International Journal of Scientific Research in Computer Science and Engineering
Research Paper Volume-3, Issue-6 ISSN: 2320-7639

# Proposed Model for Privacy Preserving in Mobile Cloud Computing

Marjan Soltani[1], Gholamreza Shahmohammadi[2*]

[1]*Department of Computer Engineering Islamic Azad University, Ashtian, Iran*
[2]**Department of Information Technology, Olum Entezami Amin University, Tehran, Iran*

www.isroset.org

Received: 27/Oct/2015 Revised: 12/Nov/ 2015 Accepted: 09/Dec/2015 Published: 30/Dec/2015

**Abstract.** Today one of the issues involved in cloud computing is the authenticity of the outsourced calculations that are delivered by cloud service providers. Moreover, controlling the penetration of the invaders in this system has a great importance. Regarding the ceaseless growing of information, protection of them in the mobile cloud environment has become more difficult. So it is necessary to use the methods such as data mining to extract knowledge and the hidden information in data and to save them. The aim of this study is to present a security architecture to keep privacy in mobile cloud computing. In this study we proposed a framework for preserving the security of the clouds' data using the data mining techniques. Then by extracting useful rules available on exchanged data on cloud environment, a useful information resource is provided for system manager to secure data and reduce the computational overhead. Various methods such as parallel executing of algorithms' associative rules extraction are used for reducing computation of cost and improving process of associative rules extraction.

**Keywords-** Cloud Computing, Security Architecture, Privacy

## I. INTRODUCTION

Today mobile devices such as mobile phones, tablets and smartphones are an important part of human life. Mobile phone users have rich experiences on working with different services and various applications. These services are on the mobile phones and service providers and they are available through Wi-Fi networks. Development in mobile computing has attracted the fields such as IT, industry and commerce.

The term "mobile cloud computing" was introduced in the mid 2002 shortly after the concept of cloud computing. Mobile Cloud Computing (MCC) has attracted the attention of entrepreneurs and it is a good choice for commerce because it reduces the cost of developments and causes the execution of mobile applications. It is a novel technology for the users to gain rich experiences while working with the low cost services of the mobile phones and for the researchers it is a solution to step towards green IT. Association of mobile cloud computing defines MCC as follows: In the most basic form, it refers to a basis in which saving and processing data is done outside a mobile device. It can transfer the applications, power of processing and data saving from mobile device to the cloud. So the transaction topic comes forward in order to outsource data. The concept of transaction has lately changed from the parallel one to distributed and web transaction and recently to cloud transaction in order to outsource. The idea of cloud transaction has recently brought forward and it's a new field in the computer science. Today cloud computing is defined as an environmental transaction between mobile and cloud in which the transactions are outsourced to be available at the times of needs for example accessing data base. Cloud transaction is a new concept in IT in which the transaction of the data is not exclusive to mobile and desktop devices and takes them to big data centers.[1]
The security of cloud transaction is of two kinds:

The security of cloud saving which is done for the assurance of the integrity of outsourced data that are saved on unreliable cloud service providers. The security of cloud computing which is done for authentication of the outsourced calculations provided by unreliable cloud service providers.

Security and privacy is the most important thing in mobile cloud computing, because data saving and processing is performed by the customer. It means that all customers are able to see the abstract basis on the unreliable physical hardware. So it is necessary to provide the privacy of the customers as a service without much additional charges.

The purposes of this study are:

1. Examining different kinds of attacks in the cloud
2. Providing security architecture to optimize the costs of calculation
3. Diagnosing Dos[1] attacks to the outsourced data base
4. Providing security protocol that increases the security of the mobile devices for connection to the data centers.

One of the solutions that are generally used for the management of the big cloud data bases is the outsourcing of data base for the purpose of technical work and reduction in

Corresponding Author: Gholamreza Shahmohammadi, Shah_mohammadi@yahoo.co.uk *Department of Information Technology, Olum Entezami Amin University, Iran*

[1] Denial of Service

the costs and data base is known as a service. The concentration of discussion on the security issues is due to the acceptance of data base model as a service based on the outsourced servicing in cloud computing. The advantages of this concept are reduction in the costs of maintenance of data, more effective data security and more accessibility to the data. The scenario of outsourced data base, the important data of an organization are not under the direct control of the owner. So some security matters such as confidentiality of the data, user privacy, authenticity of data, identification of users and controlling the accessibility must be observed.

Users usually have doubt about using cloud data bases and the origins of these doubts could be the lack of trust in the service provider for privacy matter of the outsourced data or the lack of trust in the honesty of service provider in the processing related to management. For providing security in the outsourcing of cloud data, we need a framework in order to keep the privacy or privacy-preserving data mining or PPDM[2].[1]

Most of the provided frameworks are for sharing data with a third person (such as disguising data, encoding data or reverse transition of data). In all these transitions, privacy of the personal data is the key factor.

In this study we are supposed to work exclusively on the model of attack based on repetition such as DoS. This is a kind of attack that service provider has exact information on the total data and personal data items. Now we can assume that invader can perform two types of attack in the system:

All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

## II. PROCEDURES

In this study we are following these goals:

1- First we are going to design a
2- scenario under the cloud
3- environment regarding the outsourcing of the data base transactions.
4- We provide a pattern for security architecture for data base as a service regarding the connections and entities. These entities consist of owner of data, user, applicant and executor.
5- Generally some great costs are spent on the probing for repetitive data items and important outsourced ones. To decrease these probing and to protect these data from invader, first we use clustering methods to divide the data. So we are going to use a fuzzy multi –objective genetic algorithm to increase the accuracy in clustering and to make some groups with maximum similarity.
6- On each cluster we recognize the repetitive items separately. We suggest the high backed up items to the encoding system to reduce the computational load. It means that we encode those transaction that are similar to each other in one piece so that it is not recognizable by the invader and also to reduce the cost of transition to the outsourced data base.
7- Regarding the items in different groups, we produce fake items with selection of random transactions in each cluster so that the partial effect of this item on the system is lowered and also to lessen the data dependency to the system.

## III. Research questions:

1- How clustering is done on the transactions' data bases to be prepared as some groups for the stage of encoding?
2- How can we make fake transactions?
3- Which items best suit encoding? Repetitive items or scarce ones.
4- Can this proposed approach give a numeral possibility to the user for the stealing data items? Does the number of transaction items play a role in the percentage of the possibility of attack?
5- Does the proposed approach enhance the system performance?

## IV. Research Hypothesis

1. The quality of the keeping privacy of the outsourced data base is warranted by the encoded data base.
2. Each group of transactions are saved in a dependency space to be encoded.
3. For each data item there is a probability for occurrence in real space.
4. Proposed system improves system functionality and reduces the overload of system.

## V. Introduction of some security architecture

In [18] a new comparative real time solution is presented to supervise over the measurability and reliability of data. In this study a fake positive (FB) is the load recognized by the supervised time series algorithm. The supervised algorithm is

[2] privacy-preserving data mining

done by change in the main time series and it can randomly show the load of the system with a precision less than 1. With combination of these two metrics, the estimation of F value is presented in high quality recognition. If F value is close to 1 it shows a good recognition while this number for weaker algorithms in load supervision is lowered.

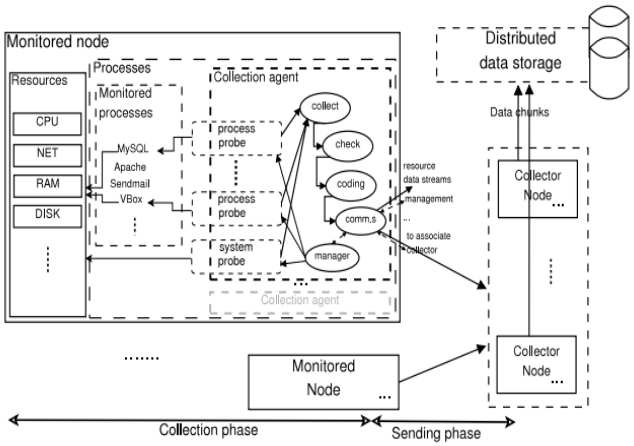



Figure 1 – Supervised architecture [18]

The purpose of [19] is to suggest a cloud computation framework that is necessary for the manipulation of big data. In the prototype of the system the identification number of Bangladesh nationality is taken into account by Bangladesh election commission.

The basis is divided to two main parts: local host cloud based on "eucalyptus" and the remote cloud that is done on Amazon web service (AWS). Some common problems in this system such as data traffic, time of service provider are also discussed.

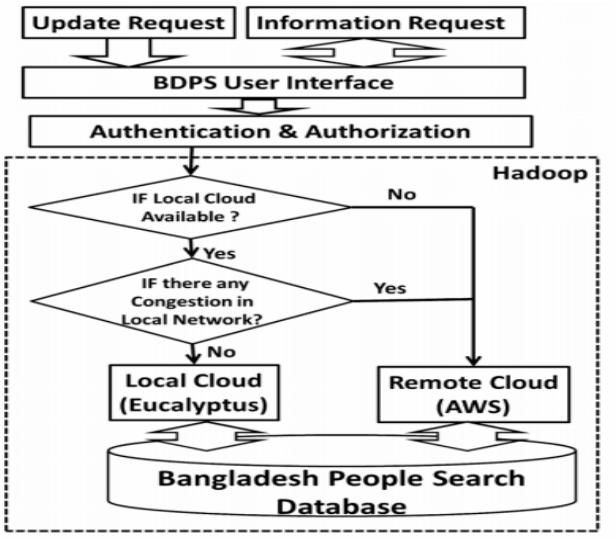



Figure 2 – Security architecture of Bangladesh system [19]

In [20] the challenges of supervision over the data saving are presented. Regarding the scalability and high performance, six different systems for saving are analyzed. In this study the concentration is on the maximum power that could be achieved by the proposed system.

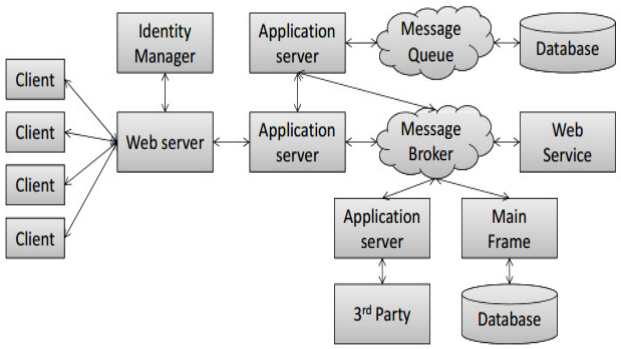



Figure 3- model of architecture for an organization system [20]

In [21] with the existence of measureable solution, migration on software architecture causes elimination of all basic needs alongside with the issuance of the software permission that is needed for their execution. After migration to the cloud each customer may begin to use the software without the help the provider of software. This process is completely automatic and it is one of the main points of migration.

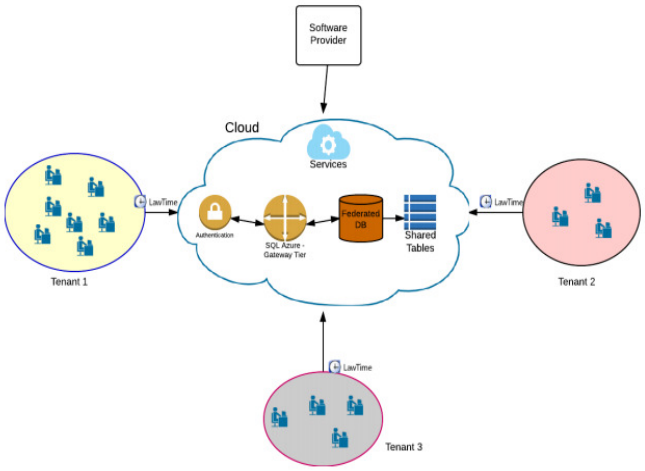



Figure 4- Method of software delivery after migration to cloud [21]

Customer tries to access the data. Identity provider is one of security matters and it is for identification of the user. Some common example of these IPs are Microsoft account, Facebook, act.

In the figure shown below a logical lookout is brought forward for identification mechanism under UML diagram. Cloud service provider has some services for controlling accessibility to authentication of users' identities from

identification providers such as Microsoft account, Google and Facebook.

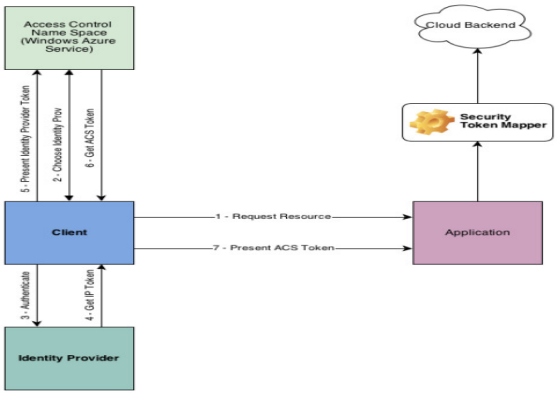



Figure 5: logical lookout (identification) [21]

In [22] the data base introduces a new transaction service called cloud connector. DbaaS consists of operations such as overload providing, configuration, scalability, regulation of function, making back up, keeping privacy and controlling the accessibility of users to data base of service provider. It also causes lower costs for the users.

The initial concerns of DbaaS consist of Amazon RDS and Microsoft Azure SQL that has considered the needs of the market for such services. However, three challenges are not considered: multi-tenant performance, resilience scalability and keeping the privacy of data base.

These three challenges should be considered and worked out for most users before the outsourcing of data base software and attraction management and it should become economical for service providers. The main technical characteristics of cloud connector are: 1- Multi tenancy that performs the recognition of heavy duties and can cooperate in data base service provider and also causes higher stability and better performance. 2- Using an algorithm for dividing data based on graph to reach resilience scalability close to linear even for heavy transactional duties. 3- An adjustable security plan that enables the enquiry of SQL before encoding such as linking and integration.

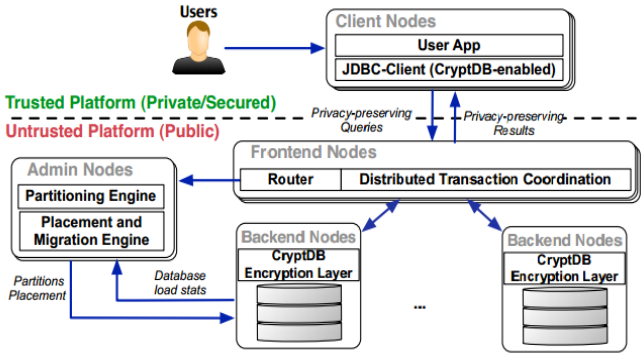



Figure 6- architecture of connector cloud.[22]

The main topic in designing different parts of connector cloud is the concept of load information. Some work is done for configuration of users and useful operators with supervision on inquiry patterns, accessible information, information system for optimization and security of various functions.

## VI. Suggested approach:

Today providers of the cloud prefer to invest on cloud services instead of working on the basic matters. The advantages of using cloud data base give this method a special priority. If organizations don't use this data base, they will be forced to spend more, establish a personal data center and recruit more manpower.

In the last decade, Wireless and mobile phones have gone through great technologic changes. For example one of these changes is wireless cloud computing. It's a concept that created a big wave of evolutions on the mobile phones. The basis of cloud computing is to provide services and processing capacity on the internet that leads to reduction of costs, increasing of savings, automatic systems, flexible and portable information. Wireless cloud computing has some benefits for users such as sharing sources and useful applications between the users, without which a great deal of cash needs to be spend on hardware and software. It causes a reduction in the price for end users of mobile phones.

The aim of this study is to present an encoding pattern that is able to provide users with their privacy and this model is done based on big scale transaction data base. In figure 7 the above architecture is introduced. Customer of owner codes the data using encoding/decoding module that is considered a black box in architecture. This module is responsible for converting the input data to encoded data base. Executor performs the data mining and sends the pattern of encoding to the owner.

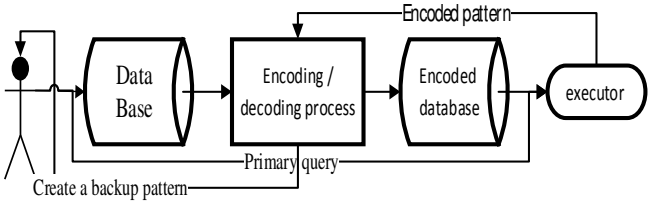



Figure 7: Architecture of suggested system [23]

## VII. Model of attack

Owner of data considers the correct items [23]

1- Each item is encoded
2- Each transaction is encoded
3- Each repetitive pattern is encoded

Consider the attack model below:

1- Attack based on item: $\forall$ encoded items $e \in \mathcal{E}$ , an invader, candidates a collection of simple items $Cand(e) \subset I$ . [23]
Probability if the items e to get defeated is Probe(e) =1/I Cand (e) I [23]
2- Attack based on set. Consider a collection of items that are encoded (E). Invader makes a collection of simple items

$\forall X \in Cand(e), X \subset I . |X| = |E|$ [23]

Probability of the defeat of the encoded items is:

$prob(E) = 1/|Cand(E)|$ [23]

### VIII. Fuzzy multi objective clustering

Imagine that X= [x1, x2,...,Xn] is a set of n categorical discrete items. Each item of Xi that i=1,2,....n is described by a set of P like A1,A2,...Ap. Imagine that $DOM(A_j)$ $1 \le j \le p$ is a domain characteristic and consists of different setting of qj. So the i-th discrete item is defined as $x_i = [x_{i1}, x_{i2}, ..., x_{ip}]$ in which $1 \le j \le p, x_{ij} \in DOM(A_j)$ . Center of cluster in FCM is replaced by the modes of each cluster in K-mode fuzzy clustering. The defined modes are as follow: Imagine that Ci is a set of discrete items and belongs to cluster i. Mode of i is a vector $m_i = [m_{i1}, m_{i2}, ..., m_{ip}]$ that $1 \le j \le p, m_{ij} \in DOM(A_j)$ in a way in a way that the below criterion is minimum:[24]

$$D(m_i, C_i) = \sum_{x \in C_i} D(m_i, x) \qquad (1)$$

$D(m_i, x)$ is the amount of heterogeneity between $m_i$ and $x$ . $m_i$ is not necessarily a member of the set $c_i$ .
The algorithm of dividing K-mode with the data of X and K clusters to get it to the minimum is:[24]

$$j_m(U, Z : X) = \sum_{k=1}^{n} \sum_{i=1}^{K} u_{ik}^m D(z_i, x_k) \qquad (2)$$

For the probability fuzzy clustering the following conditions are necessary:

$$0 \le u_{ik} \le 1, \quad 0 \le i \le K, \quad 0 \le k \le n$$

$$\sum_{i=1}^{K} u_{ik} = 1, \quad 1 \le k \le n \qquad (3)$$

$$0 \le \sum_{k=1}^{n} u_{ik} \le n, \quad 1 \le i \le K \qquad (4)$$

$m$ is the fuzzy power and $U = [u_{ik}]$ is the fuzzy dividing matrix of $K \times n$ and $u_{ik}$ is the membership degree of Kth discrete item in i-th cluster.

$Z = \{z_1, z_2, ..., z_K\}$ is the indicator id cluster centers or modes.

The algorithm of fuzzy K-mode is a part of continuous optimizing strategy that consists of calculation of the repetition of matrix and computation of the new cluster centers (modes). It starts with initial random K mode and after that in each cycle the fuzzy membership in each cluster id get by this formula:[24]

$$u_{ik} = \frac{1}{\sum_{j=1}^{K} \left( \frac{D(z_i, x_k)}{D(z_j, x_k)} \right)^{\frac{1}{m-1}}}, \quad 1 \le i \le K, \quad 1 \le k \le n \qquad (5)$$

If $D(z_j, x_k)$ is zero for some js , so that set $u_{ik} = \{0, ..., 0\}$ for $i = 1, ..., K$ and for $i = j$

Based on the values of the membership the mode of the clusters is calculated like the following. If the values of membership are static, the positions of the modes that minimize the function are like this:

$z_i = [z_{i1}, z_{i2}, ..., z_{ip}]$

In a way that : $z_{ij} = a_j^r \in DOM(A_j)$

$$\sum_{k, x_{kj} = a_j^r} u_{ik}^m \ge \sum_{k, x_{kj} = a_j^t} u_{ik}^m, \qquad 1 \le t \le q_j, \quad r \ne t \qquad (6)$$

These are the conditions for the conclusion of the time algorithm and there is no significant improvement in the value of Jm. Finally each item is dedicated to the cluster with maximum membership. The disadvantages of K-mode algorithm are: 1- It highly depends on the initial modes selection. 2- It is usually trapped in local optimization so we need to use evolutionary optimizing algorithm.

Each chromosome is a domain of values for showing K number of modes in each cluster. If each discrete item has p characteristics, the set [A1, A2....Ap] is the length of K*p chromosome in which the first P is the mode of first cluster, the send p is the mode of second cluster and suppose P=3 and K=3, them chromosome is like this:

$c_{11}$ $c_{12}$ $c_{13}$ $c_{21}$ $c_{22}$ $c_{23}$ $c_{31}$ $c_{32}$ $c_{33}$

This shows three cluster modes:

$(c_{11}, c_{12}, c_{13}), (c_{21}, c_{22}, c_{23}), (c_{31}, c_{32}, c_{33})$

$c_{ij}$ is the j-th characteristic of i-th mode of the cluster . $c_{ij} \in DOM(A_j), 1 \le i \le K, 1 \le j \le p$

**Initial population**:
The numbers of K modes are coded in each chromosome and they are selected from the set of discrete data as K random

items. This process is repeated for each chromosome in the population.

Calculation of assessment function: In this study compression and separability are two objective functions that are considered as parallel.

First the encoded modes are extracted and like $z_1, z_2, \ldots, z_k$ the membership value of $u_{ik}$ in which $i = 1,2,\ldots,K$ and $k = 1,2,\ldots,n$ are calculated by this formula: [24]

$$u_{ik} = \frac{1}{\sum_{j=1}^{K} \left( \frac{D(z_i, x_k)}{D(z_j, x_k)} \right)^{\frac{1}{m-1}}}, \quad 1 \le i \le K, \quad 1 \le k \le n \tag{7}$$

If $D(z_j, x_k)$ is zero for some js, so $u_{ik} = \{0, \ldots, 0\}$ for $i = 1, \ldots, K, i \neq j, u_{ik} = 1$ in which $m$ is the index of weight. Each one id encoded as a chromosome and $z_i = [z_{i1}, z_{i2}, \ldots, z_{iq}]$ in a way that $z_{ij} = a_j^r \in DOM(A_j)$ and

$$\sum_{k, x_{kj} = a_j^r} u_{ik}^m \ge \sum_{k, x_{kj} = a_j^t} u_{ik}^m, \qquad 1 \le t \le q_j, \quad r \neq t$$

It means that a category of $A_j$ from the center of the cluster $z_i$ is a set of values that brings the set of $u_{ij}$ (membership degree of i-th cluster) to maximum. On this basis, the value of cluster membership is calculated again. Variation of $\sigma_i$ and fuzzy cardinality of $n_i$ in i-th cluster in which $i = 1, \ldots, K$ is derived from this formula.[24]

$$\sigma_i = \sum_{i=1}^{K} \frac{\sigma_i}{n_i} = \sum_{i=1}^{K} \frac{\sum_{k=1}^{n} u_{ik}^m D(z_i, x_k)}{\sum_{k=1}^{n} u_{ik}} \tag{8}$$

For calculation of the fuzzy separable fitness function of Sep, suppose that the mode Zi on the i-th cluster is the center of a fuzzy set $\{z_i \mid 1 \le j \le K, j \neq i\}$. So the membership of each Zj to Zi in which i#j, is derived from this formula.

$$\mu_{ij} = \frac{1}{\sum_{l=1, l \neq j}^{K} \left( \frac{D(z_j, x_i)}{D(z_j, x_l)} \right)^{\frac{1}{m-1}}}, i \neq j \tag{9}$$

So the fuzzy separability is defined like this:

$$Sep = \sum_{i=1}^{K} \sum_{j=1, j \neq i}^{K} \mu_{ij}^m D(z_i, x_j) \tag{10}$$

Remember that compression of the cluster is done by minimizing $\pi$.

In contrast, for getting a good separated cluster the separability of Sep should be maximized. So these two goals are considered as optimizing functions for evolutionary algorithm. So we try to minimize the Function $\pi$ and $\frac{1}{sep}$. The purpose of the multi objective clustering is synchronizing optimization of more than one goal and increasing the function by choosing these goals. Careful selection of the goals can produce acceptable results and injudicious selection can lead to bad results.

In this study we used usual genetic operators. Binary tournament operator and single point crossover operator are used.

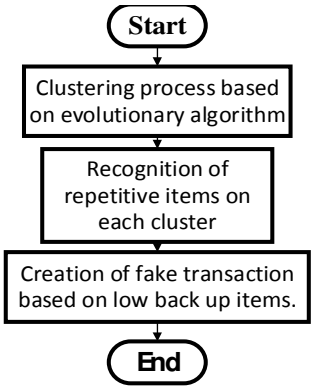



Figure 8 – Proposed Algorithm

## IX. Simulation

Security and privacy are of prime importance in cloud computing that need attention, because saving and processing of the important data are controlled by the customer himself. It means that all customers are able to see the abstract basis built on physical unreliable hardware. So the privacy of data should be provided by the cloud and with the least extra cost. Today with considering of the increase in the power of computation from one side and reduction of the prices on the other side, data mining using the parallel techniques could be a practical solution for the problem of time consuming tasks of data mining from the data base.

The data used in this research are from a chain store in Italy called COOP. In this system the technology of cloud is used. It contains 60366 customers, 4567 products and 107371973 total shopping. In this part of clustering, first we put together the data from all three available sets and we get a single data set. 4 variables are the number of customer, number of products, price of the products and the distance of the shop from customer's home. So we have a data set with 4 columns and 107371973 rows. Performance of such algorithm needs powerful computers. So, considering the current facilities, we divide the first 150000 records and we perform fuzzy multi objective clustering based on the genetic algorithm and finally we get a set of final solutions.

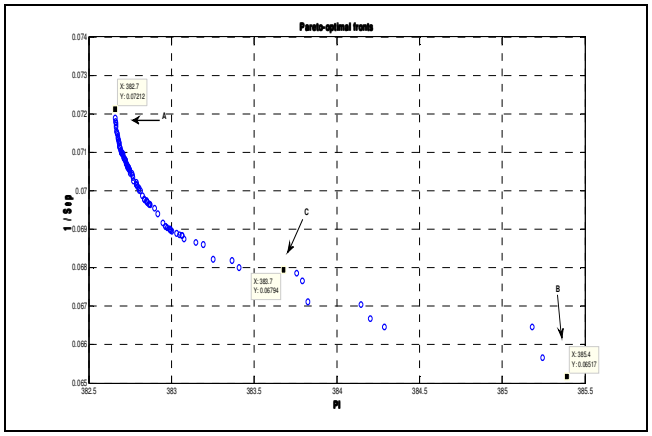



Figure 9- Pareto Front

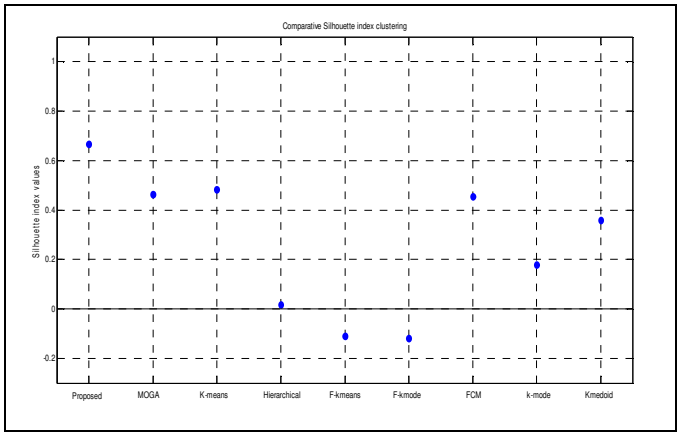



Figure 10: Comparison of authenticity method of Silhouette

For analyzing the Pareto Front in the above figure, three areas of A, B and C are specified. The points that are in A have the best resolution and the higher these points, the better resolution. The points that are in B have the best compression. But the points in C area are at a good level both in resolution and compression. In other words these points have a medium level of resolution and compression. The elbow point in the above figure show the best form of clustering. We use different criteria for assessment of the proposed clustering.

Authenticity method of Silhouette: In this part for assessing the amount of similarity between the proposed system and the reference one, we use the index of Silhouette. By comparing the proposed algorithm and the other algorithms, we see that the index of Silhouette is more in the proposed algorithm in regard of the others. So the similarity of the presented clusters in the proposed method to the reference system (correct clustering) is more than the others.

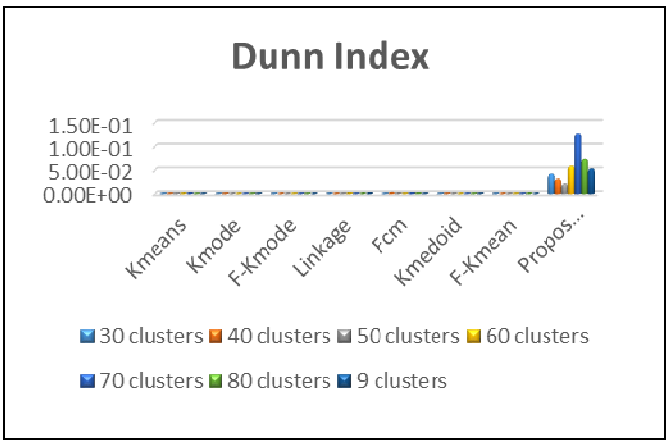
| | **Proposed** | **Kmeans** | **Hierarchical** | **F-kmeans** | **F-kmeans** | **F-kmode** | **FCM** | **k-mode** | **Kmedoid** |
|---|---|---|---|---|---|---|---|---|---|
| **Silhouette index values** | **0.6659** | 0.4822 | 0.0185 | -0.1085 | -0.1085 | -0.1180 | 0.4559 | 0.1782 | 0.3594 |

a: Comparison of authenticity method of Silhouette

The authenticity method of Dunn index: If a set of data contains separable clusters, it is expected that the distance between clusters are a lot and the diameter of the clusters are small. So a bigger amount in this criterion is better.

| Cluster | Kmeans | Kmode | F-Kmode | Linkage | Fcm | Kmedoid | F-Kmean | **Proposed** |
|---|---|---|---|---|---|---|---|---|
| **30** | 1.39E-04 | 1.07E-04 | 1.05E-04 | 4.33E-04 | 1.34E-04 | 1.42E-04 | 1.11E-04 | **0.041674** |
| **40** | 2.21E-04 | 5.57E-05 | 1.06E-04 | 8.63E-05 | 0.0012 | 1.93E-04 | 9.14E-05 | **0.031302** |
| **50** | 1.93E-04 | 5.46E-05 | 6.35E-05 | 5.30E-05 | 7.12E-04 | 1.57E-04 | 8.63E-05 | **0.021105** |
| **60** | 3.69E-04 | 1.05E-04 | 2.43E-04 | 8.65E-04 | 2.69E-04 | 1.16E-04 | 9.39E-05 | **0.058579** |
| **70** | 7.12E-04 | 9.06E-05 | 9.06E-05 | 9.84E-05 | 9.03E-04 | 1.93E-04 | 2.72E-04 | **0.126** |
| **80** | 1.88E-04 | 6.39E-05 | 2.74E-04 | 2.83E-04 | 0.001 | 3.18E-04 | 6.57E-04 | **0.073** |
| **90** | 2.54E-04 | 1.06E-04 | 1.74E-04 | 6.49E-04 | 2.01E-04 | 1.29E-04 | 1.03E-04 | **0.0529** |




b: Comparison of authenticity method of Dunn index

Figure 11: Comparison of authenticity method of Dunn index

Regarding figure 5 it is clear that the proposed method has a higher Dunn index regarding all the other methods and therefore the quality of the clusters is more. Comparison of the Dunn index between different cluster numbers from 30 to 90 shows that 70 is the lucky number.

| Cluster | Kmeans | Kmode | F-Kmode | Linkage | Fcm | Kmedoid | F-Kmean | Proposed |
|---|---|---|---|---|---|---|---|---|
| 30 | 1.7159e-55 | 7.4871e-52 | 6.9935e-61 | 1.5480e-69 | 1.1582e-53 | 7.2913e-66 | 3.92E-54 | **5.1680e-59** |
| 40 | 1.2802e-54 | 6.4155e-41 | 2.1380e-39 | 1.4290e-49 | 1.9016e-54 | 1.1580e-47 | 4.76E-50 | **1.7949e-57** |
| 50 | 1.4071e-53 | 4.2195e-36 | 7.7059e-40 | 2.1860e-79 | 4.9969e-48 | 1.4753e-55 | 1.67E-48 | **5.6419e-68** |
| 60 | 2.0972e-51 | 1.0741e-42 | 1.8236e-42 | 2.3334e-60 | 2.9709e-54 | 2.6559e-50 | 7.00E-52 | **5.1143e-51** |
| 70 | 2.5277e-60 | 5.6479e-33 | 1.2844e-38 | 1.6514e-57 | 1.6146e-52 | 8.1427e-43 | 5.38E-53 | **3.2052e-63** |
| 80 | 4.6767e-41 | 3.0741e-34 | 2.2827e-35 | 4.9750e-55 | 6.8964e-50 | 7.0340e-46 | 1.56E-41 | **3.9800e-54** |
| 90 | 2.2037e-46 | 1.3912e-26 | 1.7536e-28 | 3.3171e-63 | 9.9097e-44 | 1.2878e-52 | 3.31E-44 | **7.1482e-43** |

c: P value produced by t-test

For authenticity of the data in different clustering algorithms we do a t-test with the significance level of 5% for each clustering algorithm. P value is the probability of random data for Dunn index in each clustering. P-values for each algorithm show that the results are significant. Based on the null hypothesis, it is assumed that there is no significant difference between the amounts of Dunn index in each clustering. The other hypothesis is that there is a significant difference. Table 3-5 shows that all p-values are less than 0.05 and therefore the data are significant.

In this part, time complications for both two forms of series and parallel in the proposed framework are examined. Acceleration derives from parallel algorithm. Here we need to define these concepts:

P: the number of Slave processors

N: number of repetition

F: The average of time for performance of fitness function for each population in each repetition

G: the average of the time for performance of all the process in each repetition

C: The average of the connection time between master and slave processor

S: The average of the time for rule mining

The clustering algorithm in series form

$T_{ave}^{p}$ : The average of the time for performance of the clustering algorithm in parallel form

$S_{ave}$ : The average of acceleration

The average of the time for performance of the clustering algorithm in series form equals:

$$T_{ave}^{s} = n*(p*f+g)+s \qquad (11)$$

The algorithm of fuzzy parallel rule mining can relegate the duty of fitness function assessment to the slave processor. So the time of performance for the assessment of the fitness function in each repetition equals to the maximum of one performance of F. However, in parallel algorithm, we consider the extra computation time that is the time for the connection of master and slave processors. The average of the time for performance of the clustering algorithm in parallel form equals to:

$$T_{ave}^{p} = n*(f+g+c)+s \qquad (12)$$

Complication of the acceleration time equals to:

$$S_{ave} = \frac{T_{ave}^{s}}{T_{ave}^{p}} = \frac{n*(p*f+g)+s}{n*(f+g+c)+s} \qquad (13)$$

Since the time of data transfer from master to slave is very short, it is overlooked. On the other side the value of G is

much less than F, therefore it is possible to overlook it against the assessment time of the fitness function. After simplification we have:

$$S_{ave} = \frac{T^{s}_{ave}}{T^{p}_{ave}} = \frac{n*(p*f+g)+s}{n*(f+g+c)+s} \approx \frac{n*p*f+s}{n*f+s} \quad (14)$$

The average of the time for association rule mining is the time that is spend on finding big and small items, therefore S is smaller than F. So:

$$S_{ave} = \frac{T^{s}_{ave}}{T^{p}_{ave}} = \frac{n*(p*f+g)+s}{n*(f+g+c)+s} \approx \frac{n*p*f+s}{n*f+s} \approx p \quad (15)$$

Considering the discussed matters, acceleration of a subject is linear and it is reasonable time considering the high volume of data.

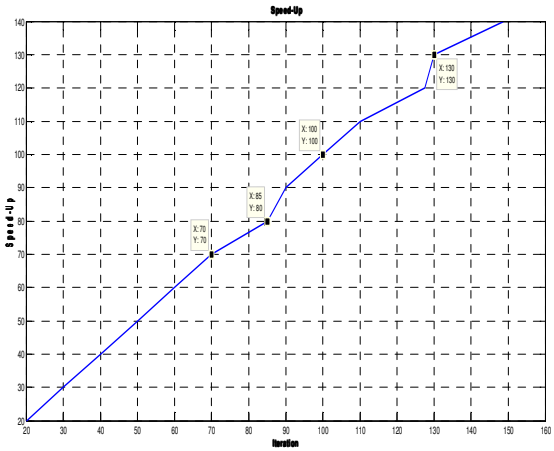



Figure 12- acceleration graph in the algorithm of association rules mining with genetic fuzzy approach

| | Proposed model [23] | Proposed model |
|---|---|---|
| Clustering | Frugal | MOGA clustering |
| Database volume | Big | Big |
| Invader knowledge context | * | * |
| Scalability | * | * |
| Complication | O(n2) | O(n) |

d: Comparison between Proposed model [23] and Proposed model

## X. Conclusion and Suggestions

Generally searching for high repetition items and mining the association rules has a high cost. So in this thesis, for the purpose of reduction of these mining, we used some dividing and paralleling methods to discrete the process of discovering the association rules. We used the algorithm of genetic fuzzy clustering to increase the accuracy in mining the data and we used the parallel approaches to increase the speed. On each transaction data base, the repetitive items were found discretely. By an optimized clustering, the items are chosen in such a way that has the least computational load for the system.

In this study genetic algorithm is used to optimize the fuzzy clustering by quantitative transactions. In this algorithm first the set of membership functions turn into a chromosome and then using the genetic algorithm they turned into the best set of membership functions. These best final sets are used for mining in the fuzzy association rules. The priority of each chromosome is derived from the priority of the set of big items.

In this study we proposed a parallel algorithm of rule mining. This parallel approach is used for solving the problem of time consuming tasks. The duty of this algorithm is to cluster the data and for the authenticity of the cluster we used two methods of Silhouette and Dunn index. The advantage of this study in comparison with the others is that it presents a better method for clustering and uses a parallel technique for computation that gives us a fairly linear time in acceleration. This study is a unique one in the field of encoding stage and the results could be extensive. So there are some suggestions for further research:

1- Using repetitive clustering cloud that considers each customer as granule and gets their data as a static clustering and enters into the data base. In the same way this cloud be used for the clustering of the products that leads to a better clustering.
2- Using the multi objective evolutionary algorithms and classification technology for reaching better and more qualitative rules.
3- Presenting the algorithms of encoding while considering the matters of personal privacy. This leads to a better security of the system and more reduction in the computational overload and time complications.

### References

[1]. Lifei Wei, Haojin Zhu, Zhenfu Cao, Xiaolei Dong, Weiwei Jia, Yunlu Chen, Athanasios V. Vasilakos, “Security and privacy for storage and computation in cloud computing”, Information Sciences 258, Page No( 371–386) ,Jan 2014.

[2]. The NIST Definition of Cloud computing version15,by peter Mell and Grance, october 7 , 2009 ,National Istitute of standards and Technology (NIST) , Information Technology Laboratory (www.csrc.nist.gov)

[3]. I .Sriram ana A.khajeh-hosseini,"Reserch Agenda in Cloud Technologies", Methodology, 2008.

[4]. L.Wang, G.V.Laszewski, A.Younge, X.He, "Cloud Computing: a Perspective Study," Provider, volume -28,Page No(137-146),Feb 2010.

[5]. A. Lenk, M. Klems, J. Nimis, S. Tai, and T. Sandholm, "What's inside the Cloud? An architectural map of the Cloud landscape," 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing, Page No(23-31), May. 2009.

[6]. Pradeep Kumar Tiwari, Dr. Bharat Mishra," Cloud Computing Security Issues, Challenges and Solution", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 2, Issue 8, August 2012.

[7]. Pankaj Sareen," Cloud Computing: Types, Architecture, Applications, Concerns, Virtualization and Role of IT Governance in Cloud", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, March 2013.

[8]. Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," Journal of Internet Services and Applications, Volume-01, Page No(7-18), Apr. 2010.

[9]. Ashutosh Kumar Singh, Dr. Ramapati Mishra, Fuzail Ahmad, Raj Kumar Sagar, Anil Kumar Chaudhary ,"A Review of Cloud Computing Open Architecture and Its Security Issues", International Journal of Scientific & Technology Research Volume-01, Issue-06, ISSN 2277-8616 July 2012

[10]. Kuyoro S.O, Ibikunle F ,Awodele o ,"Cloud Computing Security Issues and Challenges",International Journal Of Computer Network,May 2011

[11]. Flavio Lombardi, Roberto Di Pietro. "Transparent Security for Cloud" SAC '10: Proceedings of the 2010 ACM Symposium on Applied Computing, Page No( 414-415), March 2010.

[12]. Chang-Lung Tsai, 2Uei-Chin Lin,Information Security of Cloud Computing for Enterprises , Advances on Information Sciences and Service Sciences. Volume-03, Number-01, February 2011.

[13]. Waseen Iqbal,Shahid yousaf,"Formal Modeling Of Agent Based Cloud Computing Services Using Petri Nets",VFAST Transactions on Software Engineering,August 2013.

[14]. Sadeghzadeh Payam , Bahrehpoor Davood, Sadeghzadeh Peyman. " Analysis of security challenges in cloud computing", Eighth Symposium advances in science and technology,December 2014.

[15]. Sadralsadati Seyed Mohsen , Karegar Mohammad Javad ," Security challenges in cloud computing and solution to improve security in the development of e-government services" Eighth Symposium advances in science and technology, December 2014.

[16]. Talebi Samira, Khotanloo Hasan ," Check cloud security attacks and strategies to deal with them" The first national workshop on Cloud Computing, Iran, Amirkabir University of Technology, November2012.

[17]. Soltanbaghshahi Somayeh, Soltanbaghshahi Leyla,Khademzadeh Ahmad,Jobehdari Sam." Analysis of security challenges and their effect on cloud computing" The first national workshop on Cloud Computing,Iran, Amirkabir University of Technology, November2012.

[18]. Adaptive, scalable and reliable monitoring of big data on clouds, Mauro Andreolini, Michele Colajanni, Marcello Pietri, Stefania Tosi, Journal of Parallel and Distributed Computing, Available online 26 August 2014.

[19]. Narzu Tarannum and Nova Ahmed, "EFFICIENT AND RELIABLE HYBRID CLOUD ARCHITECTURE FOR BIG DATABASE", International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Volume-03, No-06, December 2013.

[20]. Tilmann Rabl, Mohammad Sadoghi, Hans-Arno Jacobsen, Sergio G ´ omez-Villamor, Victor Munt ´ es-Mulero, Serge Mankovskii, "Solving Big Data Challenges for Enterprise Application Performance Management", Proceedings of the VLDB Endowment, Volume-05, No-12,March 2012.

[21]. Halil Ibrahim Karaca, Migration of an On-Premise Single-Tenant Enterprise Application to the Azure Cloud: The Multi-Tenancy Case Study, Master Thesis, TARTU, 2013.

[22]. Carlo Curino, Evan P. C. Jones, Raluca Ada Popa, Nirmesh Malviya, Eugene Wu, Sam Madden, Hari Balakrishnan, and Nickolai Zeldovich, Relational Cloud: A Database-as-a-Service for the Cloud, In Proceedings of the 5th Biennial Conference on Innovative Data Systems Research (CIDR), Pacific Grove, CA, January 2011.

[23]. Fosca Giannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang," Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases", IEEE Systems Journal, Volume-07, No-03,Page No(385-395), September 2013.

[24]. Chao-Lung Yang , R.J. Kuo, Chia-Hsuan Chien, Nguyen Thi Phuong Quyen , "Non-dominated sorting genetic algorithm using fuzzy membership chromosome for categorical data clustering", Applied Soft Computing 30 Page No(113–122),January 2015.

[25]. http://commons.wikimedia.org/wiki/File:Mobile_Cloud_Architecture.pdf

[26]. Pawan Lingrasa, Ahmed Elagamya, Asma Ammarb, Zied Elouedib" ,Iterative meta-clustering through granular hierarchy of supermarket customers and products", Information Sciences, Volume 257, Pages No (14–31),February 2014.

## AUTHORS PROFILE

Marjan Soltani is Computer engineering master student Department of Computer Engineering Islamic Azad University Of Ashtian, Iran. Her main researches interests are Privacy in computer Sciences and Network.

.

Gholamreza Shahmohammadi received his Ph.D. degree from Tarbiat Modares University (TMU, Tehran, Iran) in 2009 and his M.Sc. degree in Computer Engineering from TMU in 2001. Since 2010, he has been Assistant Professor of information Technology Department at the Olum Entezami Amin University (Tehran, Iran). His main research interests are Software Engineering, Software Architecture, Software Metrics, Software Cost Estimation and Software Security.