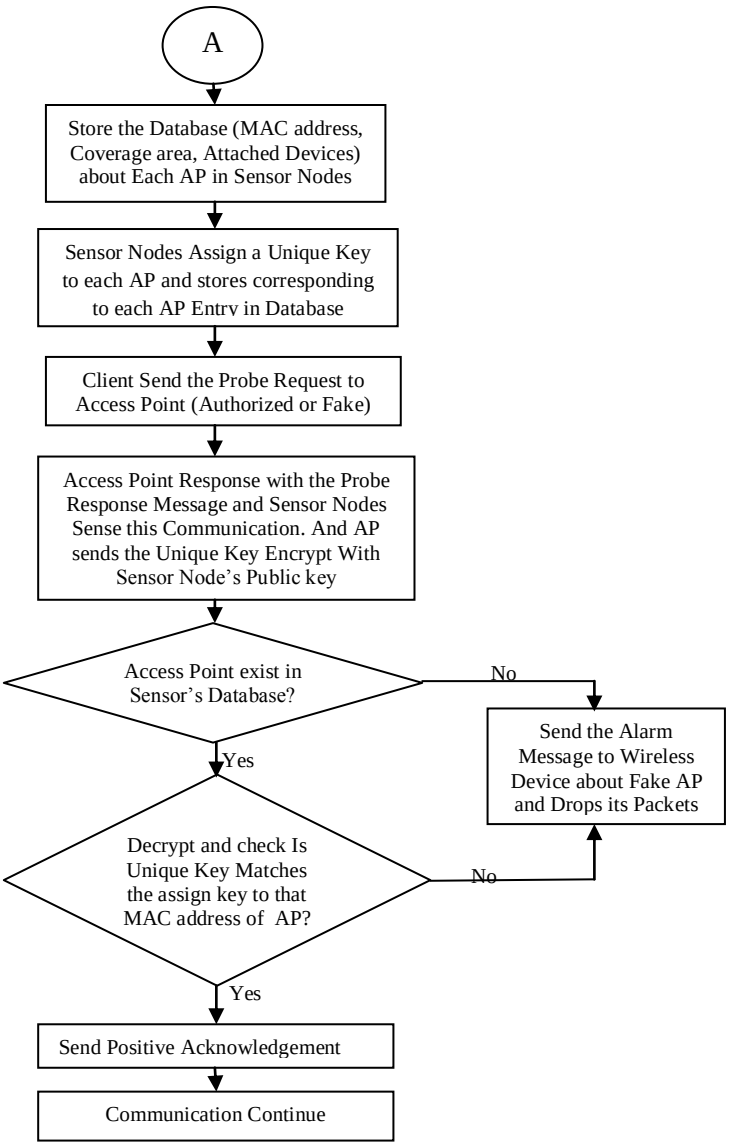



***Figure 3:*** *Flow Chart of Proposed Technique*

## V. Result

For the results we are trying to show the proposed technique to simulate through the Network Simulator 2. For this firstly make a network with few wireless nodes and access points. Fake access point is made and it start providing the services to the legitimate users figure 4 show this so that attacker can easily attack on the session of the user

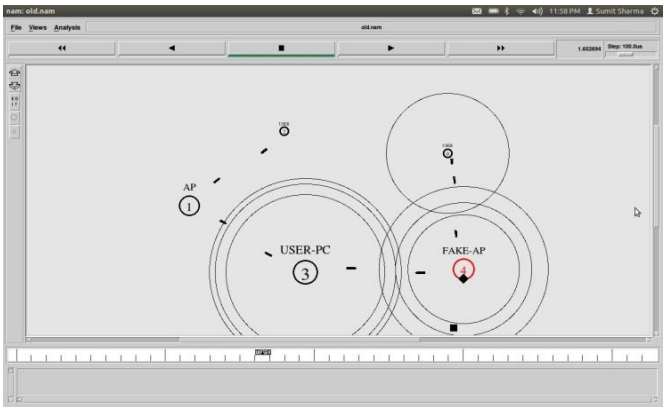



***Figure 4:*** *User accessing the services using fake AP*

Now placing some sensor node in between the network , the sensor nodes store the information about the access points in its data base. The stored information contains the MAC addresses of the access points. The users are declared which wants to communicate with the access point. And sensor nodes also give the unique key to each AP and stores into the database corresponding to each AP. Figure 5 show this

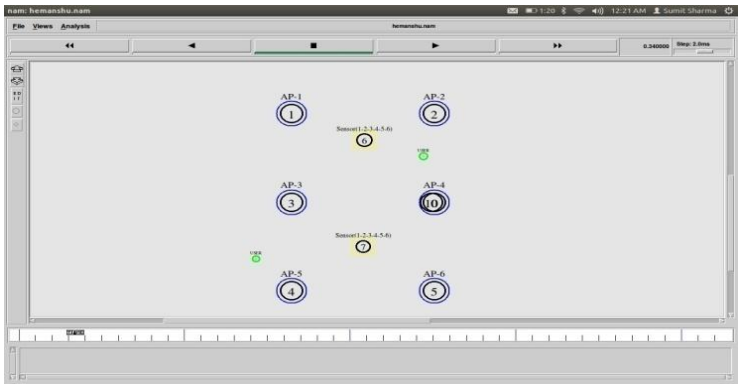



***Figure 5*** *Database maintain by sensor nodes and assign the unique key to each AP*

In figure 6, When the legitimate client send probe request message and access point reply with probe response message. It also send the unique key encrypt with the sensors public key that is near to it .The message exchange between client and access point is sensed by the sensor node. The sensor node decrypt the message that received from the Access Point. And verifies that it is same or different.

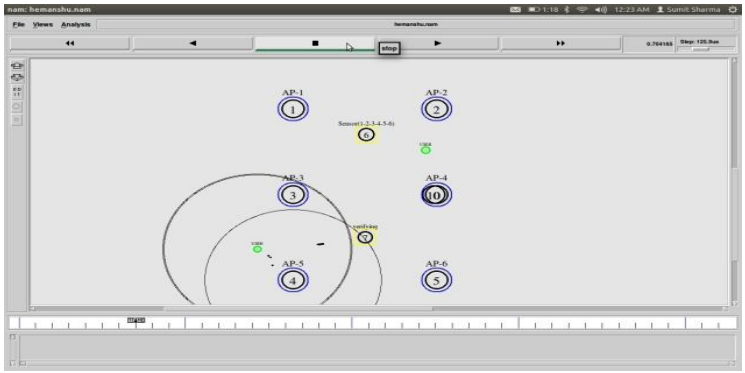



***Figure 6:*** *Verifying the access point identity by sensor node*

If the sensor node sense the communication and verifies from the database if no entry then that access point is fake to whom the client wants the communication. It will generate the alert message for the client , but if there is an entry in the database it sends for the second step of verification for the IP spoofing it decrypt the message that came from that AP and verifies that the received unique key is same or different if it is same it do nothing and if different that access point is fake to whom the client wants the communication. It will generate the alert message for the client. Figure 7 show that access point is not in the database of the sensor so it will generate the alert message for the client and declare that access point is fake.

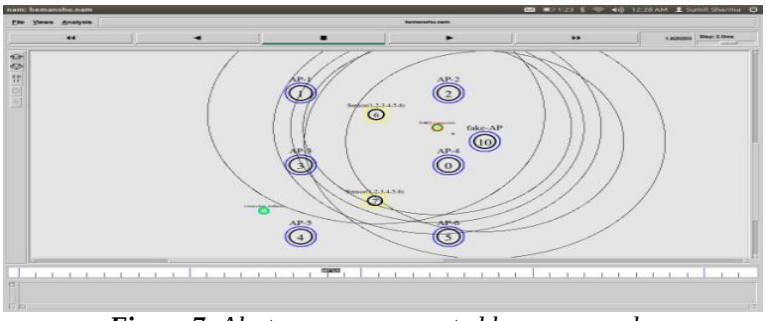



***Figure 7:*** *Alert message generated by sensor node*

## VI. Conclusion

This paper conclude that session hijacking is active type attack has very bad impact on the network .The fake access points will work like honey pot and used to gather network information. If the fake access points are detected which will work like a honey pot then session hijacking will be prevented .In this paper ,we propose sensor node based technique to prevent session hijacking by storing the information about the AP and this technique has vulnerability of MAC spoofing which is prevented through the use of public private key cryptography.

## References


[1] Thawatchai Chomsiri "A Comparative Study of Security Level of Hotmail, Gmail and Yahoo Mail by Using Session Hijacking Hacking Test" in International Journal of Computer Science and Network Security, VOL.8 No.5, May 2008.

[2] Srikanth Kamuni, S.ShreehaTejaswini, Bhaskar.J "Wavelet Based Real Time Session Hijack Detection Based On Bluetooth Signal Analysis" in GJCAT, Vol 2 (2), 2012, 1210-1213 ISSN: 2249-1945.

[3] Rupinder Gill, and Smith, Jason and Clark, Andrew "Experiences in Passively Detecting Session Hijacking Attacks in IEEE 802.11 Networks" in: Proceedings of 4th Australasian Information Security Workshop (Network Security), 16-19 January 2006, Hobart, Tasmania.

[4] Bhavna C.K. Nathani Erwin Adi "Website Vulnerability to Session Fixation Attacks" Journal of Information Engineering and Applications ISSN 2224-5782 (print) ISSN 2225-0506 (online)Vol 2, No.7, 2012.

[5] Al-Sammarraie Hosam ,Adli Mustafa and Shakeel "Exception Agent Detection System for IP Spoofing Over Online Environments" International Journal of Computer Science and Information Security, Vol. 6, No. 1, 2009.

[6] Mrs. Mridu Sahu and Rainey C. Lal "Controlling IP spoofing through packet filtering" Int.J.Computer Techology & Applications,Vol 3 (1),155-159 ISSN:2229-6093.

[7] Vimal Upadhyay , Rajeev kumar "Detection and preventing IP spoofing attack by Hashed Encryption" International Journal of Enterprise Computing and Business Systems ISSN (Online) : 2230-8849 Vol. 1 Issue 2 July 2011.

[8] Noureldien A. Noureldien, Mashair O. Hussein "Block Spoofed Packets at Source (BSPS): A method for Detecting and Preventing All Types of Spoofed Source IP Packets and SYN Flooding Packets at Source: A Theoretical Framework" International Journal of Networks and Communications 2012, 2(3): 33-37 DOI: 10.5923/j.ijnc.20120203.03.

[9] N.Arumugam, Dr.C.Venkatesh "A Dynamic Method to Detect IP Spoofing on Data Network Using Ant Algorithm" IOSR Journal of Engineering (IOSRJEN) e-ISSN: 2250-3021, p-ISSN: 2278-8719, www.iosrjen.org Volume 2, Issue 10 (October 2012), PP 09-16.

[10] Hemanshu Kamboj, Gurpreet Singh " Detection of Fake Access Point to Prevent Session Hijacking" International Journal for Advance Research and Technology Vol. 1, Issue II, Mar. 2013 ISSN 2320-6802.

[11] Kiruthiga.S and Yuvarani.G "Fast and Accurate Detection of Fake Access Points Using Non-crypto Method in WLAN" International Journal of Communications and Engineering Volume 05– No.5, Issue: 03 March2012.

[12] Jianping Song, Song Han, Aloysius K. Mok, Deji Chen, Mike Lucas, Mark Nixon "WirelessHART: Applying Wireless Technology in Real-Time Industrial Process Control" IEEE Real-Time and Embedded Technology and Applications Symposium 1080-1812/08 2008 IEEE DOI 10.1109/RTAS.2008.15.

[13] Suman Jana and Sneha k. Kasera " On Fast and Accurate Detection of Unauthorized Wireless Access Point Using Clock Skews" IEEE TRANSACTION ON MOBILE COMPUTING , VOL. 9 NO. 3 MARCH 2010.

[14] Stephen Glass NICTA "Detecting Man-in-the-Middle and Wormhole Attacks in Wireless Mesh Networks" 1550-445X/09 2009 IEEE DOI 10.1109/AINA.2009.131.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 74-82*
*www.iosrjournals.org*

# Ensuring Privacy in opportunistic Network

Er. Maggi Goyal[1,] Er. Manoj Chaudhary[2]
[1]*Research Scholar, Computer Science, Yadavindra College,Talwandi Sabo,Punjab,India.*
[2]*Lecturer,Computer Science, Yadavindra College,Talwandi Sabo,Punjab,India.*

***Abstract:*** *The emergence of extremely powerful mobile communication devices in recent times has triggered off the development of many exploitative technologies that attempt at leveraging the ever increasing processing, storage and communicating capacities of these devices. One of the most developing area of network is opportunistic network. It provides communication even in disconnected mode. Nodes are mobile and can change their location and message is forward through many intermediate nodes so identity of usersis shown to all.Any intermediate can drop the data packetsif he is not wishes to forward the data to a particular destination id. A few privacy preventing algorithms are proposed to maintain it. In this research we propose an algorithm to maintain the privacy of user if user wants it . We are ensuring the privacy of the data with the use of concept of cluster estimation. In this we use the public private cryptography technique for data encryption and decryption. Algorithm is implemented on NS2 (Network Simulator 2.35).*
***Keywords:*** *Attacks, Virtual ID, Privacy, NS2*

## I. Introduction

Opportunistic networks is a type of challenged networks. An opportunistic network is a sub-class of delay tolerance network where communication contacts are not constant, so an end-to-end path between the source and the destination may never exists. An opportunistic network may include cellular Base Stations (BSs), offering macrocell (macroBS), microcell, picocell, or femtocell (femtoBS) coverage, as well as WiFi access points (APs), mostly connected through wireless networks. The devices included in an opportunistic network can be mobile phones, personal computers, cameras, etc. In opportunistic networks each node acts as a gateway which makes it much more flexible than DTNs. Now the most basic question arises i.e. what is DTN network? How opportunistic network is different from mobile adhoc networks(MANETs)?

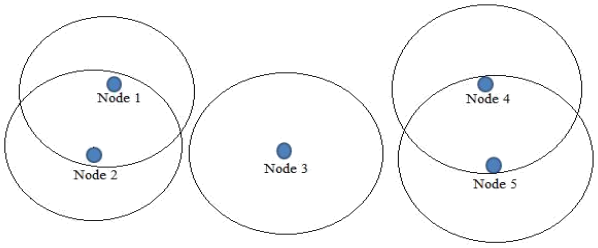



Figure 1 An Opportunistic Network

In Fig 1, Node 3 does not know the existence of Node 1 or Node 2 and Node 4 or Node 5, since they are not within its communication range. Node 1 and Node 2 do know of each other and are able to communicate. Similarly, Node 4 and Node 5 are also know each other and are able to communicate. Many opportunistic forwarding protocols are replication based. The replication factor depends on a heuristic which is used by intermediate nodes to decide either to forward the message or to drop it[7]. There are many examples of such networks in real life. For example, in north part of the Sweden[16], the communication between villages and the summer camps of the Saami population is provided when the nodes get connected. The same situation is also seen in rural villages of India and some other poor regions[17]. Other fields where this kind of communication scenarios may occur also include satellite communication[18], wildlife tracking[19], military networks[20] and vehicular ad hoc networks[21]. Ad hoc network is a decentralized type of wireless network. In ad hoc network there is no pre-existing infrastructure, such as routers in wired networks or access points in wireless networks. In ad hoc network each node participates in routing by forwarding data for other nodes, and so the determination of which nodes forward data is made dynamically based on the network connectivity. Ad-hoc networks are a new paradigm of wireless communication for mobile hosts. Basically it's a network which is used in emergency causes. Here is No fixed infrastructure in ad hoc network like base stations. Nodes within each other radio range communicate directly via wireless links while these which are far apart rely on other

nodes to relay messages. Wireless networks refer to those networks that make use of radio waves or microwaves in order to establish communication between the devices. The lack of end-to-end connectivity is a key difference between such networks and mobile ad-hoc networks (MANETs)[8].

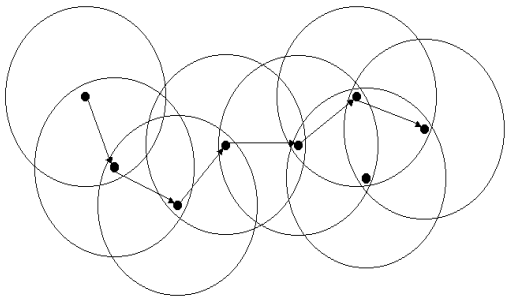


Figure 2: Mobile Adhoc Network

The complexity and uniqueness of MANETs make them more vulnerable to security threats than their wired counterparts. Attacks on ad hoc wireless networks can be classified as passive and active attacks, depending on whether the normal operation of the network is disrupted or not. A passive insider can attempt to track other nodes' movements by linking different location announcement messages. An active insider can modify, inject, and replay "genuine" messages[10]. Fig 2 shows the mobile adhoc network.

DTN is Delay Tolerant Network an approach to computer network architecture that seeks to address the technical issues in heterogeneous networks that may lack continuous network connectivity. The end-to-end path between source and the destination in DTNs is very low. Its is also referred as Intermittently Connected Mobile Networks. In delay tolerant network some problems are included in which lack of infrastructure and nodes are stationary are the main ones. In delay tolerant network routes are time dependent and requires long term storage. In Security of DTN routing, encrypted methods cannot detect packet drops in the malicious node[22]. The researchers have found the problem of Selfishness in Opportunistic networks[6]. Some main problems of opportunistic networks are given below.

1. Disclosure of the message content enabling it to be access by the malicious node or the parties which are not suppose to be read it.
2. Data is travel from many intermediate node, so there may be a threat to the integrity of the data.
3. Protection of transmitted messages in transit for malicious purposes e.g. for masquerading .

The Opportunistic network has the following features:-

- They are governed by operators through the provision of resources (e.g., spectrum available) and policies, as well as context/ profile information and knowledge, which is exploited for their creation/maintenance.
- They are extensions of the infrastructure that will include various devices and terminals potentially organized in an infrastructure-less mode, as well as elements of the infrastructure.
- They will exist temporarily, i.e. for the time frame necessary to support particular applications (requested in specific location and time). Applications can be related to the social networking and prosumer (derives from the combination of "producer" and "consumer") concepts as well as to the support of an enterprise (in a particular area and time interval) for developing and delivering products or digital services.
- At the lower layers, the operator designates the spectrum that will be used for the communication of the nodes of the opportunistic network (i.e. the spectrum derives through coordination with the infrastructure). In this respect, in principle, the bands will be licensed.
- The network layer capitalizes on context-, policy-, profile-, and knowledge-awareness to optimize routing and service/content delivery.

There are also different types of attacks in opportunistic networks:-

A) Worm Attack
B) Viruses Attack.

**Viruses Attack:-** A computer virus is a type of malware that propagates by inserting a copy of itself into and becoming part of another program. It spreads from one computer to another, leaving infections as it travels. Viruses can range in severity from causing mildly annoying effects to damaging data or software and causing denial-of-service (DoS) conditions. Almost all viruses are attached to an executable file, which means the virus may exist on a system but will not be active or able to spread until a user runs or opens the malicious host file or

program. When the host code is executed, the viral code is executed as well. Normally, the host program keeps functioning after it is infected by the virus. However, some viruses overwrite other programs with copies of themselves, which destroys the host program altogether. Viruses spread when the software or document they are attached to is transferred from one computer to another using the network, a disk, file sharing, or infected e-mail attachments.

**Worms Attacks:-** Computer worms are similar to viruses in that they replicate functional copies of themselves and can cause the same type of damage. In contrast to viruses, which require the spreading of an infected host file, worms are standalone software and do not require a host program or human help to propagate. To spread, worms either exploit a vulnerability on the target system or use some kind of social engineering to trick users into executing them. A worm enters a computer through a vulnerability in the system and takes advantage of file-transport or information-transport features on the system, allowing it to travel unaided.

## II. Background And Related Work

The opportunistic networks (OppNets) are characterized as a most challenging evolution of Mobile Ad – Hoc Networks (MANET). OppNet provide possibility to exchange messages between mobile nodes (users) even in such a disconnected environment by opportunistically selection any nearby device to move messages closer to the final nodes. The privacy that we use in opportunistic network is very different form than the privacy in opportunistic networks has been given by various researchers.

### A. Routing Protocols in Opportunistic Network

**A.1** Anna Scaglione in year 2003 proposed a technique called Opportunistic Large Array for sending the data to very long distance this technique allows efficient flooding of a wireless network with information from a source, which we refer to as the leader. At the same time, it permits us to transmit reliably to far destinations that the individual nodes are not able to reach without consuming rapidly their own battery resources, even when using multihop links (the reach-back problem). The synchronization constraints are extremely loose and can be fulfilled in a distributed manner. The key idea is to have the nodes simply echo the leader's transmission operating as active scatterers while using adaptive receivers that acquire the equivalent network signatures corresponding to the echoed symbols. The active nodes in the network operate either as regenerative or nonregenerative relays. The intuition is that each of the waveforms will be enhanced by the accumulation of power due to the aggregate transmission of all the nodes while, if kept properly under control, the random errors or the receiver noise that propagate together with the useful signals will cause limited deterioration in the performance. The avalanche of signals triggered by the network leaders form the so-called opportunistic large array (OLA)[1].

**A.2** Chiara Boldrini, Marco Conti, Andrea Passarella proposed a technique for routing and forwarding in opportunistic network in year 2008 Opportunistic networks allow content sharing between mobile users without requiring any pre-existing Internet infrastructure, and tolerate partitions, long disconnections, and topology instability in general. In this paper they propose a context-aware framework for routing and forwarding in opportunistic networks. The framework is general, and able to host various flavors of context-aware routing.

In this work they also present a particular protocol, HiBOp, which, by exploiting the framework, learns and represents through context information, the users' behavior and their social relations, and uses this knowledge to drive the forwarding process. The comparison of HiBOp with reference to alternative solutions shows that a context-aware approach based on users' social relations turns out to be a very efficient solution for forwarding in opportunistic networks[2].

**A.3** Zehua Wang gives a noble approach in year 2012 for routing in opportunistic network. this paper proposed that The link quality variation of wireless channels has been a challenging issue in data communications until recent explicit exploration in utilizing this characteristic. The same broadcast transmission may be perceived significantly differently, and usually independently, by receivers at different geographic locations. Furthermore, even the same stationary receiver may experience drastic link quality fluctuation over time. The combination of link-quality variation with the broadcasting nature of wireless channels has revealed a direction in the research of wireless networking, namely, cooperative communication. Research on cooperative communication started to attract interests in the community at the physical layer but more recently its importance and usability have also been realized at upper layers of the network protocol stack. In this article, we tackle the problem of opportunistic data transfer in mobile ad hoc networks. Our solution is called Cooperative Opportunistic Routing in Mobile Ad hoc Networks (CORMAN). Nodes in the network use a lightweight proactive source routing protocol to determine a list of intermediate nodes that the data packets should follow en route to the destination. Here, when a data packet is broadcast by an upstream node and has happened to be received by a downstream node further along the route, it continues its way from there and thus will arrive at the destination node sooner[3].

**A.4** J.P. Tower and T.D.C. Little proposed in 2008 Opportunistic networking is emerging as a technique to exploit chance encounters among mobile nodes, and is distinct from previously studied behaviors found in sensor and ad hoc networking research. In this paper, explore contributions in epidemic data dissemination and mobile ad hoc networks applicable to opportunistic networking, and propose an extension to prior work based on *active cures*. This scheme, called SERAC, increases the rate at which cure messages are propagated in a fragmented network for the purpose of reducing the overhead of outstanding yet incompletely disseminated messages.
Preliminary analyses demonstrate the feasibility of performance gains under the opportunistic networking model[4].

**A.5** Gokce Gorbil, Erol Gelenbe proposed opportunistic communication for Emergency Support Systems in 2011and Opportunistic communications (oppcomms) use low-cost human wearable mobile nodes allowing the exchange of packets at a close range of a few to some tens of meters with limited or no infrastructure. Typically cheap pocket devices which are IEEE 802.15.4-2006 compliant can be used and they can communicate at 2m to 10m range, with local computational capabilities and some local memory. In this paper we consider the application of such devices to emergency situations when other means of communication have broken down. This paper evaluates whether oppcomms can improve the outcome of emergency evacuation in directing civilians safely. We describe an autonomous emergency support system (ESS) based on oppcomms to support evacuation of civilians in a built environment such as a building or supermarket. The proposed system uses a fixed infrastructure of sensor nodes (SNs) to monitor the environment[5].

**B. Security in Opportunistic Network**

**B.1** B.Poonguzharselvi1 and V.Vetriselvi proposed a trusted framework for data forwarding in opportunistic network in year 2012 and Opportunistic networks are usually formed spontaneously by mobile devices equipped with short range wireless communication interfaces. The idea is that an end-to-end connection may never be present. Designing and implementing a routing protocol to support both service discovery and delivery in such kinds of networks is a challenging problem on account of frequent disconnections and topology changes. In this network one of the most important issues relies on the selection of the best intermediate node to forward the messages towards the destination. This paper presents a trust framework for opportunistic network where the nodes in the network follow the trace based mobility model. The selection of next hop to forward the data packets is based on the trust value as well as the direction of movement of node towards the destination.

The trust value is obtained from the trust framework of the data forwarding node. The direction of destination is obtained from the movement trace file that is maintained by the nodes in the network. In this proposed framework, the message is encrypted to secure both the data and path information. The effectiveness of this proposed framework is shown using simulation[6].

**B.2** Abdullatif Shikfa , Melek Onen , Refik Molva proposed in year 2010 and give privacy and confidentiality in context-based and epidemic routing in opportunistic network and Autonomic and opportunistic communications require specific routing algorithms, like replication-based algorithms or context-based forwarding. In addition to confidentiality, privacy is a major concern for protocols which disseminate the context of their destination. In this paper, we focus on the confidentiality 20 and privacy issue inherent to context-based protocols, in the framework of an original epidemic forwarding scheme, which uses context as a heuristic to limit the replication of messages. We define the achievable privacy level with respect to the trusted communities assumption, and the security implications. Indeed, privacy in such an environment raises challenging problems, which lead us to a solution based on refinements of two pairing-based encryption, namely searchable encryption and identity-based encryption. This new solution enables forwarding while preserving user privacy by allowing secure partial matches in the header and by enforcing payload confidentiality[7].

**B.3** Abdullatif Shikfa and Melek Önen and Refik Molva proposed in year 2012 gives a technique for securing data forwarding in opportunistic network by Local key management and Opportunistic networks are a new and specific type of mobile peer-to-peer networks where end-to-end connectivity cannot be assumed. These networks present compelling challenges, especially from a security perspective, as interactive protocols are infeasible in such environments. In this article, focus on the problem of key management in the framework of content-based forwarding and opportunistic networks. After analysing this issue and identifying specific security threats such as Sybil attacks, propose a specific key management scheme that enables the bootstrapping of local, topology-dependent security associations between a node and its neighbours along with the discovery of the neighbourhood topology[8].

**B.4** Enrico Scalavino, Giovanni Russello and Rudi Ball proposed a novel approach for security in opportunistic network in year 2010 that a novel version and implementation of the Policy-based Authority Evaluation Scheme (PAES) to protect data disseminated amongst the responders to an emergency situation when no network connectivity is available. In such situations Delay Tolerant Networks (DTN) are used to disseminate the data by exploiting the peers' mobility in the area. However, existing DTN protection models

require recipients to be known in advance. In emergency situations the data may instead be received by unknown responders who might need it while carrying out their duties. Existing data dissemination solutions such Enterprise Rights Management (ERM) systems rely on centralized architectures where recipients must contact the authorities that can grant access to data. Such centralized solutions cannot be deployed when connectivity cannot be guaranteed. Our solution combines data protection schemes such as ERM systems with DTNs[9].

Table 1: Table of Techniques Used in Related Work

| Author(s) | Year | Paper Name | Technique | Result |
|---|---|---|---|---|
| **A. Routing Protocols in Opportunistic Network** | | | | |
| Anna Scaglione[1] | 2006 | Opportunistic Large Arrays: Cooperative Transmission in Wireless Multihop *Ad Hoc* Networks to Reach Far Distances | Opportunistic Large Array | application may arise in joint control systems and security or military scenarios |
| Chiara Boldrini, Marco Conti, Andrea Passarella[2] | 2008 | Exploiting users' social relations to forward data in opportunistic networks | context-aware approach | this approach allows automatically control congestion in opportunistic networks |
| J.P. Tower and T.D.C. Little[4] | 2008 | A Proposed Scheme for Epidemic Routing with Active Curing for Opportunistic Networks | Active Curing | We must use ER-based routing such as SERAC in a low-power radio mode for delay tolerant traffic |
| Gokce Gorbil, Erol Gelenbe[5] | 2011 | Opportunistic Communications for Emergency Support Systems | Sensor Nodes | Not secure |
| Zehua Wang[3] | 2012 | CORMAN: A Novel Cooperative Opportunistic Routing Scheme in Mobile Ad Hoc Networks | lightweight proactive source | |
| **B. Security in Opportunistic Network** | | | | |
| Abdullatif Shikfa , Melek Onen , Refik Molva[7] | 2010 | Privacy and confidentiality in context-based and epidemic forwarding | Two pairing-based encryption, | specific use of PEKS allows intermediate nodes to securely discover partial matches between their profile |
| Enrico Scalavino, Giovanni Russello and Rudi Ball[9] | 2010 | An Opportunistic Authority Evaluation Scheme for Data Security in Crisis Management Scenarios | Policy-based Authority Evaluation Scheme (PAES) | |
| Abdullatif Shikfa and Melek Önen and Refik Molva[8] | 2012 | Local key management in opportunistic networks | | This also prevents Sybil attacks |
| B.Poonguzharselvi1 and V.Vetriselvi[6] | 2012 | Trust Framework for Data Forwarding in Opportunistic network Using Mobile Traces | Selection of Next Hop on trust Value | High delivery probability |

## III. Problems Formulation

**a) Routing problems**

In Opportunistic Networks (OppNets), routing is one of the main challenges. The protocols can be Epidemic Routing, PROPHET, Spray and Wait. The Epidemic routing [13] protocol is a flooding based scheme [15]. In Epidemic routing protocol each node receives a request packet and forwards the packet on its entire outgoing links except the one corresponding to the incoming link on which the packet arrives. Each request packet may reach the destination node along a different route at a different time [10][14]. The context information used in PRoPHET is the frequency of meetings between nodes, as is also seen in the MV (Meeting and Visits) and MaxProp protocols [11][12]. The design of efficient routing strategies in conventional networks is usually based on the knowledge of the available infrastructure and the network topology, whether physical or logical. Unfortunately, such knowledge is not available in such networks as the formation of data path is entirely

opportunity based. For achieving a reliable data path a trade off against the performance of the network must be met before designing the routing strategy. As the nodes are mobile and not aware about any other node until comes in his range, here finding an efficient routing protocol is difficult.

**b) Privacy**

The opportunistic networks present compelling challenges, especially from a security perspective. The key problem in opportunistic network is privacy. Privacy of user location, identity and the message confidentiality is the main problem. For data confidentiality we can use encryption, but we have to maintain the key of each node in the diameter of opportunistic network. And the nodes are not so big to store a number of keys and data which is carry by the nodes. Location of the user is shown to all because it can only pass the message when some other node comes in the range, hence it is easy to find that both node are at the same location and time while exchange the data. User identity is also degrading this network performance, a selfish node which is not interested to forward the message of a particular sender of his cluster or receiver of his cluster, so he drops the packet. This leads to loss of data packet.

## IV. New Proposed Technique

In this research an algorithm is propose to maintain the privacy of user if user wants it. To divide the network into the clusters our research uses the concept of cluster estimation.

In this the network is initialized with the finite number of nodes. The whole network is divided into clusters. In each cluster one fixed node is defined. In fixed node database is maintained and ID of each node and password is stored. The node which wants to communicate to the other node will first communicate with the fixed node to get the virtual ID. The source node send its credentials (USER_NAME & PASSWORD) which authenticate the valid node of group and also with which node it want to communicate) to the fixed node. When fixed node verifies the credentials ,fixed node communicate with the stable node of the cluster in which destination is present and will send the virtual ID's of the source and destination , to the source node + the secure session key. Once the source node authenticates the user then stable node communicate with the stable node of destination cluster and exchange information and update their table. Now stable node of source cluster will sends a new virtual ID to the source , new ID of destination and a session key with which the source node encrypt the message. This message is now encrypt by the public key of the source node. And stable node also sends a new virtual ID to the destination, virtual id of source node and session decryption key. This message is encrypted by the public key of destination node. Source node now send message with new ID and encrypt the message with session key.

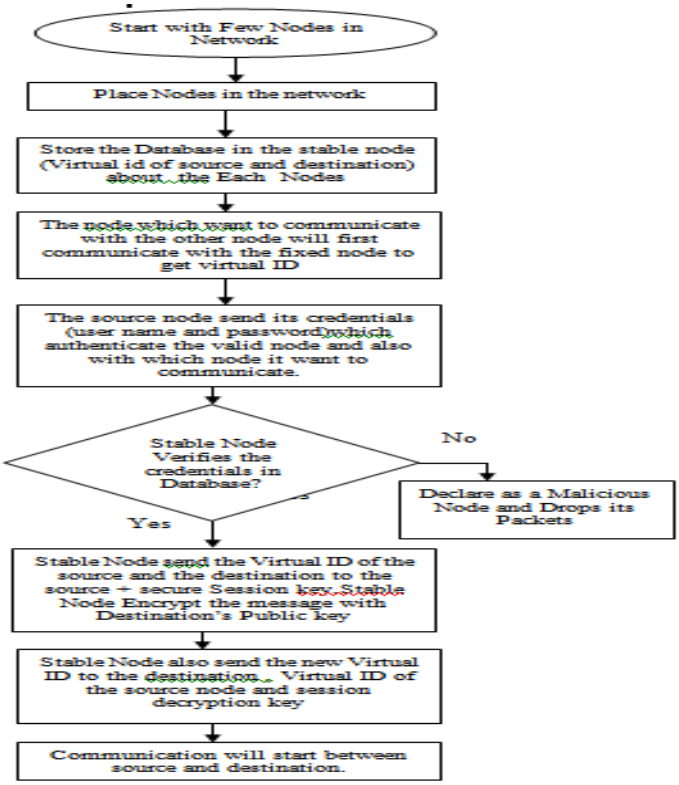



*Figure 3 Flowchart of Proposed Technique*

## V. Results And Discussion

**Results**

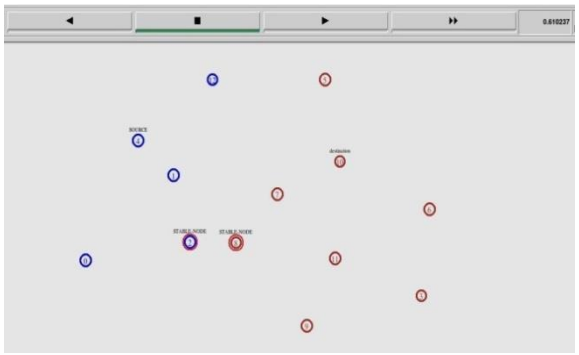


*Figure 4 Proposed network designs*

We consider two cluster differentiated by color and two stable node, one for each cluster shown by thick outline. Take Node4 as source node and Node10 as destination node (both are lies in different cluster). Features of the stable node is same as the other node it also transfer a message only when other nodes comes in his communication range. ID Source node request to stable node for a Virtual ID. Stable node check the table to authenticate the valid user.

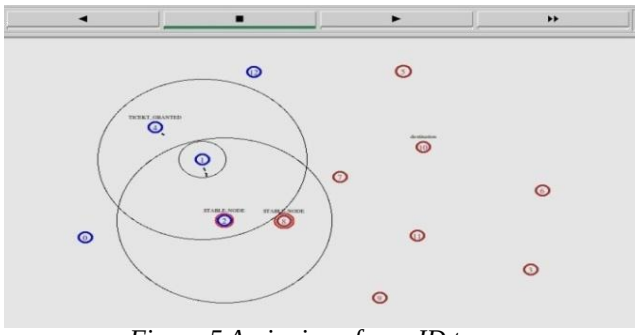


*Figure 5 Assigning of new ID to source*

After sharing data, stable node provides a new id to the source node. Stable node sends ticket to the source node after authentication and sharing information with the stable node of cluster in which destination is present.

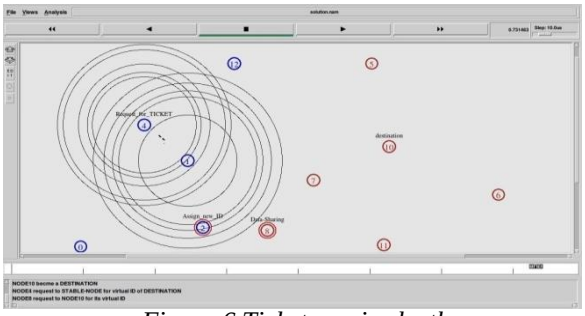


*Figure 6 Ticket receive by the user*

Ticket contains:

- A new virtual ID
- New destination ID
- A session key to encrypt the message.

This ticket is encrypted by the public key of source node which is present in the table of stable node. After getting the new ID, source node send the message and encrypt with the session key provides by the stable node. The message is encypted by the public key of the destination node. Destination node also gets a new ID by stable node8.

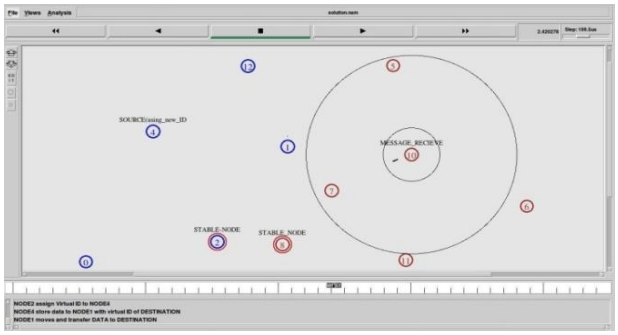



*Figure 7 Message forward by intermediate node to the closer one in his range*

Now the data carrying node (Node1) forward data to Node7 when appears in his communication range. Message reaches at the destination.Node7 forward the message to the destination. It will send the data to a particular destination in a very secure manner.

## VI. Discussion

a) Comparison of both techniques in term of packet loss

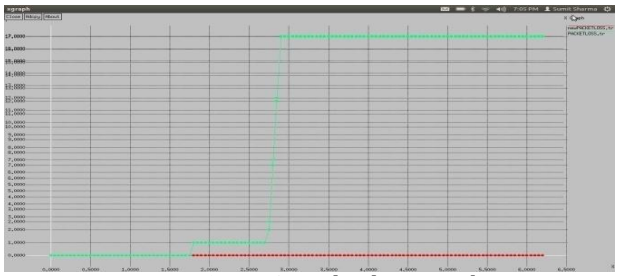


*Figure 8 Packet loss graph*

- Graph is plotted between the number of packet (y-axis) and the time (x-axis).
- Green line shows the packet loss by the previous method.
- Red line shows the packet loss by proposed algorithm.

b) Comparison of both methods in the form of throughput

*Figure 9 Throughput graph*

- Red line shows throughput by net algorithm
- Green line show throughput by previous algorithm
- Throughput is lies between 0 and 1 along y-axis
- Time along x-axis

## VI. Conclusion And Future Scope

**Conclusion**

In this research our work finds that opportunistic network is very useful if privacy is maintained. Due to inefficient in privacy, many false node or selfish nodes in this network do not want to forward the data to the destination. This results the increase in packet loss and decrease in throughput. To reduce the packet loss and increase the throughput, this research propose a network architecture in which a node want to send a message to

a destination and he doesn't want to explore his identity as well as destination identity, then he first communicate with stable node(trusted node) and get a virtual id for a period of time. Stable node act as special node, which contains information of every node of the cluster and authenticate the nodes who wishes to communicate and provide virtual id to that nodes. And also a session key is provided for encryption of message to the source and decryption key to the destination for maintaining the confidentiality of the message. The public private cryptography technique is used for data encryption and decryption. This approach provides privacy to the user and reduces the packet loss by a selfish node.

**Future Work**

In our work we purpose a technique to provide privacy to a user in opportunistic network on the basis of providing virtual id by a stable node, which is present in every cluster. But this technique increases the work load of a sender who wishes to communicate. In future our research tries to find a way which reduces the user work by providing some new mechanism to hide the id of user.

Due to infrastructure less architecture and mobility of the nodes, opportunistic network faces many problems related to security, privacy, nodes authentication and efficient routing protocol.

## References


[1]. Anna Scaglione "Opportunistic Large Arrays: Cooperative Transmission in Wireless Multihop Ad Hoc Networks to Reach Far Distances" IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 51, NO. 8, AUGUST 2003.

[2]. Chiara Boldrini, Marco Conti, Andrea Passarella "Exploiting users' social relations to forward data in opportunistic Networks" 1016/j.pmcj.2008.04.003 2008 Elsevier.

[3]. Zehua Wang "CORMAN: A Novel Cooperative Opportunistic Routing Scheme in Mobile Ad Hoc Networks" IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 30, NO. 2, FEBRUARY 2012.

[4]. J.P. Tower and T.D.C. Little "A Proposed Scheme for Epidemic Routing with Active Curing for Opportunistic Networks" In Proc. 1st IEEE Intl. Workshop on Opportunistic Networking, Okinawa, Japan, March 2008.

[5]. Gokce Gorbil, Erol Gelenbe "Opportunistic Communications for Emergency Support Systems" 1877–0509 © 2011 Published by Elsevier 10.1016/j.procs.2011.07.008.

[6]. B.Poonguzharselvi1 and V.Vetriselvi "Trust Framework for Data Forwarding in Opportunistic network Using Mobile Traces" International Journal of Wireless & Mobile Networks (IJWMN) Vol. 4, No. 6, December 2012.

[7]. Abdullatif Shikfa , Melek Önen , Refik Molva "Privacy and confidentiality in context-based and epidemic forwarding" 2010 Published by Elsevier B.V:10.1016/j.comcom.2010.04.035

[8]. Abdullatif Shikfa and Melek Önen and Refik Molva "Local key management in opportunistic networks" Int. J. Communication Networks and Distributed Systems, Vol. 9, Nos. 1/2, 2012.

[9]. Enrico Scalavino, Giovanni Russello and Rudi Ball "An Opportunistic Authority Evaluation Scheme for Data Security in Crisis Management Scenarios" ASIACCS'10 April 13–16, 2010, Beijing, China.

[10]. Ram Ramanathan, Richard Hansen "Prioritized Epidemic Routing for Opportunistic Networks" June11, 2007, San Juan, Puerto Rico, USA.

[11]. Pelusi, L., Passarella, A. and Conti, M. (2006) "Opportunistic networking: data forwarding in disconnected mobile ad hoc networks", IEEE Communications Magazine, Vol. 44, pp.134–141.

[12]. Thrasyvoulos Spyropoulos, Konstantinos Psounis, Cauligi S. Raghavendra "Efficient Routing in Intermittently Connected Mobile Networks: The Multiple-Copy Case" IEEE/ACM Transactions on Networking, Vol. 16, No. 1, February 2008.

[13]. Mamoun Hussein Mamoun, Saud Barrak "Adaptive Priority Routing Protocol for DTN Networks" International Journal of Engineering and Technology Volume 3 No. 3, March, 2013.

[14]. A. Vahdat and D. Becker, "Epidemic routing for partially-connected ad hoc networks" Duke University, Tech. Rep. CS-2000-06, Jul. 2000.

[15]. S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," in Proc. ACM SIGCOMM, Oct. 2004.

[16]. A. Doria, M. Uden, and D. P. Pandey, Providing connectivity to the saami nomadic community, in Proceedings of the 2nd International Conference on Open Collaborative Design for Sustainable Innovation (dyd 02), Bangalore, India, Dec 2002.

[17]. A. Pentland, R. Fletcher, and A. A. Hasson, A road to universal broadband connectivity, in Proceedings of the 2nd International Conference on Open Collaborative Design for Sustainable Innovation (dyd 02), Bangalore, India, Dec 2002.

[18]. G. E. Prescott, S. A. Smith, and K. Moe, Real-time information system technology challenges for NASAs earth science enterprise, in Proceedings of The 20th IEEE Real-Time Systems Symposium, Phoenix, Arizona, Dec 1999.

[19]. P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebranet, in Proceedings of ACM ASPLOS, 2002.

[20]. Disruption tolerant networking, http://www.darpa.mil/ato/solicit/DTN/.

[21]. J. Ott and D. Kutscher, A disconnection-tolerant transport for drive-thru internet environments, in Proceedings of IEEE INFOCOM, 2005.

[22]. Eyuphan Bulut and Boleslaw K. Szymanski, "On Secure Multi-copy based Routing in Compromised Delay Tolerant Networks," Workshop on Privacy, Security and Trust in Mobile and Wireless systems at 20th IEEE International Conference on Computer Communications, ICCCN, Maui, Hawaii, July 31,2011.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 83-92*
*www.iosrjournals.org*

# An Overview of TRIZ Problem-Solving Methodology and its Applications

Hajar Mat Jani
*(College of Information Technology, Universiti Tenaga Nasional, Malaysia)*

***Abstract:*** *TRIZ, which is a Russian word that stands for "Theory of Inventive Problem Solving", is a problem-solving methodology that was invented based on the belief that there are universal principles of invention that are the foundation for creative innovations that help in advancing technology. One of the most widely used approaches to problem-solving in most technology-related fields is the structured problem-solving methodology. Actually, TRIZ enhances the structured problem-solving methodology by applying its principles to the first few phases of the conventional structured methodology with more creative and advanced steps that make the problem-solving process more efficient and effective. A brief description of the structured problem-solving methodology is presented since it has a big influence on the original TRIZ methodology. TRIZ's ways of solving problems are explained in detail, and several main applications of TRIZ in the technology-related and other new fields (non-technology) are discussed. Several suggestions are put forward in order to overcome some of the main issues faced by TRIZ in order to improve its effectiveness in solving problems, especially in non-technical and non-scientific domains.*

***Keywords:*** *Inventive Problem Solving, Problem-Solving Methodology, Structured Problem-Solving, TRIZ*

## I. INTRODUCTION

In general, a problem is any difficulty, obstacle or issue that needs to be analyzed and overcome using factual knowledge when solving the problem. On the other hand, problem solving is a cognitive process in which one is required to identify the exact problem and find the solution to the identified problem [1]. Normally, a series of steps should be followed systematically even though sometimes certain steps are skipped or repeated several times depending on the types of problems at hand. It is also important to note that problem solving only occurs whenever an individual or a group of people wants to move from a given undesirable or problematic current state to a desired state or goal [1].

The following is a list of steps that is normally used in problem-solving [2]:

- Identify the problem
- Define the problem
- Determine the best strategy that suits the problem
- Organize and gather facts and knowledge regarding the current problem
- Resource allocation (time, money, people, etc.)
- Monitor progress
- Evaluate the results

There are many techniques and methodologies to problem solving, and one of them is TRIZ methodology. TRIZ is a Russian word that stands for "Theory of Inventive Problem Solving" or TIPS, which is the equivalent phrase for TRIZ in Russian [3]. TRIZ was developed in 1946 by Genrich Altshuller and his colleagues in the former USSR, and it is now being used widely throughout the world in solving complex inventive problems [4][5][6].

TRIZ was created based on the theory or belief that "there are universal principles of invention that are the basis for creative innovations that advance technology, and that if these principles could be identified and codified, they could be taught to people to make the process of invention more predictable [4]". Altshuller discovered that invention is nothing more than the removal of technical contradiction with the assistance of a set of known principles. He emphasized that one does not have to be born an inventor in order to be a good inventor, and he criticized the trial and error method that are normally used to make discoveries [7].

The main rule is that the progress and evolution of technological systems is governed by a set of objective laws, which Altshuller called *Laws of Technological System Evolution* [5][8]. To devise these laws, Altshuller originally started by analyzing around 200,000 patterns and invention descriptions from various fields of engineering from available world-wide patent databases. Altshuller also made a conclusion that there were around 1,500 technical contradictions that could be solved easily by simply applying the discovered principles [7].

From his thorough study and analysis, Altshuller selected and examined the most effective solutions - "the breakthroughs [5][8]." As a result, the following three main findings are concluded [4][6][8]:

- Repetitive problems and solutions occurred across industries and sciences
- Patterns of technical evolution and advancement were repeated across industries and sciences
- Innovations used scientific effects outside the field where they were developed

The above main findings are applied in TRIZ for creating new products or inventions and also to improve current products, systems, and services.

In addition, based on the analysis done on the selected 40,000 innovative patterns, 40 *Inventive Principles* were formulated. In fact, it was also discovered that inventiveness could be taught and trained to others. To date, almost 3 millions existing patents were examined and studied, classified by their level of inventiveness and analyzed thoroughly with the intention of finding new useful principles of innovation [8].

In the application of TRIZ all of the above findings are employed to create and to improve products, services, and systems. In addition, new creations of products are also possible by observing past inventions patterns in different technology fields. It is very important to note that, in this information technology (IT) era, most products' life spans are shortened since at almost any time around the world someone is introducing something new and because of that newer products must be produced the soonest possible in order to have the products in the market faster than the competitors. This leads to very short development time and rapid development of products is required. And to ensure that the products can be marketed faster, efficient and effective methodologies are required, and this is where TRIZ is very useful.

## II. MOTIVATION

The main goal of this research is to perform a study on TRIZ problem-solving methodology by reviewing its fundamental concepts and the various TRIZ applications in solving engineering-related, technology-related or scientific-related, and also non-technology-related problems. Several objectives of this paper are as the following:

- To review past research and works on TRIZ problem-solving methodology.
- To highlight in detail several engineering applications and non-engineering or non-technical applications that have used TRIZ.
- To propose several suggestions that can help improve TRIZ problem-solving effectiveness.

## III. TRIZ BACKGROUND

TRIZ is a Russian acronym for "Teoriya Reshiniya Izobreatatelskikh Zadatch", which means 'Theory of Inventive Problem Solving" in English [8]. TRIZ can be regarded as a philosophy, a process, and a series of tools based mainly on the notion of resolving contradictions [8]. As widely known, problem solving is the "heart of improving designs and the processes to make them [5]." Innovation involves *continuous improvements* to existing designs and processes. TRIZ is a systematic approach for finding advanced and creative solutions to difficult problems in a more efficient and effective manner to ensure that the solutions are up-to-date and still relevant during its launching.

In TRIZ, it is assumed that the degree of complexity of a problem mainly depends on the way the problem is formulated [5], and the clearer the formulation, the most likely and easily that the solution is going to be found. In order to have a good formulated problem, a series of successive reformulations of the initial problem is conducted until an initially ill-defined problem is transformed into a much clear formulated problem with obvious solution or it becomes clearer that the problem cannot be resolved because lacking in the required technology or scientific knowledge [5].

It was discovered that Engineering Systems progress towards "*Ideality*" by overcoming existing contradictions within the systems and these Engineering System evolutions are driven by objective laws [8]. Problems tend to repeat across industries and sciences and it was found that the solutions used to resolve these problems are also repeated correspondingly.

Basically, a general TRIZ problem-solving methodology is shown in Fig. 1 [5]:

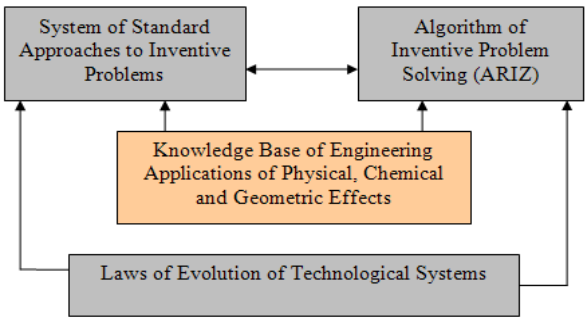



Figure 1. TRIZ problem-solving methodology

The four main components of the TRIZ methodology along with brief descriptions of each component are as follows [5]:

- *Laws of Evolution of Technological Systems*: In TRIZ it is believed that the evolution of technological systems is governed by a set of objective laws that can be used as the basis for problem solving These formulated laws can be reused instead of searching blindly to deliberately develop technological systems (or to solve new problems) [5].
- *System of Standard Approached to Inventive Problems*: "A set of rules for problem solving based on the laws established by Altshuller stating that many problems from different areas of technology can be solved by the same conceptual approaches [5]."
- *Algorithm of Inventive Problem Solving (ARIZ)*: A set of sequential and logical procedures aimed at solving and minimizing the system conflict at the core of the problem.
- *Knowledge Base of Engineering Applications of Physical, Chemical and Geometric Effects*: Contains knowledge of past solutions to similar problems; can facilitate problem solving by suggesting analogies from previous creative solutions.

## IV. STRUCTURED AND TRIZ PROBLEM-SOLVING METHODOLOGIES

Various techniques and approaches are used in solving complex problems, but the most common approach used in solving software development or engineering problem is the structured methodology. Ideally, the structured methodology consists of several phases or steps that must be followed in order to solve problems systematically. Fig. 2 illustrates the various phases within the conventional structured problem-solving methodology [6] that is normally used in many problem-solving domains.

A variation of the structured methodology to problem solving is presented in Fig. 3 [8] where an iteration is added to check the effectiveness of the solution. If the implemented solution does not solve the problem, then a new solution is generated. This repetitive step in carried out until a satisfactory solution that solves the problem is generated.

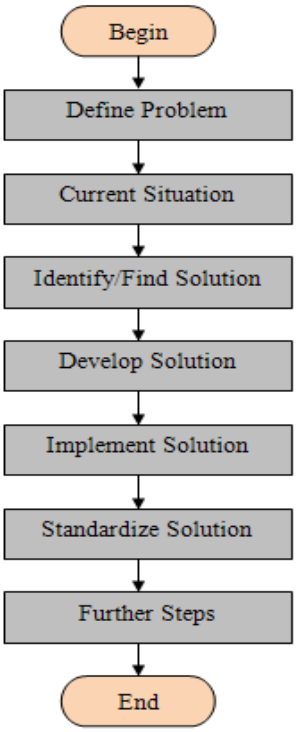



Figure 2. Conventional structured problem-solving methodology's phases [6]

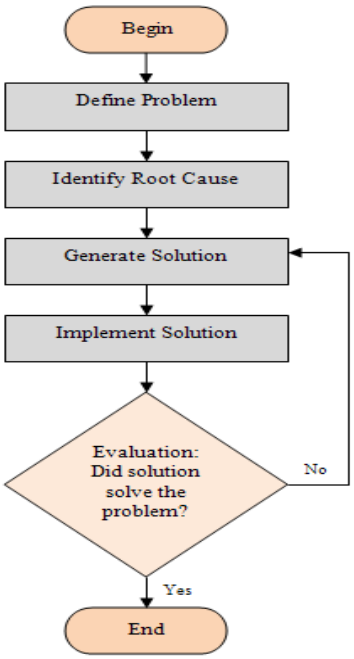



Figure 3. Structured problem-solving methodology's phases with an iteration [8]

When TRIZ was first introduced, it was used to complement the conventional structured problem-solving methodology because the two methodologies have a lot in common. Several earlier phases of TRIZ are almost the same as the structured methodology. It has been proven that using both the TRIZ and structure problem-solving methodologies in finding solutions to problems has resulted in better innovative results. Fig. 4 presents the steps taken in TRIZ problem-solving methodology [8][9].

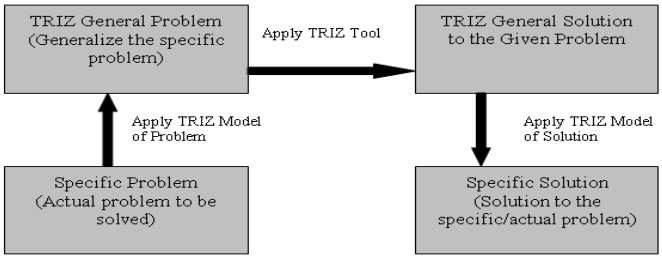



Figure 4. TRIZ's way of solving a problem

Based on Fig. 4, it is obvious that TRIZ methodology of problem solving is not the same as the normal problem-solving process. In a normal problem-solving scenario, the problem solver directly tries to find a specific solution to a specific problem, and in most cases, this is difficult to accomplish because of the complexity of the problem. In addition, finding a specific solution from scratch is very time-consuming. Most of the time, there exist contradictions among the various parameters that prevent the generation of good solutions [8] to the problems faced, and TRIZ simplifies this process.

Fig. 5 presents a modification of the diagram in Fig. 4 that shows what tool and what principles are used in solving the general problem. It also shows where the 39 *System Parameters* are used [10][11] within TRIZ's steps.

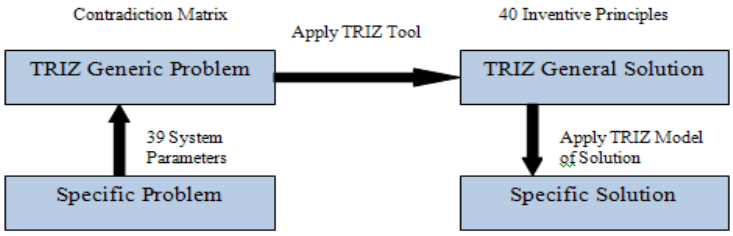



Figure 5. Schema of solution of problems using TRIZ [10][11]

The first step in TRIZ methodology is to convert the specific problem into a TRIZ general problem that basically models the problem [12]. The general problem is considered as the *Model of Problem*, and TRIZ has the required *Tool* for resolving this *Model of Problem*. An example of a *Tool* is the *Contradiction Matrix* [8]. A *Contradiction Matrix* is a TRIZ Tool that is used for generating potential *Inventive Principles* that can assist in finding the right solution. A contradiction in engineering problems is a situation where an attempt to improve a specific characteristic or parameter of the system causes another characteristic to worsen [8]. In TRIZ, an engineering system's parameters must be mapped to the pre-determined 39 *System Parameters* or *Contradiction Parameters* [8][10][13][14]. Once all the improving and worsening *System Parameters* are identified, then only the *Contradiction Matrix* is used in generating potential solution to the problem. A specific solution must be derived by the user based on the suggested general TRIZ solution by referring to the *40 Inventive Principles* [8][10][13][14] and the *76 Inventive Solutions*. This would result in the *Model of Solution* to the specific problem.

TABLE I. TRIZ 39 SYSTEM PARAMETERS [13][14]

| | | | |
|---|---|---|---|
| 1. Weight of moving object | 11. Tension, pressure | 21. Power | 31. Harmful side effects |
| 2. Weight of nonmoving object | 12. Shape | 22. Waste of energy | 32. Manufacturability |
| 3. Length of moving object | 13. Stability of object | 23. Waste of substance | 33. Convenience of use |
| 4. Length of non-moving object | 14. Strength | 24. Loss of information | 34. Repairability |
| 5. Area of moving object | 15. Durability of moving object | 25. Waste of time | 35. Adaptability |
| 6. Area of nonmoving object | 16. Durability of nonmoving object | 26. Amount of substance | 36. Complexity of device |
| 7. Volume of moving object | 17. Temperature | 27. Reliability | 37. Complexity of control |
| 8. Volume of nonmoving object | 18. Brightness | 28. Accuracy of measurement | 38. Level of automation |
| 9. Speed | 19. Energy spent by moving object | 29. Accuracy of manufacturing | 39. Productivity |
| 10. Force | 20. Energy spent by nonmoving object | 30. Harmful factors acting on object | - |

Table I [8][10][13][14] presents a full list of the *39 System Parameters*, while Table II [8][10][14][15] contains the *40 Inventive Principles*. A unique integer value is assigned to each *Inventive Principle*, which is used in the *Contradiction Matrix* to indicate what *Inventive Principle(s)* is (are) to be applied in solving the generic problem at hand. Since the *Contradiction Matrix*'s table is quite complex, only a portion of the table is presented in Table III to illustrate how it is used.

TABLE II. TRIZ 40 INVENTIVE PRINCIPLES [14][15]

| | | | |
|---|---|---|---|
| 1. Segmentation | 11. Cushion in Advance | 21. Rushing through | 31. Use of porous materials |
| 2. Extraction | 12. Equipotentiality | 22. Convert harm into benefit | 32. Changing the color |
| 3. Local Quality | 13. Inversion | 23. Feedback | 33. Homogeneity |
| 4. Asymmetry | 14. Spheroidality | 24. Mediator | 34. Rejecting and regenerating parts |
| 5. Combination | 15. Dynamicity | 25. Self-service | 35. Transforming physical or chemical states |
| 6. Universality | 16. Partial, overdone or excessive action | 26. Copying | 36. Phase transition |
| 7. Nesting | 17. Moving to a new dimension | 27. Inexpensive short life | 37. Thermal expansion |
| 8. Counterweight | 18. Mechanical vibration | 28. Replacement of a mechanical system | 38. Use strong oxidisers |
| 9. Prior Counteraction | 19. Periodic action | 29. Use pneumatic or hydraulic systems | 39. Inert environment |
| 10. Prior Action | 20. Continuity of useful action | 30. Flexible film or thin membranes | 40. Composite materials |

TABLE III. AN EXTRACTION OF THE CONTRADICTION MATRIX [8][15]

| | Worsening Feature/Parameter → / Improving Feature/Parameter ↓ | Weight of Moving Object | Weight of Stationary Object | Length of Moving Object | Length of Stationary Object |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| 1 | Weight of Moving Object | + | - | **15, 8, 29, 34** | |
| 2 | Weight of Stationary Object | - | + | - | **10, 1, 29, 35** |
| 3 | Length of Moving Object | **8, 15, 29, 34** | - | + | - |
| 4 | Length of Stationary Object | | **35, 28, 40, 29** | - | + |

The *Contradiction Matrix* was developed based on thorough studies on roughly 40,000 innovative patents. Based on Table III, if the Improving Feature/Parameter is *Length of Moving Object* (Row 3) and the Worsening Feature/Parameter is *Weight of Moving* Object (Column 1), then the set of *Inventive Principles* to be used is 8, 15, 29, and 34 (refer to Table II for the details). The proposed set of *Inventive Principles* to be used is based on the most probable set of *Inventive Principles* to solve the contradiction.

Another concept in TRIZ that must be emphasized is the *Ideality* of the system [16], which basically means that "the ideal system" that needs to be developed where all of its components perform at the greatest possible capacity. *Ideality* measures how close a system is to the "ideal machine" and it is normally expressed as follows:

$$Ideality = \frac{\sum Benefits}{(\sum Costs + \sum Harms)}$$

All useful functions that result from the system are considered as the system's *benefits*. *Harms* are any unwanted or undesirable outputs of the system including any waste products or side effects produced by the system. One of TRIZ's main objectives is to increase *Ideality* by moving closer towards the ideal final result (IFR), which is considered to be the most optimal situation. And based on the above mathematical expression, IFR can be achieved by increasing the *benefits* of the system, reducing its *harmful* outputs, and also by reducing costs of producing the system towards achieving its *benefits*. Consequently, the end product is automatically the result of an innovative problem-solving approach, which is considered as an invention.

## V. TRIZ APPLICATIONS

TRIZ is normally used to solve engineering-related, technology-related, and scientific problems. In the past many applications in science and technology employed TRIZ in getting results effectively and efficiently with the assistance of the various proven steps used within TRIZ. TRIZ problem-solving methodology is famous for its ability to produce solutions to problems based on past related technologies and at the same time allows users to come up with innovative products really fast. Nowadays, TRIZ has been applied to solve various types of problems ranging from engineering to problems that are not technology-related. Several applications of TRIZ in various problem domains are explained in more detail below just to give some ideas on how TRIZ can be applied in solving inventive or creative problems.

One of the applications of TRIZ is in Computer Aided-Design (CAD). In a paper [17], the authors applied TRIZ and evolutionary algorithms (EA) to solve inventive problems based on dialectical negation. TRIZ and EA are integrated for creating a new conceptual framework that will enhance computer-aided problem solving. The two basic ideas being presented in this paper are "the *inversion* of the traditional EA selection ("survival of the fittest"), and the incorporation of new dialectical negation operators in evolutionary algorithms based on TRIZ principles [17]."

The three laws of dialectics used are given below [18]:

- *The law of the transformation of quantity into quality and vice versa*
- *The law of the interpenetration of opposites*
- *The law of the negation of the negation*

The results showed that TRIZ and evolutionary algorithms (EA) support the idea that inventiveness can be learned, implemented, and developed systematically using some known principles and this approach will save a lot of the inventors' precious time.

Fig. 6 illustrates a *concept map* [17][19] of the inventive problem (IP) solving process performed under the dialectic negation perspective.

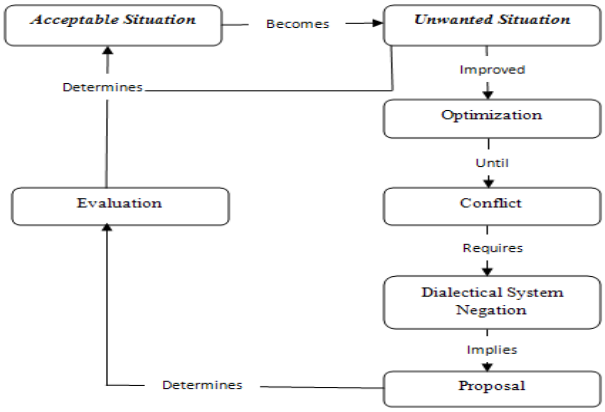



Figure 6. Inventive problem (IP) solving generic stages [17]

The cycle will start again when a new unwanted or undesirable situation reoccurs, and more contradictions will be tackled to make the system become more ideal.

In another paper [20], in an attempt to overcome conflicts between the advancement in technology and the environment, TRIZ was introduced in balancing between technical innovation and its environmental impact. Chang and Chen [20] presented a conflict-problem-solving CAD software that integrated TRIZ into the eco-innovation idea. Design engineers normally use eco-design methods in reducing the product's environmental impact that will occur throughout its life cycle. The harmful impact of the product must be minimized. Using the proposed 'Eco-Design Tool' software, the software engineers can acquire the most feasible or optimal solution efficiently. The five major functions of 'Eco-Design Tool' are as follows [20]:

- *Eco-design targets search by the analytic hierarchy process technique*
- *Product evaluation*
- *TRIZ engineering parameter recommendation reflecting eco-efficiency elements*
- *TRIZ inventive principle exploration by the statistic technology*
- *TRIZ inventive principle interpretation*

Based on the authors' eco-innovative product examples, it was shown that the proposed 'Eco-Design Tool' software was able to assist design engineers, in particular those novices, in producing products that are more environmental-friendly.

TRIZ has also been used in education systems to increase students' problem-solving ability. In a paper [21], at RMIT, a group of forty-two engineering students were enrolled in a course on TRIZ for a duration of 13 weeks, and it was discovered that most of the students were not aware of any other problem-solving methodology or tools before learning TRIZ problem-solving methodology.

As a result of this experiment, the students' perceptions of their capabilities in solving problems have changed significantly, and their thinking abilities have also changed (improved) in such a way that they were able to come up with better ideas that they would never thought of while doing their final project [21]. It was also discovered that the course on TRIZ tools has greater impact on students' problem-solving ability much more than the disciplines based courses [21]. Some of the most significant findings from this research of teaching TRIZ are as follows [21]:

- *Improved ability to attempt open-ended problems*
- *Improved structured/systematic thinking*
- *Able to look beyond the current knowledge*
- *Changed in thinking style*
- *Acquired good problem-solving skills (after completing the course)*

Obviously, from the above results of teaching TRIZ to engineering students, it can be concluded that TRIZ was able to improve the students overall problem-solving ability and also able to increase the students' level of self-confidence in tackling new problems.

In another paper [22], a framework that integrated both the structured and TRIZ methodologies to problem solving was proposed. In this study, the structured methodology was enhanced in such a way that the phases within the structured methodology were slightly modified to consider the integration of the TRIZ approach to problem solving.

This framework considered the various software development methodologies, such as the following [23]: *Waterfall*, *Prototyping*, *Spiral*, *Incremental*, *Rapid Application Development*, *Object-Oriented*, *Extreme Programming*, *Agile*, and many more. Basically, each methodology has the standard software development phases such as *Requirements Specifications*, *Analysis & Design*, *Implementation*, *Testing*, *Verification & Validation*, *Documentation* and *Maintenance*.

Fig. 7 illustrates the TRIZ problem-solving methodology's framework as proposed in [22] for software development and programming problems. The framework was adapted from the structured problem-solving methodology's framework and TRIZ methodology's framework given in [6][8].

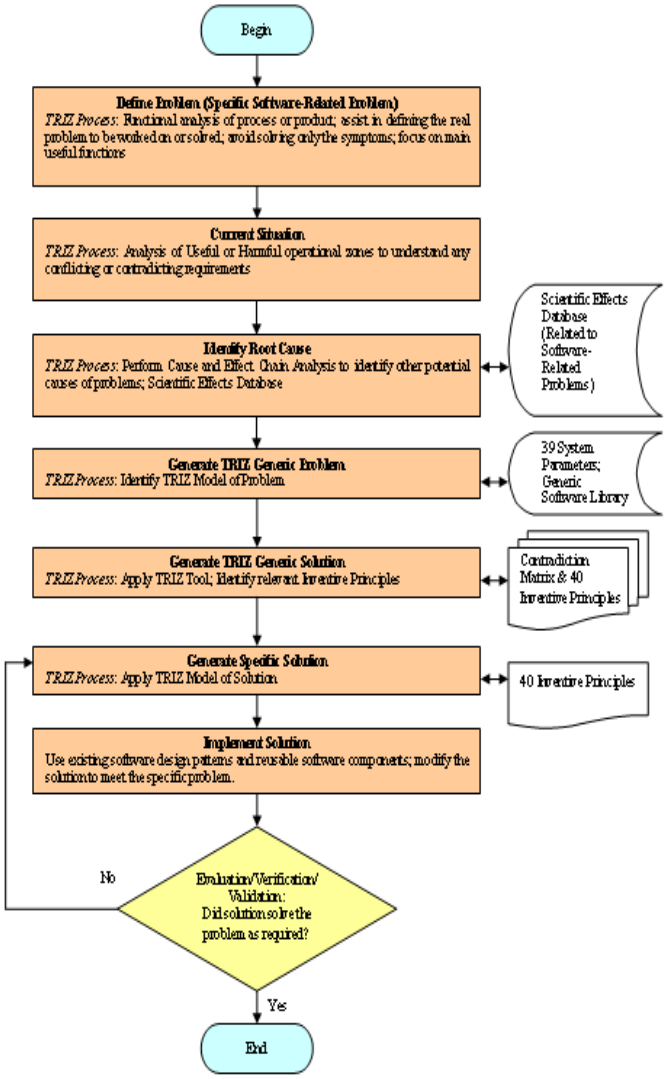



Figure 7. TRIZ problem-solving methodology's framework for software development and programming problems [22]

There are several other applications that used TRIZ in solving inventive problems such as in mobile network industry, control strategies, semiconductor manufacturing, product R & D design, quality assurance, and also in business modeling.

## VI. SUGGESTIONS FOR IMPROVING TRIZ

When TRIZ was first introduced, it was meant to be used to solve engineering-related problems, and in the past many applications had employed TRIZ in getting results effectively and efficiently. TRIZ problem-solving methodology is famous for its ability to produce solutions to problems based on past related technologies and at the same time allows users to come up with innovative products really fast.

But, as described earlier, quite a number of research works also used TRIZ in problem domains or areas that are non-engineering or non-technology-related (non-technical). Based on this new development, the following are suggestions that can further improve the application of TRIZ in wider areas of problem domains, especially those that are not considered as technology-oriented problem areas.

- Add several more innovative steps or processes to the existing general steps of the original TRIZ problem-solving methodology to make the methodology more comprehensive. The steps should be general enough to cater for both technical and non-technical problems.
- Include an intelligent component (with heuristic algorithms) in the overall TRIZ methodology that will try to derive new solutions based on the available knowledge stored in the knowledge base using inferences and reasoning similar to human's ways of reasoning.
- Include a method that allows automatic expansion or growth of the knowledge base used by TRIZ in deriving the solutions.

- Integrate TRIZ with other problem-solving methodologies that have not been considered in past research works.

## VII. FUTURE WORK AND CONCLUSION

Problem solving nowadays is getting more and more challenging because all components or parts that need to be assessed or analyzed are changing rapidly, and there are always possibilities that by the time a solution is formulated for the current state of a system, new problems may arise. So, in the future, an intelligent reasoning and learning component (as suggested earlier) could be embedded within TRIZ so that new solutions that are being devised also consider the most recent developments within similar problem domains and be more pro-active in projecting potential future problems. It is better if TRIZ includes a step that uses artificial intelligence techniques such as case-based reasoning, genetic algorithms, and neural networks in generating several alternatives to the problem at hand, and later the most optimal solution is derived from a pool of possible solutions.

Based on the review on TRIZ and its application performed in this paper, it can be concluded that TRIZ has an enormous influence on the problem-solving and decision-making process. TRIZ problem-solving methodology is still gaining further popularity since its application has been extended beyond its initial problem domain, which is engineering and technology-related. TRIZ is now widely used in education in instilling more systematic problem-solving strategy in students so that they are able to effectively and efficiently solve new problems without depending too much on others.

In conclusion, if we want to solve problems by studying certain patterns (models) in the current problem domain and try to find previously solved solutions that are similar to the conceptual design of the problem at hand, and from there a specific solution can be derived, then TRIZ is the most suitable problem-solving methodology. Most problem-solving methodologies attempt to solve problems by trying to directly find specific solutions to the problems, and this approach is very time-consuming, costly, and also will not guarantee that the specific solution is usable or suitable because it has not been tested in the past. TRIZ on the other hand, uses past conceptual or general solutions to derive a specific solution for the current problem. And, this method of problem solving definitely saves significant amount of time and energy in inventing new products.

## Acknowledgement

This research project was partially funded by the Ministry of Higher Education (MOHE) Malaysia under the Fundamental Research Grant Scheme (FRGS).

## REFERENCES

[1] Absolute Astronomy, Problem solving, Retrieved 8 July, 2013, from http://www.absoluteastronomy.com/topics/Problem_solving.
[2] K. Cherry, What Is Problem-Solving?, About.com Guide, Retrieved 26 June, 2013, from http://psychology.about.com/od/problemsolving/f/problem-solving-steps.htm.
[3] Sixsigmatriz.com, Six sigma TRIZ, Retrieved 17 July, 2013, from http://www.sixsigmatriz.com/.
[4] TRIZ Journal, What is TRIZ?, Retrieved 2 August, 2011, from http://www.triz-journal.com/whatistriz_orig.htm.
[5] V. Fey and E. Rivin, TRIZ: A new approach to innovative engineering & problem solving, Retrieved 11 August, 2011, from http://www.trizgroup.com/articles/TRIZ-ANewApproach.pdf.
[6] Malaysia TRIZ, TRIZ, Retrieved 12 August, 2011, from http://www.mytriz.com.my/home/problemsolving/.
[7] L. Lerner, Genrich Altshuller: Father of TRIZ, Russian Magazine Orgonek, 1991, Retrieved 8 July, 2013, from http://asoft616.securesites.net/articles/altshuller.pdf.
[8] T.S. Yeoh , T.J. Yeoh Tay Jin, and C.L. Song, Theory of Inventive Problem Solving TRIZ – Systematic Innovation in Manufacturing, (Firstfruits Publishing, 2009).
[9] D. Mann, Hands-on Systematic Innovation, (CREAX Press, Ieper, Belgium, 2002), Cited in Yeoh Teong San, Yeoh Tay Jin, and Song Chia Li.
[10] J.L. Moya, A.S. Machado, Reiner Robaina, J.A. Velázquez, R Mestizo, J. A. Cárdenas, R. A. Goytisolo, Application of TRIZ principles to spur gear design, Retrieved 27 August, 2011, from http://www.trizsite.com/articles/aug2010/Application%20of%20TRIZ%20principles%20to%20gear%20design.pdf.
[11] W.C. Ames, ADMINISTRACIÓN TRIZ, la herramienta del pensamiento e innovación sistemática, 2008, Cited in J.L. Moya et al., Retrieved 27 August, 2011, from http://www.trizsite.com/articles/aug2010/Application%20of%20TRIZ%20principles%20to%20gear%20design.pdf.
[12] GEN3, GEN3 Training Manual, 2006, Cited in Yeoh Teong San, Yeoh Tay Jin, and Song Chia Li.
[13] Trizsite.com, 39 contradiction parameters, Retrieved 27 August, 2011, from http://www.trizsite.com/triztools/parameters.asp.
[14] M.C. Maldonado, Innovación sistemática mediante TRIZ, 2005, Cited in J.L. Moya et al., Retrieved 27 August 2011, Available: http://www.trizsite.com/articles/aug2010/Application%20of%20TRIZ%20principles%20to%20gear%20design.pdf.
[15] Trizsite.com, 40 inventive principles, Retrieved 27 August, 2011, from http://www.trizsite.com/triztools/principles.asp.
[16] G. Altshuller, The Innovation Algorithm: TRIZ, systematic innovation, and Technical Creativity. Technical Innovation Centre, Originally published in Russian 1969 and 1973, Worcester, Massachusetts, 1999.
[17] R. Duran-Novoa, N. Leon-Rovira, H. Aguayo-Tellez, D. Said, Inventive problem solving based on dialectical negation, using evolutionary algorithms and TRIZ heuristics, Computers in Industry, Volume 62, Issue 4, May 2011, Pages 437-445,
[18] F. Engels, C. Dutt, J.B. Haldane, Dialectics of Nature, Progress Publishers, 1964, cited in Ref_18 (Roberto Duran-Novoa)
[19] Novak, J. D. & A. J. Cañas, The Theory Underlying Concept Maps and How to Construct and Use Them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008, Available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf

[20] H.T. Chang, J.L. Chen, The conflict-problem-solving CAD software integrating TRIZ into eco-innovation, Advances in Engineering Software, 35(8-9), 2004, 553-566.
[21] I. Belski, TRIZ course enhances thinking and problem solving skills of engineering students, Procedia Engineering, 9, 2011, 450-460.
[22] H.M. Jani, Improving Software Development and Programming Effectiveness through TRIZ Problem-Solving Methodology, In Proc. ICACT2011: The 2nd International Conference on Advancements in Computing Technology, Korea, 2011, 1089-1093.
[23] Centers for Medicare & Medicaid Services (CMS) Office of Information Service, Selecting a development approach, Webarticle, United States Department of Health and Human Services (HHS), 2008, Retrieved 1 September, 2011, from https://www.cms.gov/SystemLifecycleFramework/Downloads/SelectingDevelopmentApproach.pdf.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 93-96*
*www.iosrjournals.org*

# Cloud Information Accountability Frameworks for Data Sharing in Cloud

[1] C. Madhuri, [2] A. Krishna chaitanya
[1]*CSE Dept, Vardhaman College of Engineering, Hyderabad, India,*
[2]*IT Dept, VCE, Hyderabad, India,*

***Abstract:*** *The ability to hold individuals or organizations accountable for transactions is important in most of the commercial and legal or e-transactions. Cloud computing is a new term that is introduced in business environment where users can interact directly with the virtualized resources and safe the cost for the consumers. Where user have everything to be deal on internet on e-transaction to save work and time but we fear for our confidential and individual data is in safe or not, is it been hacked and misused is a fear that stops to use of modern technology. Moreover, users may not know the machines which actually process and host own data. While enjoying the convenience brought by this new technology, end users also start worrying about security of their own personal and important data. Accountability traces every important aspect of any data sharing on data usage in cloud where it accounts every action in the system can be traced back to some entity with assuring the safety and security including the handling of personally identifiable information. In this paper we review the cloud information accountability framework for the data sharing in which procedural and technical solutions are co-designed to demonstrate accountability by the various researchers to resolving privacy and security risks within the cloud and presents a review on new way to supplement the current consumption and delivery model for IT services based on the Internet, by providing for dynamically scalable framework.*
***Keywords****: e-transactions, Accountability, TPA, Tracing Authority.*

## I. Introduction

Cloud computing is the access to computers and their functionality via the Internet or a local area network where clients request clouds access from a set of web services that manage a pool of computing resources (i.e. machines, network, storage, operating systems, application development environments, application programs). Requests are dedicated to user until he or she releases them. Cloud computing presents a new way to supplement the current consumption and delivery model for IT services based on the Internet, by providing for dynamically scalable and often virtualized resources as a service by cloud. Today, there are a number of notable commercial and individual cloud computing services, including Amazon, Google, Microsoft, Yahoo, and Salesforce etc. Users may not know the machines which actually process and host their data. While enjoying the convenience brought by this new technology, they also start worrying about losing control of their own data. The data processed on clouds are often outsourced, leading to many issues related to accountability, including the handling of personally identifiable information and these fears are becoming a significant obstacle to the wide acceptance of cloud services.

Accountability [2] is the obligation to act as a responsible for preserving the personal information of others and appropriate use of that information beyond mere legal requirements, and to be accountable for any misuse of that personal information. Accountability places a legal responsibility on an organization to guarantee that the contracted partners to whom it supplies data are compliant and privacy. We also develop two distinct modes for auditing: push mode and pull mode. The push mode refers to logs being periodically sent to the data owner or stakeholder while the pull mode refers to an alternative approach whereby the user can retrieve the logs as needed.

A Cloud Information Accountability (CIA) framework [1], based on the notion of information accountability which focuses on keeping the data usage transparent and trackable.CIA framework provides end-to end accountability in a highly distributed fashion that influence and expand the programmable capability of JAR (Java Archives) files to automatically log the usage of the users' data by any entity in the cloud. Users will send their data along with any policies such as access control policies and logging policies that they want to enforce, enclosed in JAR files, to cloud service providers. Any access to the data will trigger an automated and authenticated logging mechanism local to the JARs.

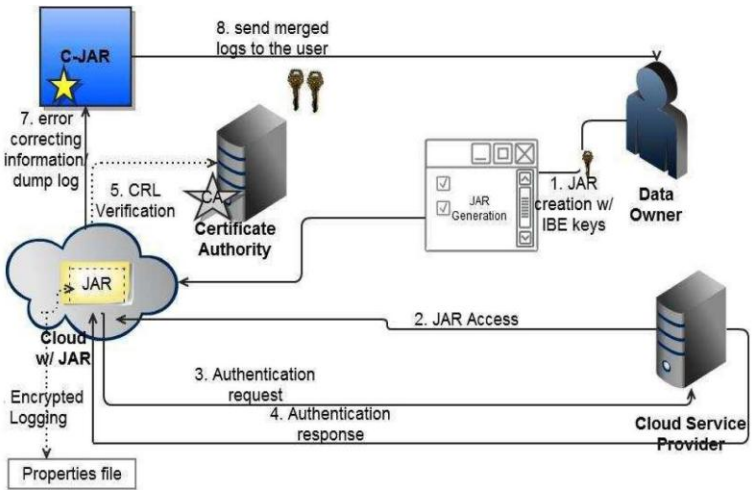



Fig. 1: A framework for cloud computing accountability

Cloud computing has raised a range of important privacy and security issues. The basic idea is that the user's private data are sent to the cloud in an encrypted form, and the processing is done on the encrypted data. The work of A. Squicciarini, S. Sundareswaran, and D. Lin [5] a Java-based approach is provided to prevent privacy leakage from indexing, which could be integrated with the CIA framework proposed in this work since they build on related architectures. We do not rely on IBE to bind the content with the rules. Instead, we use it to provide strong guarantees for the encrypted content and the log files, such as protection against chosen plaintext and cipher text attacks. To its accountability plays a major role in electronic transaction that takes place in reality via internet in our day to day life is explained in his paper with multi party transactions and users confidential data [6].

## II. Cloud Scenario

We have a cloud computing scenario as a significance context to describe Electronic Data Sharing Agreements i.e. e-DSAs related policy definitions and their enforcement [3]. The e-DSAs and their automated enforcement are keys to further enable business interactions and information flows within the Cloud, by providing more assurance and control on data where multiple Cloud Service Provider (CSP) available in the Internet. A customer services are supplied by a specific CSP to access online travelling, printing, office applications, etc. To require their access services, customers need to register and disclose personal data, inclusive of address, financial details, etc and to provide the required functionalities, a CSP might need to interact with other Service Providers and share relevant data to enable the business transaction. For example, e-banking services using credit cards buying online may require users account details to be disclosed , e-transactions in order to supply the required service to the customers where it can potentially be analyzed, processed and exchange the information between parties.[7]

The key issue is that both the customer and service providers may lose control on data when this data is exchanged between different parties in chains of interactions. Customers might desire to know control details about: How their data should be used; who can access it, etc. (i.e. accountability) ;The purposes for which data can be disclosed to third parties.(i.e. sharing information to other organization);Impose constraints on the retention time, notifications, etc. Similar comments apply to a service provider disclosing information to third parties.

Including privacy preferences on how their personal and confidential data should be handled along with access control and obligation constraints for examples of authorization policies for access control and obligation policies like **Authorization Policies and obligation policies**

- Data of my credit card can be accessed by Service Provider 1(SP1) only for Business Transaction purpose.
- My email address can be shared with SP2 and SP3 only for business transaction and goods delivery purpose (For businesses: defined legal environment, allowing risk assessments. For individuals: maintenance of societal rights, privacy, and right to time and memory loss but as consumers: defined legal environment Multi party security requirements.)
- My email address details must not be shared with SP4.

The obligation policies on the users data would be like where I want to be notified by email every time my data is accessed; I want to be notified every time my credit card is disclosed to another Service Provider; I want my data to be deleted after 1 year if not accessed/used.

Interestingly, the stated constraints might need to be enforced by all the entities involved in a chain of data disclosures, e.g. in the example, by the banking Service, the Printing Service, the Flight Booking Service, etc where the customer might change their mind and modify some of their preferences and constraints. These changes should be passing through the chain of disclosures as well. Security in cloud computing consist of security abilities of web browsers and web service structure.

## III. Related Work On Information Security

In this section we try to highlight the framework suggested by of Marco Casassa Mont, Siani Pearson, Pete Bram hall discussed some problems related with the personal information security. [3] In order to discuss the involved problem, we refer to an e-commerce scenario. By providing damage recovery options mainly: contracts, legal entities, activity logs, defined and agreed transactions .we initially provides personal digital identity and profile information to an e-commerce site in order to access their services, possibly after negotiations about which privacy policies need to be applied . Then the user logs in and interacts with these services: it might happen that in so doing he/she needs to involve other web sites or organizations. The user might be conscious of this or this might happen behind the scenes, for example due to the fact that the e-commerce site interacts with partners and suppliers. The e-commerce site might need to disclose personal data to third parties (such as suppliers, information providers, government institutions etc.) in order to fulfill the specific transaction. This involved e-commerce sites do not necessarily have prior agreements or belong to the same web of trust. Such scenario highlights a few key issues: how to fulfill users' privacy rights and make users be in control of their information. At the same time users' interactions need to be simple and intuitive Privacy and data protection laws that regulate this area do exist but it is hard to monitor them, especially when private information spread across organizations and nations' boundaries. In addition, further complexity arises due to the fact that privacy laws can differ quite substantially depending on national and geographical aspects. For example in US privacy laws restrict what the government can do with personal data but they introduce few restrictions on trading of personally identifiable information private enterprises. [7]

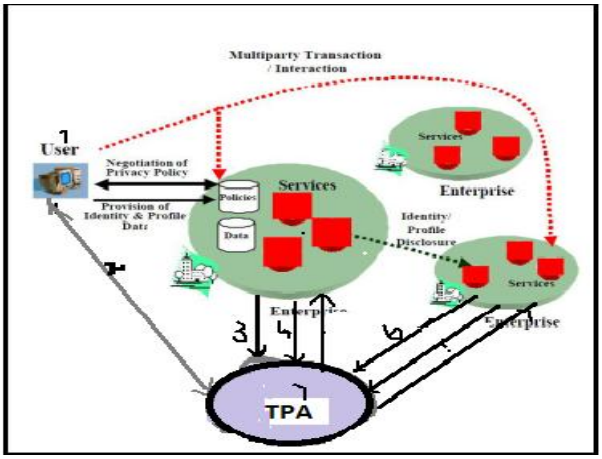



Fig 2 A scenario where users deal with e-transactions that span across multiple ecommerce sites. In Europe (EU) people can consent to have their personally identifiable information used for commercial purposes but the default is to protect that information and not allow it to be used indiscriminately for marketing purposes.

In this model people Graphical tools (1) locally author their disclosure policies (i.e. sticky polices) in a fine-grained way; Obfuscate their confidential data by directly using these disclosure polices; · Associate these policies to the obfuscated data. Some of the above activities can be automated by using predefined policy templates and scripts. Digital packages: (2) containing obfuscated data along with their sticky polices can be provided to requestors such as ecommerce sites. These digital packages might contain a superset of the required information, Selective disclosure of (part of) their contents will be authorized, depending on needs. A requestor (3) has to demonstrate to the Tracing Authority that he/she understands the involved terms and conditions. A Tracing Authority checks trustworthiness of the requestor's credentials and their IT environment (4) accordingly to the disclosure policies. The owner of the confidential information can be actively involved in the disclosure process (5) by asking for his authorizations or by notifications, according to the agreed disclosure policies. The actual disclosure (6) of any obfuscated data to a requestor (for example the e-commerce site) only happens after the requestor demonstrates to a trusted third party – i.e. the "Tracing Authority" - that it can satisfy the associated sticky policies[3]. Disclosures of confidential data are logged and audited by the Tracing Authority (7).In our model nothing prevents the owner of the confidential information from running a Tracing Authority. (8)This increases the accountability of the requestors by creating evidence about their knowledge of users'

confidential data. In particular this applies when confidential information is in discriminately disclosed to third parties, as this evidence can be used for forensic analysis. In case a requestor sends the obfuscated data package to a third party the same process, described above, applies. Multiple trusted third parties can be used in the above process in order to minimize the risks involved in the management of trust, for example having to rely only on one entity. Once the authentication succeeds, the service provider (or the user) will be allowed to access the data enclosed in the JAR. The accountability in distributed data sharing mechanism with auditing modes and logging mechanism as referred in [6].

## IV. Conclusion

It is more and more important to defend and preserve people's privacy on the Internet, against unwanted and unauthorized disclosure of their confidential data. Throughout this paper, the authors have systematically studied and review the security and privacy issues in cloud computing. We propose a novel highly decentralized information accountability framework to keep track of the actual usage of the users' data in the cloud. In particular, we propose an object centered approach that enables enclosing our logging mechanism together with users' data and policies. We have identified the most representative security/privacy attributes (e.g., confidentiality, integrity, availability, accountability, and privacy-preservability). Cloud computing is a new term that is introduced in business environment where users can interact directly with the virtualized resources and safe the cost for the consumers. It has model to protect its data for the business users. An organization used private clouds within its organization to prevent from loss of data.

## References

[1] S. Sundareswaran, A. Squicciarini, D. Lin, " Distributed Accountability for Data Sharing in the Cloud," *Proc. IEEE Transactions on Dependable and Secure Computing, Vol. 9, No.4,* Aug. 2012.
[2] S. Pearson, "Towards Accountability in the Cloud ," *Proc. IEEE Internet Computing,* pp. 64-69, 2011
[3] Mr. Marco Casassa Mont, Siani Pearson, Pete Bramhall , Towards Accountable Management of Identity and Privacy: Sticky Policies and Enforceable Tracing Services HPL-2003-49 March 19th , 2003.
[4] Siani Pearson and Andrew Charles worth " Accountability as a Way Forward for Privacy Protection in the Cloud".
[5] A. Squicciarini, S. Sundareswaran, and D. Lin, "Preventing Information Leakage from Indexing in the Cloud," *Proc. IEEE Int'l Conf. Cloud Computing,* 2010.
[6] S. Sundareswaran, A. Squicciarini, D. Lin, and S. Huang, "Promoting Distributed Accountability in the Cloud," *Proc. IEEE Int'l Conf. Cloud Computing,* 2011.
[7] Shraddha B. Toney, Sandeep U.Kadam "Cloud Information Accountability Frameworks for Data Sharing in Cloud." 13th May 2012.

*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 13, Issue 2 (Jul. - Aug. 2013), PP 97-110*
*www.iosrjournals.org*

# Medical Image Processing in Nuclear Medicine and Bone Arthroplasty

Joyjit Patra[1], Himadri Nath Moulick[2], Arun Kanti Manna[3]
[1,2](C.S.E,.Aryabhatta Institute Of Engineering And Management,Durgapur,West Bengal,India)
[3](C.S.E,Modern Institute of Engineering And Technology,Hoogly,West Bengal,India)

***Abstract:*** *Computers have become indispensable in all domains, and the medical segment does not represent an exception. The need for accuracy and speed has led to a tight collaboration between machines and human beings.Maybe the future will allow the existence of a world where the human intervention won't be necessary, but for now,the best approach in the medical field is to create semiautomatic applications, in order to help the doctors with the diagnoses, with following the patients' evolution and managing them and with other medical activities. Our application is designed for automatic measurements of orthopedic parameters, and allows the possibility of human intervention in case the parameters have not been detected properly. The segment of the application is Hip Arthroplasty.*

***Keywords:*** *Hip Arthroplasty, Canny Edge Detection, DICOM, Hough Transform, Radiographic Image Processing.*

## I. INTRODUCTION

Medical image processing is an area of increasing interest. It includes a wide range of methods and techniques, starting with the acquisition of images using specialized devices (for example, CT devices), image enhancement and analysis, to 3D model reconstruction from 2D images. Thus, the research in this field represents a point of interest for both doctors and engineers, in their attempt to improve medical techniques, with computer assistance, in order to obtain more accurate results in treating the patients.Among many research projects in this area of interest some of the most relevant are: - The SCANIP [11] image processing software that provides a broad range of image visualization, processing and segmentation tools for medical purposes. This software has been created by Simpleware in Great Britain. The SCANIP programs ensure the conversion of 3D medical images into quality meshes. These meshes can be used in future processing in analysis programs, in fluid dynamics, CAD and the creation of Rapid Prototyping models. The sources for these programs come from MRIs, CTs or MicroCTs.

The 3D-DOCTOR Project [12] that comes with an advanced 3d modeling software, with strong processing and measurement functions for MRI-s, CT-s, PET-s, and other types of medical images.The possible applications of this software are in the scientific and medical domain, but also in the image processing industrial field. The functioning principle of this software is based on edge detection techniques using 3D image segmentation functions and on the construction of 3D surfaces and volumes, that are afterwards visualized and measured for the purpose of a quantitative and qualitative analysis. - The Hip-OpCT software [13] that allows importing CT images in DICOM format. Once imported, theCT dataset is visualized through several modalities from which the doctors can plan the size and the position of the prosthesis. Advanced techniques of image processing and analysis find widespread use in medicine. In medical applications, image data are used to gather details regarding the process of patient imaging whether it is a disease process or a physiological process.

Information provided by medical images has become a vital part of today's patient care. The images generated in medical applications are complex and vary notably from application to application. Nuclear medicine images show characteristic information about the physiological properties of the structures-organs. In order to have high quality medical images for reliable diagnosis, the processing of image is necessary. The scope of image processing and analysis applied to medical applications is to improve the quality of the acquired image and extract quantitative information from medical image data in an efficient and accurate way.MatLab (**Mat**rix **Lab**oratory) is a high performance interactive software package for scientific and engineering computation developed by MathWorks (Mathworks Inc., 2009).MatLab allows matrix computation, implementation of algorithms, simulation, plotting of functions and data, signal and image processing by the Image Processing Toolbox. It enables quantitative analysis and visualisation of nuclear medical images of several modalities, suchas Single Photon Emission Computed Tomography (SPECT), Positron Emission Tomography (PET) or a hybrid system (SPECT/CT) where a Computed Tomography system (CT) is incorporated to the SPECT system. The Image Processing Toolbox (Mathworks Inc., 2009) is a comprehensive set of reference-standard algorithms and graphical tools for image processing, analysis, visualisation and algorithm development. It offers the possibility to restore noisy or degraded images, enhance images for

improved intelligibility, extract features, analyse shapes and textures, and register two images. Thus, it includes all the functions that MatLab utilises in order to perform any sophisticated analysis needed after the acquisition of an image. Most toolbox functions are written in open MatLab language offering the opportunity to the user to inspect the algorithms, to modify the source code and create custom functions (Wilson et al., 2003, Perutka, 2010).This chapter emphasises on the utility of MatLab in nuclear medicine images' processing. It includes theoretical background as well as examples. After an introduction to the imaging techniques in nuclear medicine and the quality of nuclear medicine images, this chapter proceeds to a study about image processing in nuclear medicine through MatLab. Image processing techniques presented in this chapter include organ contouring, interpolation, filtering, segmentation, background activity removal, registration and volume quantification. A section about DICOM image data processing using MatLab is also presented as this type of image is widely used in nuclear medicine.

## II. NUCLEAR MEDICINE IMAGING

Nuclear Medicine is the section of science that utilises the properties of radiopharmaceuticals in order to derive clinical information of the human physiology and biochemistry. According to the examination needed for each patient, a radionuclide is attached to a pharmaceutical (tracer) and the whole complex is then delivered to the patient intravenously or by swallowing or even by inhalation. The radiopharmaceutical follows its physiological pathway and it is concentrated on specific organs and tissues for short periods of time. Then, the patient is positioned under a nuclear medicine equipment which can detect the radiation emitted by the human body resulting in images of the biodistribution of the radiopharmaceutical. In Nuclear Medicine, there are two main methods of patient imaging, the imaging with Planar Imaging, Dynamic Imaging or SPECT and the PET.

During the last decade, hybrid systems have been developed integrating the CT technique with either SPECT or PET resulting in SPECT/CT and PET/CT respectively. This chapter will concentrate on the implementation of MatLab code in gamma camera planar imaging, SPECT and SPECT/CT methods. The gamma camera is composed of a collimator, a scintillator crystal usually made of NaI (or CsI), the photomultiplier tubes, the electronic circuits and a computer equipped with the suitable software to depict the nuclear medicine examinations. In planar imaging, the patient, having being delivered with the suitable radiopharmaceutical, is sited under the gamma camera head. The gamma camera head remains stable at a fixed position over the patient for a certain period of time, acquiring counts (disintegrations). These will constitute the radiopharmaceutical distribution image. The counts measured in a specific planar projection originate from the whole thickness of patient (Wernick & Aarsvold, 2004). In SPECT, the gamma camera head rotates around the patient remaining at well defined angles and acquiring counts for specific periods of time per angle. What makes SPECT a valuable tool in nuclear medicine is the fact that information in the three dimensions of the patient can be collected in a number of slices with a finite known volume (in voxels). Thus, SPECT technique is used to display the radiopharmaceutical distribution in a single slice removing the contribution from the overlying and underlying tissues. In order to obtain the most accurate quantitative data from SPECT images, two issues that have to be resolved are the attenuation correction and the Compton scattering that the photons are undergone until reach and interact with the slice of interest tissues. As an examining organ has certain dimensions, each slice along the axis of the gamma camera has different distance from the detector. Thus, each photon experiences different attenuation. These two phenomena usually lead to distortion of the measured activity concentration (Wernick & Aarsvold, 2004). The acquired data are processed in order to correct and compensate the undesired effect of these physical phenomena. The projection data of each slice constitute the sinogram. As a result, a series of sinograms is the files acquired. However, this kind of files needs reconstruction in order to get an image with diagnostic value. The most known reconstruction methods are the Filtered Back-Projection (FBP) and the Iterative methods. Attenuation correction is resolved by using the constant linear attenuation coefficient (μ) method or using the transmission source method. In the first one, the distance that each photon has travelled is calculated based on the patient geometry and the exponential reduction of their intensity. Then, considering the human body as a uniform object, an attenuation map is implemented in the reconstructed image. The latter method utilises a transmission source which scans the patient. This depicts each pixel or voxel of the patient with a specific μ producing an attenuation coefficient map. Finally, the attenuation map isimplemented on the image resulting in a more accurate diagnosis. The second issue of scatter correction can be resolved by the electronics of the gamma camera and the filtering process during reconstruction. When a photon undergoes scattering, its energy reduces. So, a well defined function can accept for imaging photons with energy at a certain narrow energy window around the central photopeak of the γ- emission. A hybrid SPECT/CT scanner is capable of implementing both a CT scan and a SPECT scan or it can be used for each of these scans separately.

Using the CT scan, the anatomy of a specific patient area can be imaged while the SPECT scan can depict the physiology of this area. Then, the registration of the two images drives at an image of advanced diagnostic value. Moreover, the CT data is used for the implementation of attenuation correction. (Delbeke et al., 2006) The range of nuclear medicine examinations is fairly wide. It includes, among others, patients'

studies, as myocardium perfusion by 99mTc-Tetrofosmin or 99mTc-Sestamibi, striatum imaging in brain by 123I-Ioflupane (DaTSCAN), renal parenchyma imaging by 99mTc-De-Methylo-Sulfo-Acid (DMSA) and 99mTc-Methylo-Di-Phosphonate (MDP) for bone scintigraphy. Fundamental image analysis methods of myocardium, brain, kidneys, thyroid, lungs and oncological (e.g. neuroblastoma) nuclear medicine studies include regions' properties, boundary analysis, curvature analysis or line and circle detection. Image processing serves in reconstruction of images acquired using SPECT techniques, in improvement of the quality of images for viewing and in preparation of images for quantitative results. Data of the mentioned examinations are used in the following applications of MatLab algorithms to make the image processing and analysis in nuclear medicine clear and show the MatLab utility for these studies.

## III. IMAGE ANALYSIS AND PROCESSING IN NUCLEAR MEDICINE

In the last several decades, medical imaging systems have advanced in a dynamic progress. There have been substantial improvements in characteristics such as sensitivity, resolution, and acquisition speed. New techniques have been introduced and, more specifically, analogue images have been substituted by digital ones.

As a result, issues related to the digital images' quality have emerged. The quality of acquired images is degraded by both physical factors, such as Compton scattering and photon attenuation, and system parameters, such as intrinsic and extrinsic spatial resolution of the gamma camera system. These factors result in blurred and noisy images. Most times, the blurred images present artefacts that may lead to a fault diagnosis.

In order the images to gain a diagnostic value for the physician, it is compulsory to follow a specific series of processing. Image processing is a set of techniques in which the data from an image are analysed and processed using algorithms and tools to enhance certain image information that is more useful to human interpretation (Nailon, 2010). The processing of an image permits the extraction of useful parameters and increases the possibility of detection of small lesions more accurately. Image processing in nuclear medicine serves three major purposes: a) the reconstruction of the images acquired with tomographic (SPECT) techniques.

b) the quality improvement of the image for viewing in terms of contrast, uniformity and spatial resolution and,

c) the preparation of the image in order to extract useful diagnostic qualitative and quantitative information.

### ❖ Digital images

In all modern nuclear medicine imaging systems, the images are displayed as an array of discrete picture elements (pixels) in two dimensions (2D) and are referred as digital images. Each pixel in a digital image has an intensity value and a location address (Fig. 1). In a nuclear medicine image the pixel value shows the number of counts recorded in it. The benefit of a digital image compared to the analogue one is that data from a digital image are available for further computer processing.

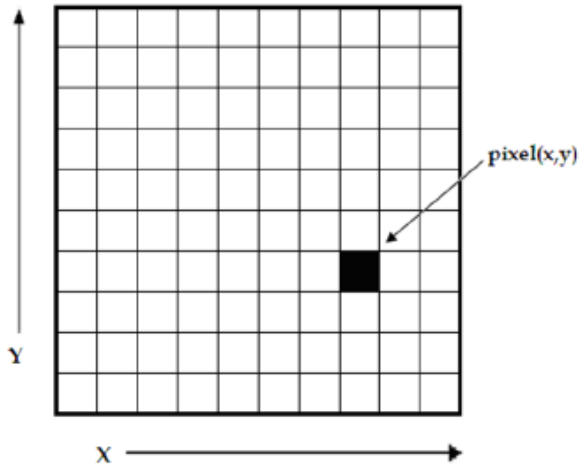



Fig. 1. A digital image is a 2D array of pixels. Each pixel is characterised by its (x, y) coordinates and its value.

Digital images are characterised by matrix size, pixel depth and resolution. The matrix size is determined from the number of the columns (m) and the number of rows (n) of the image matrix (m×n). The size of a matrix is selected by the operator. Generally, as the matrix dimension increases the resolution is getting better (Gonzalez et al., 2009). Nuclear medicine images matrices are, nowadays, ranged from 64×64 to 1024×1024 pixels. Pixel or bit depth refers to the number of bits per pixel that represent the colour levels of each pixel in an image. Each pixel can take 2k different values, where k is the bit depth of the image. This

means that for an 8-bit image, each pixel can have from 1 to 28 (=256) different colour levels (grey-scale levels). Nuclear medicine images are frequently represented as 8- or 16- bit images. The term resolution of the image refers to the number of pixels per unit length of the image.

In digital images the spatial resolution depends on pixel size. The pixel size is calculated by the Field of View (FoV) divided by the number of pixels across the matrix. For a standard FoV, an increase of the matrix size decreases the pixel size and the ability to see details is improved.

❖ **Types of digital images – MatLab**

MatLab offers simple functions that can read images of many file formats and supports a number of colour maps. Depending on file type and colour space, the returned matrix is either a 2D matrix of intensity values (greyscale images) or a 3D matrix of RGB values. Nuclear medicine images are grey scale or true colour images (RGB that is Red, Green and Blue).

The image types supported from the Image Processing Toolbox are listed below:

• Binary Images. In these, pixels can only take 0 or 1 value, black or white.

• Greyscale or intensity images. The image data in a greyscale image represent intensity or brightness. The integers' value is within the range of [0… 2k-1], where k is the bit depth of the image. For a typical greyscale image each pixel can represented by 8 bits and intensity values are in the range of [0…255], where 0 corresponds to black and 255
to white.

• True color or RGB. In these, an image can be displayed using three matrices, each one corresponding to each of red-green-blue colour. If in an RGB image each component uses 8 bits, then the total number of bits required for each pixel is 3×8=24 and the range of each individual colour component is [0…255].

• Indexed images. Indexed images consist of a 2D matrix together with an m×3 colour map (m= the number of the columns in image matrix). Each row of map specifies the red, green, and blue components of a single colour.

An indexed image uses direct mapping of pixel values to colour map values. The colour of each image pixel is determined by using the corresponding value of matrix as an index into map.The greyscale image is the most convenient and preferable type utilised in nuclear medicine image processing. When colouring depiction is needed, the RGB one should be used and processed. The indexed type images should be converted to any of the two other types in order to be processed. The functions used for image type conversion are: rgb2gray, ind2rgb, ind2gray and reversely. Any image can be also transformed to binary one using the command: im2bw.

Moreover, in any image, the function impixelinfo can be used in order to detect any pixel value. The user can move the mouse cursor inside the image and the down left corner appears the pixel identity (x, y) as well as the (RGB) values. The pixel range of the image can be displayed by the command imdisplayrange.

❖ **Image processing techniques - MatLab**

Image processing techniques include all the possible tools used to change or analyse an image according to individuals' needs. This subchapter presents the most widely performed image processing techniques that are applicable to nuclear medicine images. The examples used are mostly come from nuclear medicine renal studies, as kidneys' planar images and SPECT slices are simple objects to show the application of image processing MatLab tools.

❖ **Contrast enhancement**

One of the very first image processing issues is the contrast enhancement. The acquired image does not usually present the desired object contrast. The improvement of contrast is absolutely needed as the organ shape, boundaries and internal functionality can be better depicted. In addition, organ delineation can be achieved in many cases without removing the background activity. The command that implements contrast processing is the imadjust. Using this, the contrast in an image can be enhanced or degraded if needed. Moreover, a very useful result can be the inversion of colours, especially in greyscale images, where an object of interest can be efficiently outlined. The general function that implements contrast enhancement is the following:

J = imadjust(I,[low_in high_in],[low_out high_out],gamma);

while the function for colour inversion is the following:

J = imadjust(I,[0 1],[1 0],gamma); or J = imcomplement(I);

suppose that J, is the new image, I, is the initial image and gamma factor depicts the shape of the curve that describes the relationship between the values of I and J. If the gamma factor is omitted, it is considered to be 1.

❖ **Organ contour**

In many nuclear medicine images, the organs' boundaries are presented unclear due to low resolution or presence of high percentage of noise. In order to draw the contour of an organ in a nuclear medicine image,

the command imcontour is used. In addition, a variable n defines the number of equally spaced contoursrequired. This variable is strongly related with the intensity of counts. For higher n values, the lines are drawn with smaller spaces in between and depict different streaks of intensity. The type of line contouring can be specified as well. For example, when a contour of 5 level contours, drawn with solid line, is the desirable outcome, the whole function is:

```
Example 1I = imread('kindeys.jpg');
figure, imshow(I)
J = imcontour(I,5,'-');
Figure, imshow(J)
```

Where J and I stands for the final and the initial image respectively and the symbol ('-') stands for the solid line drawing. An example of the initial image, the contour with n=15 and n=5 respectively, follows.

Fig. 2. (a) Original image depicting kidneys, (b) organs contoured with n = 15, (c) organs contoured with n = 5.

## IV. IMAGE INTERPOLATION

Interpolation is a topic that has been widely used in image processing. It constitutes of the most common procedure in order to resample an image, to generate a new image based on the pattern of an existing one. Moreover, re-sampling is usually required in medical image processing in order to enhance the image quality or to retrieve lost information after compression of an image (Lehmann et al., 1999). Interpreting the interpolation process, the user is provided with several options. These options include the resizing of an image according to a defined scaling factor, the choice of the interpolation type and the choice of low-pass filter. The general command that performs image resizing is imresize. However, the way that the whole function has to be written depends heavily on the characteristics of the new image. The size of the image can be defined as a scaling factor of the existing image or by exact number of pixels in rows and columns. Concerning the interpolation types usually used in nuclear medicine, these are the following: a) nearest-neighbour interpolation ('nearest'), where the output pixel obtains the value of the pixel that the point falls within, without considering other pixels, b) bilinear interpolation ('bilinear'), where the output pixel obtains a weighted average value of the nearest 2x2 pixels, c) cubic interpolation ('bicubic'), where the output pixel obtains a weighted average value of the nearest 4x4 pixels (Lehmann et al., 1999).When an image has to resize in a new one, with specified scaling factor and method, then the function Implementing that, is the following:

```
NewImage = imresize(Image, scale, method);
```

For example, for a given image I, the new image J shrunk twice of the initial one, using the bilinear interpolation method, the function will be: J = imresize(I, 0.5, 'bilinear');

This way of image resizing contributes to the conversion of image information during any such process, a fact that is valuable in the precision of a measurement. Bilinear interpolation is often used to zoom into a 2D image or for rendering, for display purposes. Apart from the previous methods, the cubic convolution method can be applied to 3D images.

## V. IMAGE FILTERING

The factors that degrade the quality of nuclear medicine images result in blurred and noisy images with poor resolution. One of the most important factors that greatly affect the quality of clinical nuclear medicine images is image filtering. Image filtering is a mathematical processing for noise removal and resolution recovery. The goal of the filtering is to compensate for loss of detail in an image while reducing noise. Filters suppressed noise

as well as deblurred and sharpened the image. In this way, filters can greatly improve the image resolution and limit the degradation of the image. An image can be filtered either in the frequency or in the spatial domain. In the first case the initial data is Fourier transformed, multiplied with the appropriate filter and then taking the inverse Fourier transform, re-transformed into the spatial domain. The basics steps of filtering in the frequency domain are illustrated in Fig. 3.

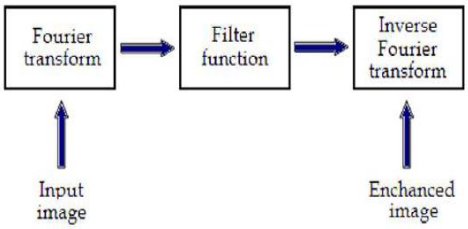



Fig. 3. Basics steps of frequency domain filtering.

The filtering in the spatial domain demands a filter mask (it is also referred as kernel or convolution filter). The filter mask is a matrix of odd usually size which is applied directly on the original data of the image.

The mask is centred on each pixel of the initial image. For each position of the mask the pixel values of the image is multiplied by the corresponding values of the mask. The products of these multiplications are then added and the value of the central pixel of the original image is replaced by the sum. This must be repeated for every pixel in the image. The procedure is described schematically in Fig. 3. If the filter, by which the new pixel value was calculated, is a linear function of the entire pixel values in the filter mask (e.g. the sum of products), then the filter is called linear. If the output pixel is not a linear weighted combination of the input pixel of the image then the filtered is called non-linear. According to the range of frequencies they allow to pass through filters can be classified as low pass or high pass. Low pass filters allow the low frequencies to be retained unaltered and block the high frequencies. Low pass filtering removes noise and smooth the image but at the same time blur the image as it does not preserve the edges. High pass filters sharpness the edges of the image (areas in an image where the signal changes rapidly) and enhance object edge information. A severe disadvantage of high pass filtering is the amplification of statistical noise present in the measured counts. The next section is referred to three of the most common filters used by MatLab: the mean, median and Gaussian filter.

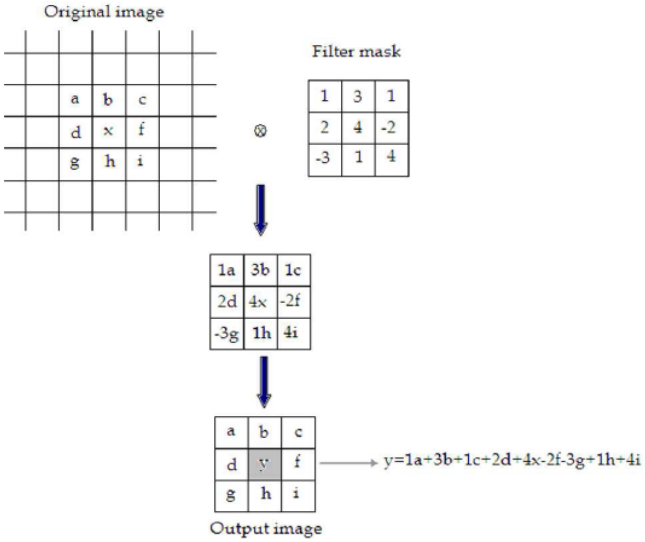



Fig. 4. Illustration of filtering process in spatial domain.

❖ **Mean filter**

Mean filter is the simplest low pass linear filter. It is implemented by replacing each pixel value with the average value of its neighbourhood. Mean filter can be considered as a convolution filter. The smoothing effect depends on the kernel size. As the kernel size increases, the smoothing effect increases too. Usually a 3×3 (or larger) kernel filter is used. An example of a single 3×3 kernel is shown in the Fig. 5.

| | | |
|---|---|---|
| $a$ | $b$ | $c$ |
| $d$ | $e$ | $f$ |
| $g$ | $h$ | $i$ |

$$\longrightarrow \frac{1}{9}(a+b+c+d+e+f+g+h+i)$$

Fig. 5. Filtering approach of mean filter.

The Fig.5 depicts that by using the mean filter, the central pixel value would be changed from "e" to '' (a+b+c+d+e+f+g+h+i) 1/9''.

❖ **Median filter**

Median filter is a non linear filter. Median filtering is done by replacing the central pixel with the median of all the pixels value in the current neighbourhood. A median filter is a useful tool for impulse noise reduction (Toprak & Göller, 2006). The impulse noise (it is also known as salt and paper noise) appears as black or (/and) white pixels randomly distributed all over the image. In other words, impulse noise corresponds to pixels with extremely high or low values. Median filters have the advantage to preserve edges without blurring the image in contrast to smoothing filters.

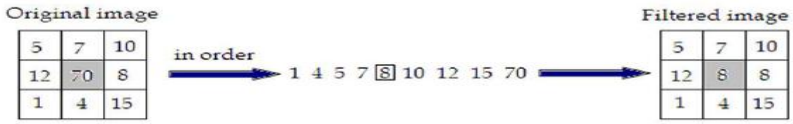



Fig. 6. Filtering approach of Median Filter.

❖ **Gaussian filter**

Gaussian filter is a linear low pass filter. A Gaussian filter mask has the form of a bellshaped curve with a high point in the centre and symmetrically tapering sections to either side (Fig.6). Application of the Gaussian filter produces, for each pixel in the image, a weighted average such that central pixel contributes more significantly to the result than

pixels at the mask edges (O'Gorman et al., 2008). The weights are computed according to the Gaussian function (Eq.1):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / (2\sigma^2)} \quad \text{.....(1)}$$

where μ, is the mean and σ, the standard deviation.

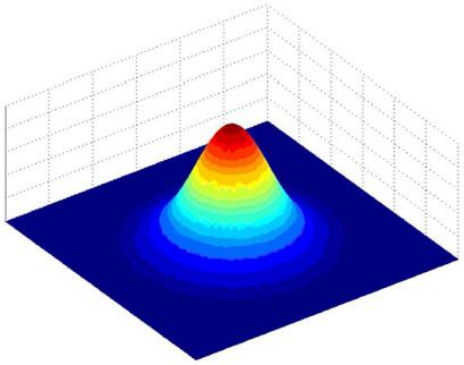


Fig. 7. A 2D Gaussian function.

The degree of smoothing depends on the standard deviation. The larger the standard deviation, the smoother the image is depicted. The Gaussian filter is very effective in the reduction of impulse and Gaussian noise. Gaussian noise is caused by random variations in the intensity and has a distribution that follows the Gaussian curve.

## VI. IMAGE SEGMENTATION

The image segmentation describes the process through which an image is divided into constituent parts, regions or objects in order to isolate and study separately areas of special interest. This process assists in detecting critical parts of a nuclear medicine image that are not easily displayed in the original image. The process of segmentation has been developed based on lots of intentions such as delineating an object in a gradient image, defining the region of interest or separating convex components in distance-transformed images.

Attention should be spent in order to avoid 'over-segmentation' or 'under-segmentation'. In nuclear medicine, segmentation techniques are used to detect the extent of a tissue, an organ, a tumour inside an image, the boundaries of structures in cases that these are ambiguous and the areas that radiopharmaceutical concentrate in a greater extent. Thus, the segmentation process serves in assisting the implementation of other procedures; in other words, it constitutes the fundamental step of some basic medical image processing (Behnaz et al., 2010). There are two ways of image segmentation:

a) based on the discontinuities and,

b) based on the similarities of structures inside an image. In nuclear medicine images, the discontinuity segmentation type finds more applications. This type depends on the detection of discontinuities or else, edges, inside the image using a threshold.

The implementation of threshold helps in two main issues:

i) the removal of unnecessary information from the image (background activity) and,

ii) the appearance of details not easily detected. The edge detection uses the command edge.

In addition, a threshold is applied in order to detect edges above defined grey-scale intensity. Also, different methods of edge detection can be applied according to the filter each of them utilises. The most useful methods in nuclear medicine are the 'Sobel', 'Prewitt', 'Roberts', 'Canny' as well as 'Laplacian of Gaussian'. It is noted that the image is immediately transformed into a binary image and edges are detected. The general function used for the edge detection is the following:

[BW] = edge (image, 'method', threshold) Where [BW] is the new binary image produced, image is the initial one; 'method' refers to the method of edge detection and 'threshold' to the threshold applied. In nuclear medicine, the methods that find wide application are the *sorbel*, *prewitt* and *canny*. In the following example, the *canny* method is applied in order to detect edges in an image.

Example 4

```
I = imread('kidneys.jpg');
figure, imshow(I)
J = edge(I,'canny', 0.048);
figure, imshow(J)
```

Another application of segmentation in nuclear medicine is the use of gradient magnitude. The original image is loaded. Then, the edge detection method of *sobel* is applied in accordance with a gradient magnitude which gives higher regions with higher grey-scale intensity. Finally, the foreground details are highlighted and segmented image of the kidneys is produced. The whole code for that procedure is described below.

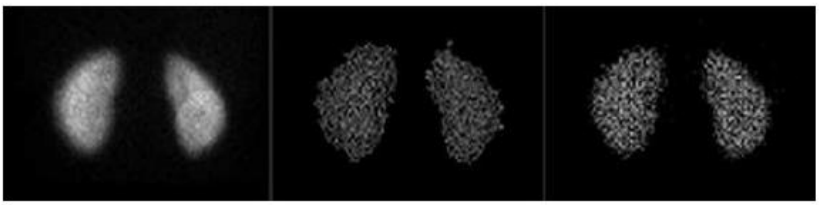


Fig. 8. Edge detection (a) Original kidneys image, (b) edge detection with *canny* method and threshold 0.2667, (c) edge detection with *prewitt* method and threshold 0.038. *[(a) to (c) from left to right]*

Example

```
I = imread('kidneys.jpg');
Figure, imshow(I)
hy = fspecial('sobel');
hx = hy';
Iy = imfilter(double(I), hy, 'replicate');
Ix = imfilter(double(I), hx, 'replicate');
gradmag = sqrt(Ix.^2 + Iy.^2);
figure, imshow(gradmag,[])
se = strel('disk', 20);
K = imopen(I, se);
figure, imshow(K).
```

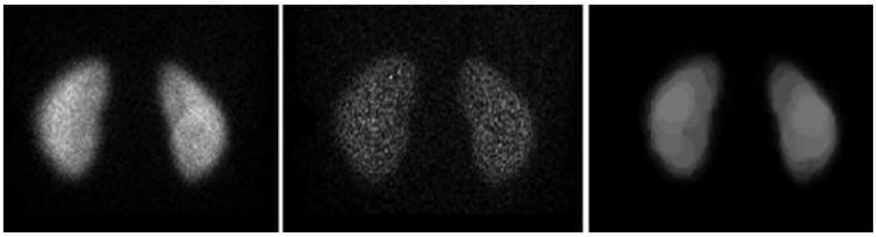


Fig. 9. Gradient Magnitude process: (a) Original image, (b) image after implementation of filter and gradient magnitude, (c) image after masking of foreground objects *[(a) to (c) from left to right]*

In the final image, the outline of the organs is depicted. The area inside the kidney has been separated into larger parts with grey-scale intensity weighted and decided from the closest 20 pixels in a circular region. In

the areas of kidney that have higher activity concentrated, more than one layer of circular regions have been added presenting a final lighter region.

## VII. BACKGROUND ACTIVITY REMOVAL

One of the first steps to be completed in the medical image processing is removing the background activity. This procedure is based on image segmentation as in order to achieve the background activity removal, the organs' boundaries are first defined. The steps in this procedure are the following:

i) the image is read,
ii) the image is appeared,
iii) a grey level threshold is decided by MatLab,
iv) the image is transformed into binary image in order to isolate the two kidneys,
v) the binary image is multiplied by the initial one,
vi) the final image is appeared,
vii) the colour can change (or not) according to individuals' needs. The following example of kidneys image describes the process.

Example
I = imread('kidneys.jpg');
figure, imshow(I) (fig.10a)
graythresh(I) and the value of the threshold is calculated: ans = 0.2667
I2 = im2bw(I, 0.2667) (fig.10b)
I3 = immultiply(I2, I)
imshow(I3) (fig.10c)
colormap(hot) (fig.10d)

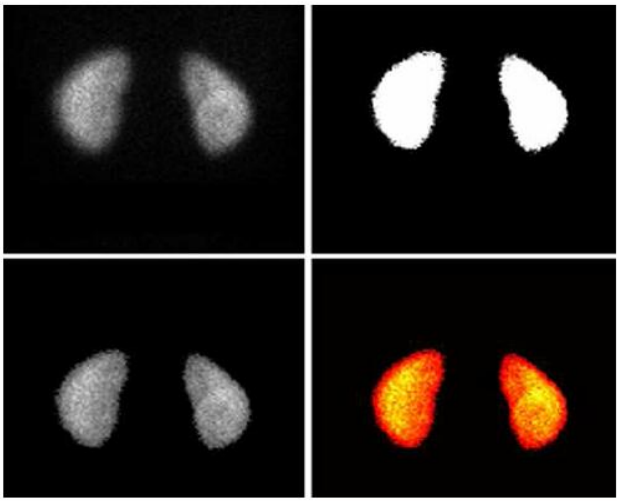


Fig. 10. Background subtraction: (a) Original image, (b) segmented binary image after thresholding depicting only sharp organ boundaries, (c) image after background removal,(d) change of colour to nuclear medicine pattern. [*(a) to (d) from left to right]*

## VIII. IMAGE REGISTRATION

Image registration is used for aligning two images of the same object into a common coordinate system presenting the fused image. The one image is usually referred as reference and the other as sensed (or referred).

Image registration is a spatial transform. The images can be acquired from different angles, at different times, by different or same modalities. A typical example of the use of image registration from different modalities in nuclear medicine is the combination of SPECT and CT images (SPECT/CT) or PET and CT (PET/CT). Image registration is used mainly for two reasons: i) to obtain enhanced information and details from the image for more accurate diagnosis or therapy (Li & Miller, 2010) and, ii) to compare patient's data (Zitova & Flusser, 2003). MatLab can be used in order to perform such a process. The whole procedure shall follow a specific order.

The first step of the procedure includes the image acquisition. After that, each image is reconstructed separately. Any filters needed are applied as well as enhancements in brightness and contrast. The process of filter application has been described in a previous section. The next step includes the foundation of a spatial transformation between the two images, the one of SPECT and the other of CT. The key figure in this step concerns about the alignment of the two images. A spatial transformation modifies the spatial relationship between the pixels of an image relocating them to new positions in a new image. There are several types of

spatial transformation including the affine, the projective the box and the composite (Delbeke et al., 2006). The final step in image registration is the overlapping of the two images allowing a suitable level of transparency. A new image is created containing information from both pictures from which, the first has been produced. The whole procedure can be described with a set of commands which is user customised as different registration function packages can be constructed for different uses.

## IX. ARTHROPLASTY – GENERAL PRESENTATION

Arthroplasty [10] represents a surgical procedure in which the arthritic or dysfunctional joint surface is replaced with prosthesis or by remodeling or realigning the joint. The important joint for this article is the one located at hip area. This is the reason why the article details the parameters that belong to the thigh-bone and the pelvis. Fig.11 presents the most important parameters in Hip Arthroplasty, extracted from an anterior-posterior radiography. Fig.12 presents the parameters extracted from an anterior-lateral radiography representing the hip after the insertion of the prosthesis.For the scope of this article, three areas of the thighbone are analyzed: the femoral head (the nearest part to the pelvis), the femoral neck, and the femoral shaft or body (the longest part of the thigh-bone).

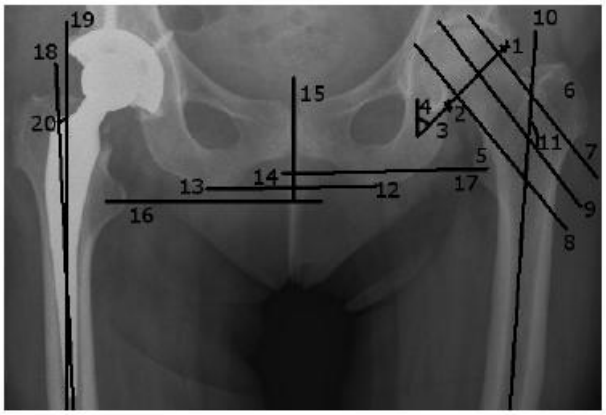



Fig.11. Parameters important in Hip Replacement, extracted from an anterior-posterior radiography

The parameters of interest for the Hip Replacement [7] are listed below:
1) the superior margin of the acetabulum (the superior point in which the thigh-bone meets the pelvis).
2) the inferior margin of the acetabulum (the inferior point in which the thigh-bone (the femoral head) meets the pelvis).
3) the femoral head axis or the acetabulum axis (the line determined by the two points that represent the superior margin and the inferior margin of the acetabulum).
4) the angle created by the acetabulum axis with the vertical line. This angle has to be the same for both of the femoral bones.
5) the lesser trochanter (located in the upper left part of the femoral body).
6) the greater trochanter (located in the upper right part of the femoral body).
7) the tangent to the superior cortical of the femoral neck
8) the tangent to the inferior cortical of the femoral neck
9) the femoral neck axis (the axis of the cylinder determined by the tangents 7 and 8)
10) the femoral body axis or the diaphyseal axis (the axis of the cylinder that approximates the femoral body)
11) the most important parameter, extracted from the x-ray before the surgical intervention is represented by the cervicodiaphyseal angle (the angle determined by the neck axis and the diaphyseal axis).Depending on the value of this angle, it can be determined whether the patient needs or not a prosthesis. If the angle has values between 125 and 135 degrees, the thigh-bone is considered to be in normal ranges.
12) the right ischiadic tuberosity (the lowest right part of the pelvic bone).
13) the left ischiadic tuberosity (the lowest left part of the pelvic bone).
14) the ischiadic line or the horizontal reference line (the line determined by the two ischiadic tuberosities).
15) the vertical reference line, that is perpendicular on the ischiadic line, in its middle.
16) the line starting from the center of the lesser left trochanter, parallel to the ischiadic line
17) the line starting from the center of the right lesser trochanter, parallel to the ischiadic line; the distance between lines 16 and 17 represents the vertical distance between the two thigh-bones. If this distance is greater

than a chosen threshold, there is an indicationof a difference between the lengths of the two femoral bones that has to be resolved surgically (in most cases).
After inserting the prosthesis, some new parameters must be taken into consideration:
18) the diaphyseal axis of the femoral bone (the same as parameter 10).
19) the diaphyseal axis of the prosthesis or the axis of the prosthesis' body (the axis of the cylinder that approximates the prosthesis' body).
20) The deviation of the prosthesis (the angle determined by lines 18 and 19).

This parameter will be computed in several radiographic images, following the evolution of the same patient. The important parameters extracted from an anteriorlateral radiography, after inserting the prosthesis, are shown in Fig.12:

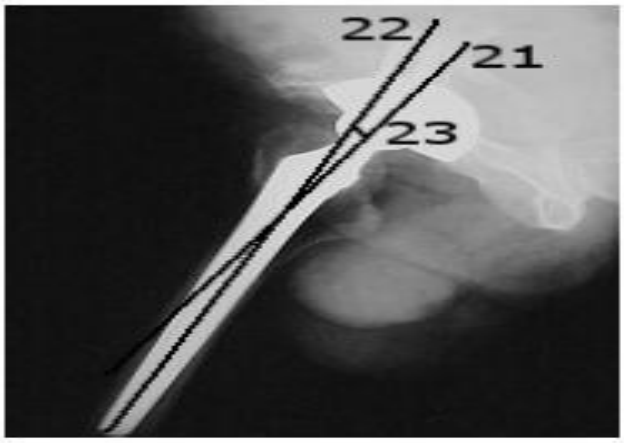



Fig.12. Parameters important in Hip Replacement,extracted from an anterior-lateral radiography, after the insertion of the prosthesis.

21) The axis of the prosthesis' neck
22) The axis of the prosthesis' body
23) The anteversion angle (the angle determined by the axes 21 and 22).
If this angle has its value situated between 5 and 10 degrees, it's considered to be in normal ranges.

## X. The Dicom Standard

In order to manage and interpret the x-rays data in a simple and organized manner, a standard for the x-ray files is needed. This is the reason why the most popular standard for medical images was chosen: DICOM [14]. The *Digital Imaging and Communications in Medicine(DICOM)* standard is a detailed specification of the coding and transfer of medical images and their associated information.

### ❖ REPRESENTING DATA IN THE DICOM FORMAT

The clinical data are represented in a variety of formats:the distances are measured in millimeters, the time in seconds, etc. The PS 3.5 part of the standard, entitled. Data Structure and Their Encoding, defines 27 types of standard data, known as "value representations" (VR),that include all types of data that can appear in the medical domain. Any information encoded in a DICOM file has to belong to one of these predefined types.

Some of the most important standard data are: Person Name (PN), Date Time (DT), and Age String (AS).A DICOM file has the following structure:
- A preamble of 128 bytes
- A prefix (4 bytes) for retaining the letters 'D', 'I',
'C', 'M' (the signature of a DICOM file)
- A data set that contains information like: patient's name, image type, image dimension, etc.
- Pixels that form the image (or the images) contained in the file.

### ❖ EXTRACTING DATA FROM DICOM FILES

Extracting data from a DICOM file can be made using the tags defined in the DICOM dictionary. Every tag is searched in the file, and, if found, is interpreted.
The steps in extracting the information are:
- Checking the existence of the characters 'D', 'I',
'C', 'M'
- Determining of the VR type - Setting the order of the bytes (Big Endian or Little Endian)
- Searching for a tag in the DICOM file,corresponding to the order of bytes and the VR type

- Extracting the values corresponding to that tag Some characteristics of the DICOM files, important when extracting data, are:
- the number of images contained in the DICOM file
- the number of bits per pixel: 8, 12, 16, 24
- the compression
- the photometric interpretation: shades of gray or color images In case of images without compression, the extraction of the images is made pixel by pixel, according to the
number of bits per pixel. For images with compression,a decompression step should be previously performed.

## XI. THE STRUCTURE OF THE CR DICOM FILES

Computer radiographic images (CR) stored in DICOM files are accompanied by general identification elements and some specific information. For example, the Patient module contains: the name of the patient, the patient's ID, the patient's date of birth,
etc. Another module, specific for CR, CR Series, contains information about the examined body part, the view position, etc. Our application extracts from every module the important elements for managing and interpreting the patient's data.

### ❖ Radiographic Image Processing

As in any image analysis application, the first step is a preprocessing step, needed to improve the image by noise removal, contrast improvement, edge enhancement and others [4]. In our application, this step is followed by a contour extraction step, which helps in the arthroplasty parameters' extraction.

### ❖ Image Enhancement

One reason why the automatic interpretation of radiographic images doesn't give accurate results is the fact that the radiographic images are blurred. This is why enhancing images before applying contour detection algorithms is a step that should not be omitted.In the case of our application, the radiographic images are enhanced by noise removal, edge enhancement and
contrast improvement. We will detail each method in the following subsections.

### ❖ Contour Extraction

Most of the contour extraction algorithms which are based on edge detection follow these steps:
- detecting the edge pixels (pixels where the intensity changes abruptly)- eliminating the edge pixels which are not also contour pixels - connecting the contour pixels using local methods (based on the pixels' relations to their neighbor pixels) or global methods (based on global information, for example the shape of a bone, in a computer radiography).After trying a series of methods, the Canny algorithm [5] has been chosen in order to extract the contour lines,because this produced the best results. The Canny algorithm will be briefly described in the following lines. The Canny edge detection algorithm is a very well known algorithm and is considered by many the optimal edge detector. The algorithm is structured into 6 steps:
1) filter out any noise in the original image with the Gaussian filter.
2) apply the Sobel operator on the resulting image, estimating the gradient in the horizontal direction (*Gx*) and in the vertical direction (*Gy*). The magnitude, or the edge strength is approximated by the sum between *Gx* and *Gy*
3) find the edge direction , as the arctangent of *Gy/Gx*.
4) Once the direction is known, relate the edge direction to a discretized direction (all the angles between 67.5 and 112.5 will be considered to be of 90 degrees, all angles between 112.5 and 157.5 are set to 135 degrees, etc). Fig. 13 shows the possible discretized edge directions, previously determined in step 3.

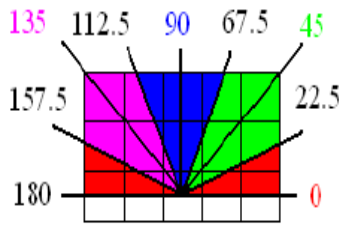



Fig. 13. Gradient direction

5) Non-maximum suppression (trace along the edge in the edge direction and suppress any pixel that is not considered to be an edge - that is not a local maximum)

6) Hysteresis for eliminating streaking, using two thresholds, *T1* (high) and *T2* (low). Any pixel with a value greater than *T1* is considered to be an edge pixel. After applying the first threshold, any pixels connected

to the edge pixels that have a value greater than *T2* will also be selected as edge pixels. Usually *T1=2*T2*.The result of applying the Canny detector is a binary image (Fig. 14), where white pixels represent contour pixels. Having the contour lines of the bones, the next step in our application is the extraction of the important parameters in hip arthroplasty (automatic and semiautomatic extraction). We can observe in Fig. 14 that the contour is disconnected in some parts of the radiography, and that it cannot be reconstructed with accuracy. That is why we propose to search for some salient parameters (for example the lines representing the contour of the femoral body) that can be identified without the previous reconstruction of the entire pelvic and femoral contour. In the following section we will present the methods used after preprocessing the image (with noise removal, edge enhancement, contrast improvement and Canny edge detection), in order to extract the parameters important in arthroplasty.

Fig.14. The image after applying the Canny Edge Detector

## XII. CONCLUSION

The research in the field of medical image analysis is a continuous challenge. The need to discover new image analysis algorithms and new automatic learning techniques that would help in computer assisted diagnosis is and will be a topic of interest for researchers.

The results of our research, presented in this paper, prove that solutions do exist. Although not all the arthroplasty parameters determined automatically were 100% accurate, the application proved to be very useful to doctors. The fact that the application allows patient's data saving management, during a long period of time after the hip replacement procedure is another plus. This application can be used for a single hospital, or for an entire national/international network of hospitals, integrating other applications of diagnosis or of assisting doctors in planning certain surgeries and following the patients' evolution after the surgeries. Image processing and analysis applied to nuclear medicine images for diagnosis, improve the acquired image qualitatively as well as offer quantitative information data useful in patient's therapy and care. Advanced techniques of image processing and analysis find widespread use in nuclear medicine. MatLab and Image Processing Toolbox enable both quantitative analysis and visualization of Nuclear Medicine images acquired as planar or angle projected images to reconstruct tomographic (SPECT, PET) slices and 3D volume surface rendering images.

## REFERENCES

[1] Bankman I, Handbook of Medical Image Processing and Analysis, Academic Press, 2000.
[2] Feng D D, Biomedical Information Technology,Elsevier, 2008
[3] Chen Y, Ee X, Leow K W, Howe T S, Automatic Extraction of Femur Contours from Hip X-ray images, 2000.
[4] Gonzales R C, Woods R E, Digital Image Processing, Prentice-Hall, 2002
[5] Canny J F, A Computational Approach to Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 1986/
[6] Campilho A, Kamel M, Image Analysis and Recognition Springer.
[7] Raj K Sinha, Hip Replacement.
[8] Kennon R, Hip and Knee Surgery: A Patient's Guide to Hip Replacement, Hip Resurfacing, Knee Replacement, and Knee Arthroscopy Book Description.
[9] Botez P, Ortopedie, Bit Publishing House (Iasi), 2001.
[10] Morrey B F, Joint Replacement Arthtoplasty,
[11] SCANIP: Available at http://www.simpleware.com/software/scanip/iptechd ata.php in 10.10.2009

[12] 3D-DOCTOR: Available at http://www.ablesw.com/3d-doctor/index.html in 10.10.2009
[13] Hip-OpCT: Available at http://www.hipop.it/hipopct.html in 20.10.2009
[14] The official page of the DICOM standard. Available at http://dicom.nema.org in 20.05.2009
[15] Mohapatra S, Kumar Sa P, Majhi B, Impulsive Noise Removal Image Enhancement Technique, 6th WSEAS International Conference on CIRCUITS, SYSTEMS, ELECTRONICS, CONTROL & SIGNAL PROCESSING, Cairo, Egypt, 2007.
[16] Chen T, Wu HR, Adaptive impulse detection using center weighted median filters, IEEE Signal Process Lett, 2001.
[17] Georgakopoulos S, Andreadis A, Image Noise Removal Using Graph Theory Concepts, WSEAS Multiconference: Signal, Speech and Image Processing, Rethymno, Greece, 2003
[18] Mohammad F, Al-Otum H M, Oraiqat M T, A Comparison of Image Enhancement using Curvelet Transform with Multiscale Gradient and Retinex Operators, WSEAS Multiconference: Signal, Speech and Image Processing, Izmir, Turkey, 2004
[19] Djekoune O, Achour K, Halimi M, Kahlouche S, Incremental Hough Transform: An Improvement Algorithm for Digital Devices Implementation, WSEAS Int. Conference on Electronics, Control & Signal Processing and E-Activities, Singapore, 200
[20] Bidgood, D. & Horii, S. (1992). Introduction to the ACR-NEMA DICOM standard. RadioGraphics, Vol. 12, (May 1992), pp. (345-355)
[21] Delbeke, D.; Coleman, R.E.; Guiberteau M.J.; Brown, M.L.; Royal, H.D.; Siegel, B.A.; Townsend, D.W.; Berland, L.L.; Parker, J.A.; Zubal, G. & Cronin, V. (2006). Procedure Guideline for SPECT/CT Imaging 1.0. The Journal of Nuclear Medicine, Vol. 47, No. 7, (July 2006), pp. (1227-1234). Gonzalez, R.; Woods, R., & Eddins, S. (2009) Digital Image Processing using MATLAB, (second edition), Gatesmark Publishing, ISBN 9780982085400, United States of America
[22] Lehmann, T.M.; Gönner, C. & Spitzer, K. (1999). Survey: Interpolation Methods in Medical Image Processing. IEEE Transactions on Medical Imaging, Vol.18, No.11, (November 1999), pp. (1049-1075), ISSN S0278-0062(99)10280-5
[23] Lyra, M.; Sotiropoulos, M.; Lagopati, N. & Gavrilleli, M. (2010a). Quantification of Myocardial Perfusion in 3D SPECT images – Stress/Rest volume differences, Imaging Systems and Techniques (IST), 2010 IEEE International Conference on 1-2 July 2010, pp 31 – 35, Thessaloniki, DOI: 10.1109/IST.2010.5548486
[24] Lyra, M.; Striligas, J.; Gavrilleli, M. & Lagopati, N. (2010b). Volume Quantification of I-123 DaTSCAN Imaging by MatLab for the Differentiation and Grading of Parkinsonism and Essential Tremor, International Conference on Science and Social Research, Kuala
[25] Lumpur, Malaysia, December 5-7, 2010. http://edas.info/p8295 Li, G. & Miller, R.W. (2010). Volumetric Image Registration of Multi-modality Images of CT,
[25] MRI and PET,Biomedical Imaging, Youxin Mao (Ed.), ISBN: 978-953-307-071-1, InTech,Available from:http://www.intechopen.com/articles/show/title/volumetric-image-registrationof- multi-modality-images-of-ct-mri-and-pet
[26] O' Gorman, L.; Sammon, M. & Seul M. (2008). Practicals Algorithms for image analysis, (second edition), Cambridge University Press, 978-0-521-88411-2, United States of America
[27] Nailon, W.H. (2010). Texture Analysis Methods for Medical Image Characterisation, Biomedical Imaging, Youxin Mao (Ed.), ISBN: 978-953-307- 071-1, InTech, Available from: http://www.intechopen.com/articles/show/title/texture-analysis-methodsfor-
[28] medical-image-characterisationMathWorks Inc. (2009) MATLAB User's Guide. The MathWorks Inc., United States ofAmerica.
[29] Perutka K. (2010). Tips and Tricks for Programming in Matlab, Matlab - Modelling, Programming and Simulations, Emilson Pereira Leite (Ed.), ISBN: 978-953-307-125-1, InTech, Available from: http://www.intechopen.com/articles/show/title/tipsand-tricks-for-programming-in-matlab.
[30] Toprak, A. & Guler, I. (2006). Suppression of Impulse Noise in Medical Images with the Use of Fuzzy Adaptive Median Filter. Journal of Medical Systems, Vol. 30, (November2006), pp. (465-471)
[31] Wernick, M. & Aarsvold, J. (2004). Emission Tomography: The Fundamentals of PET and SPECT,Elsevier Academic Press, ISBN 0-12-744482-3, China Wilson, H.B.; Turcotte, L.H. & Halpern, D. (2003). Advanced Mathematics and Mechanics Applications Using MATLAB (third edition), Chapman & Hall/CRC, ISBN 1-58488- 262-X, United States of America
[32] Zitova, B. & Flusser J. (2003). Image Registration methods: a survey. Image and Vision Computing. Vol 21, (June 2003), pp. (977-1000).