International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Privacy Preservation of Healthcare Data in Hybrid Cloud Using a Hybrid Meta-Heuristics Based Sanitization Technique

Sridhar Reddy Vulapula, Srinivas Malladi

Abstract*: Over the recent years, the expansion of cloud computing services enable hospitals and institutions to transit their healthcare data to the cloud, thus it provides the worldwide data access and on-demand high quality services at a cheaper rate. Despite the benefits of healthcare cloud services, the associated privacy issues are widely concerned by individuals and governments. Privacy risks rise when outsourcing personal healthcare records to cloud due to the sensitive nature of health information and the social and legal implications for its disclosure. Over the recent years, a privacy-preserving data mining (PPDM) technique has become a critical issue for the problems. Our goal is to design a privacy-preserving outsourcing framework under the hybrid cloud model. In this work we propose a Hybrid Ant Colony Optimization and Gravitational Search Algorithm (ACOGSA) to express the problem of hiding sensitive data through transaction deletion. Thus, it reduces the side effects of the hybrid cloud. Substantive experiments will be carried to compare the performance of the designed algorithm with the state-of-the-art approaches in terms of the side effects and database similarity (integrity). Over the past to sanitize the databases used for hiding sensitive information, a few heuristic approaches have been proposed. The method used for the comparison involves GA, PSO, ACO, and Firefly framework.

I. INTRODUCTION

Cloud systems are strong computing resources to provide vast storage capacity and large computation power [1], [2]. In data subscription and publishing cloud computing resources are naturally used [3]. Accessing the cloud platform with huge volume of data one of the main issues raised is privacy. To avoid these issues, privacy protection is introduced in sectors like as healthcare and business [4, 5]. With the rapid growth of technology, the healthcare industry transformed in to a big way [3]. Healthcare includes the complex process of prevention of diseases, diagnosis, injury, treatment and other impairments. Informational collections like Electronic Health Records (EHR) in such applications frequently contain protection delicate data, which achieves security concerns conceivably if the data is discharged or shared to outsiders in cloud. To protect some secret information, data sanitization process is used i.e., transformation of original database into a modified database. In data sanitization, the impacts of hiding process based on the item and transaction selection. The main problem occurred in selection is to find an optimal solution

[6, 7]. On the other hand, a Privacy preserving data mining (PPDM) technique is been used for hiding private and difficult information in a dataset. To discover o useful knowledge from big collections of data several data mining techniques are introduced [8]. Several algorithms are found in PPDM to hide sensitive information. Most of these are applying addition and deletion operation to unsettle database and hide sensitive data.

When sensitive data extend over so as to cover partly with important but non-sensitive data, to produce a correct item sets by minimizing the side effect is a major challenge in PPDM algorithms. Sometimes inauthentic information may be produced or hidden data has to be modified due to the time of hiding and securing information. An extension algorithm is suitable for such NP-hard problems [9]. The extension algorithms such as are given below: Genetic Algorithm (GA), Particle Swarm Optimization (PSO) or Ant Colony Algorithm (ACO) could be used to reach the best solution. But these methods find the solutions with some side effects. For privacy preservation on cloud, a dyadic product and crow lion algorithm is developed by Ashok George and Sumathi. Also for the construction of database privacy, a dyadic square matrix is been developed using the dyadic product of vectors sensitive utility (SU) coefficient and the cumulative data key product. C-lion algorithm is used for selection of SU coefficient vector [3]. Data anonymization refers to hiding identity or sensitive data of user, while in diverse analysis and mining tasks some certain information can be still exposed to data user [4]. A multi-agent architecture was designed for privacy preservation and data sharing. This framework consists of federation of agents that could be customized based on the data source characteristics and heterogeneity of data. But this framework worked well in limited framework [7].

To overcome the difficulties and reducing the side effects obtained in preservation we proposed a hybrid model for private preservation. In the work, A hybrid ant colony optimization and gravitational search algorithm (ACOGSA) is used to reduce the side effects in hybrid cloud as well as to notify the problem in the hidden sensitive data through transaction deletion.

Manuscript published on November 30, 2019.
* Correspondence Author
Sridhar Reddy Vulapula*, Department of CSE, KL University, Andhra Pradesh, India, Email: vsridharreddy0102@gmail.com
Dr Srinivas Malladi, Department of CSE, KL University, Andhra Pradesh, India, Email: srinu_cse@kluniversity.in

In our work, a set of Electronic Medical record (EMR) are taken and applied to cloud. Transaction database are extracted from the record and the sanitization process is executed based on proposed technique. The sanitized information is accessed by user from cloud.

The rest of the paper is depicted as follows: Section 2 talks about the related work. Section 3 consists of our proposed work and the brief description. The proposed work performance results and the relevant discussions are given in section 4 and the paper is concluded in section 5.

II. RELATED WORK

Abdul majeed [10] implemented an attribute centric anonymization to improving the privacy of e-health data. Their approach consists of following steps such as preprocessing the original health records, user ranking, formation and analysis of equivalence classes, attributes classification and data anonymization. Simulation result shows that the data privacy is improved than existing methods.

Peng cheng et al. [11] introduced a rule hiding method for privacy preservation. Their approach includes two phases for protection. First phase followed the Apriori algorithm for determining the frequency pattern and corresponding association rule. Second phase was reducing sensitive rules by relevance sorting algorithm. For experimental analysis it hides sensitive rules and data loss compared with SIF-IDF and it is a time consuming process.

In 2017 Kalyani et al. [12] designed an adapted binary firefly algorithm for privacy preserving classification Rule mining for balance data utility. In their method initially a set of classification rules would be invented with a classification algorithm from original data. Using a set of identified sensitive classification rules as input then their method transforms the original data. Their approach hides the sensitive rule by selection of best possible transactions.

For hiding sensitive datasets, Chun wei lin et al. [13] presented a GA based algorithm, in which it initially defines the type of Chromosome used to represent the possible solutions and then it has been determined with the help of a designed fitness function. The major operations used in their algorithm are selection, crossover and mutation. Their approach hides dataset through deletion. For performance analysis it has been used for variation of NP hard problems and related applications.

Lin et al. [8] implemented a particle swam optimization algorithm for hiding sensitive item sets. In their approach firstly calculated the no of transaction to be deleted. Then initialize each particle and evaluate fitness. They determines the global best (gbest) and the personal best (pbest), then choose gbest as a solution when the condition was satisfied else, updating velocity and position for fitness evaluation. Experimental analysis shows that the method was much faster and generates little N-T-H side effects, when compared to other existing methods.

Fabian Prasser et al. [14] introduced a scalable and pragmatic method for safe sharing of health data. Their approach follows the onion-skin principle for privacy. Their method consists of accompanying steps like as implementing the original data, analyze its properties and finally de-identification method. When the population characteristic was not known then the pragmatic solution was implemented. It was a highly scalable method compared to other procedures.

In 2017 Wu et al. [9] implemented an ant colony system for hiding sensitive item sets. From the original database during the process, each ant would build a tour for iteration and every tour represents a deleted transaction. Their approach includes following steps Ant routing map, Termination condition for each tour, Fitness function, Heuristic function and Delete transactions. Simulation results perform in more complicated computation.

Sabin Begum and Sugmar [15] presented a novel entropy based approach for cost effective privacy preservation. In their method adaptive particle swarm optimization was used in optimal entropy value process. Their technique consists of joint entropy model and database difference model. For testing purpose, evaluation metrics was taken from entropy and database difference. In performance their method has minimum access time, better memory usage and good privacy.

A. Problem Statement

In hybrid cloud, there is need for protect information of users. The following are the some pitfalls of most privacy protection methods. They are, Third party interfering that is API (Interfaced application infection), Loss of data in mass storage area, lack of privacy at service provider and user/client level, disclosure of sensitive private information and unauthorized access to personal data. Due to hiding sensitive items in a database and produce a sanitized database, the problem is to decrease the sensitive item set $S_i \in si$ support count such that its support count become less than the minimum support count and it is given as, $\sup(S_i) < \delta * |d|$

To find the quality of the approach the side effect measures are utilized. If the number of F-T-H is very large, it means that more sensitive item sets are still present in the sanitized database. If N-T-H is very high, it denotes that there is some necessary decision making information may be missing in the final database. If N-T-G is very large, it represents that some number of unwanted artificial information may be introduced in the sanitized database. In the transactions deletion, the number of transactions changes and also the minimum support count. So the infrequent item sets become frequent in the end of process. Thus there is a trade-off relationship between the three side effects (F-T-H, N-T-H and N-T-G). An NP-hard problem is used to determine the best solution for hiding and minimizing the side effects.This journal uses double-blind review process, which means that both the reviewer (s) and author (s) identities concealed from the reviewers, and vice versa, throughout the review process. All submitted manuscripts are reviewed by three reviewer one from India and rest two from overseas. There should be proper comments of the reviewers for the purpose of acceptance/ rejection. There should be minimum 01 to 02 week time window for it.

III. PROPOSED METHODOLOGY

A. Preliminaries

Preliminaries and hiding sensitive data item definitions in diminishing the side effects are introduced in this section.

Let $i = \{I_1, I_2, I_3, \ldots, I_x\}$ be a finite set of x distinct items. A database $d = \{t_1, t_2, \ldots, t_n\}$ is a set of transactions. For each transaction $t_q \in d$, t_q is a subset of *i*, and q is a unique identifier for t_q. For a specific transaction this identifier is called Transaction Identifier (TID). Assume a minimum support threshold δ is set by user or experts manually. An item set *I* is a subset of *i*. The support count of an item set is the number of transactions present in the item set. The minimum support count represents the product of the minimum support threshold and the number of transactions in the database.

The frequent item sets is expressed as $fi = \{F_1, F_2, \ldots, F_k\}$ and support count of a frequent item set ($\sup(F_i)$) is greater than the minimum support count in the database and is given as $\sup(F_i) \geq |d| * \delta$. The set of sensitive item sets is represented as $si = \{S_1, S_2, \ldots, S_p\}$ and it is a subset of frequent item sets that is $si \subseteq fi$ and it can be specified by users or experts.

Using our proposed method, the goal is to find a sanitized database d' from a database d has been obtained by removing some transactions/item sets in an original database. The main aim of PPDM is to hide much sensitive information of database in the cloud so the preserved data cannot be discovered using data mining techniques. To hide the sensitive item sets, it is important to remove transactions containing sensitive item sets. However, finding a set of transactions or item sets to be deleted that minimize side effects is a NP-hard problem. Three important side effects are the failure to hide some sensitive information (called fail to be hidden, F-T-H, or hiding failure), hiding information that is important but not sensitive (called not to be hidden, N-T-H, or missing cost), and the introduction of artificial information (called not to be generated, N-T-G, or artificial cost).

The side effects in F-T-H is been represented as β , and is defined as the amount of sensitive item sets present in the sanitized database d', that is: $\beta = |si \cap fi'|$. The side effect of N-T-H is represented as γ , and it is defined as the number of hidden non-sensitive item sets in the sanitized database d', that is: $\gamma = |\sim si - fi'| = |(fi - si) - fi'|$. The side effect in N-T-G is represented as α , is defined as the number of frequent item sets in the new database d' that is: $\alpha = |fi' - fi|$

B. System Model of Proposed Work

The architecture of our method is as shown in figure 1. Our model consists of three main parts such as data owner, hybrid cloud and authorized user. Hybrid cloud is a combination of private and public cloud. The data owner may be medical institution or organization. Before upload the database to cloud, data owner protect the health data from unauthorized user to access data. The access control strategy has been used to prevent the unknown user. Initially data owner would be applying the data to the private cloud. Sanitization process of database would be done in the inter relation of clouds. Finally the authorized users access the sanitized data from the public cloud.

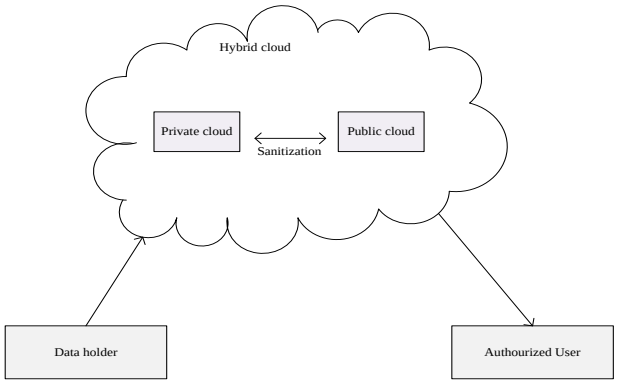

Fig.1. System model

C. Proposed ACOGSA Approach

This section described the proposed ant colony optimization and gravitational search algorithm (ACOGSA) in hybrid cloud for hiding sensitive item sets through transaction deletion. The proposed work follows the traditional Ant Colony Optimization (ACO) and Gravitational search Algorithm (GSA). An ant leaves the pheromone in case of searching food; other ants follow the path according to the concentration of pheromones. By placing pheromones on the path, the ants find the little path from nest its destination. In our proposed work ACO is utilized to find the suitable solutions. In our work, a heuristic function is introduced using pheromone information to select a direction. The concept of GSA was used to update the pheromone that is to find the best solution. Computing the values of force, mass and updating velocity, position of the agent (suitable solution) the best solution would be determined. The brief step of our proposed work is given as follows:

1) Initialization: In this step, first initialize a projected database d^* from the original database d . To hide sensitive item sets through deletion operation, the transaction must contain at least one sensitive item set for deletion. The projected database is a subset of transactions from the original database, which contains at least one sensitive item set. It can be given as,

$$d^* \leftarrow \left\{ t_q \middle| t_q \in d, \exists S_i \in HS, S_i \subseteq t_q \right\} \quad (1)$$

2) Ant routing graph: The method does not map all the transactions, in order to diminish the search space from the routing space. So from the projected database the routing graph is encoded. In Ant routing graph each node denotes the transaction in the dataset and there is no proper destination. The graph is shown in figure 2. In the design all the nodes are edge connected to the nest node and they do not connect with each other.

An ant is been placed on the nest node (center node) at the beginning of each tour (Solution). The ant is choosing an edge and return back to the nest node. An ant selects various nodes (transactions) and thereby reaches the termination criteria after completing the tour. Thus, the completed tour of an ant represents the subset of the input transaction dataset.

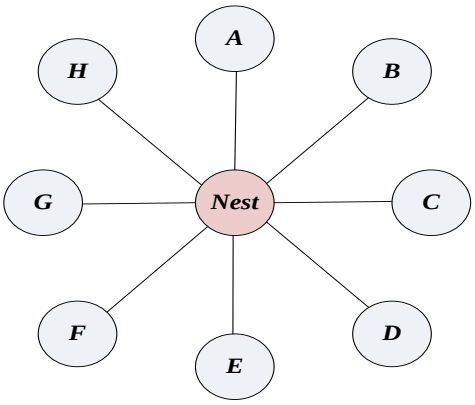

Fig. 2. Ant routing graph

The termination condition of each tour as given as below: Our graph has no destination so maximum number of selected transaction would be set in ACOGSA. The proper number of transactions is been defined to denote the maximum number of deleted transactions. It would be described as follows:

$$M = N * \left(\left\lfloor \frac{Max_{S_i \in HS} \sup(S_i) - \delta * |d|}{1-\delta} \right\rfloor + 1 \right) \tag{2}$$

Theoretically, $\left\lfloor \frac{Max_{S_i \in HS} \sup(S_i) - \delta * |d|}{1-\delta} \right\rfloor + 1$ is the minimum number of selected transactions and this process hides the entire sensitive item sets. N is a number which is greater than 1 and it is suitable for all different types of dataset. If the process that always selects a transaction containing the sensitive item sets with the highest support counts is deleted, thus the number N can be set to approach 1. If it cannot, it needs to be set as a number greater than 1, in order to hide the entire sensitive item sets.

After finding the minimum number of selected transaction, the method makes the comparison of heuristic function with current fitness function. In heuristic function there is no node better than the fitness function, thus the selection process is been disturbed and the ant thus finishes the tour.

3) Heuristic Function: Heuristic function is been used to guide an ant that searches for a better direction. Thus based on the current situation, our system chooses a suitable transaction to adjust or enhance the current solution. Assume a parameter p in which it represents the number of non-hidden transactions that are presented in the selected transaction. The hidden item sets in the selected transactions is denoted as q and r is the number of fake item sets in the transaction. Then the value of heuristic function is described as follows:

$$h(e) = p * \beta - q * \gamma + r * \alpha \tag{3}$$

Where

h Denotes the heuristic function

e Denotes the selected transactions

β Denotes the number of F-T-H item sets

γ Represents the number of N-T-H item sets

α Represents the number of N-T-G item sets.

Using this function we can select the suitable transaction to adjust the current solution and to reduce the side effects. In our system, the ants can refer the selected transaction and thereby selects the next transaction using this function.

4) Pseudo random proportional rule: Probabilistically the state transaction rule is used to find the ant's next node. For example, take the traveling salesman problem (TSP). Let us imagine that the K[th] ant is presently in the city u (node). The next city S (node) visited by the K[th] ant is been given as follows:

$$S = \begin{cases} \arg N \in \vec{r}_K(u) \max \left\{ [\tau(u,N)]^x * [\eta(u,N)]^y \right\} \\ ,if\ Q \le Q_0 \\ i\ with\ a\ probability\ p_k(u,v) \\ ,if\ Q \le Q_0 \end{cases} \tag{4}$$

where

$r_k(u)$ Denotes the set of cities that can be visited by the K[th] ant.

$\tau(u,N)$ Is the accumulated pheromone on the edge between city u and N.

x and y are the two parameters that indicate the relative influence of pheromone versus the distance between two cities in this problem.

Q indicates the random number uniformly distributed between 0 and 1.

Q_0 is a parameter usually set as a number close to 1.

$p_k(u,v)$ Represents the probability from city u to city v if $Q>Q_0$.

$$p_k(u,v) = \begin{cases} \dfrac{[\tau(u,v)]^x * [\eta(u,v)]^y}{\sum_{N \in r_K(u)} [\tau(u,N)]^x * [\eta(u,N)]^y} \\ ,if\ u \in r_K(u) \\ 0 \\ ,otherwise \end{cases} \tag{5}$$

This equation is known as random proportional rule and the combination of this selection process is called as pseudo random proportional rule. In basic, the heuristic function $\eta(u,v)$ is related to optimization.

5) Local updating rule: To adjust the density of the pheromones, local updating rule is used. It is always used to avoid the selected similar tours in the iteration.

When an ant selects an edge between the nodes u and S, the pheromone density is updated to avoid the local optimum. It can be described as follows,

$$\tau_{after}(u,S) = (1-\rho) * \tau_{before}(u,S) + \rho * \tau_0 \quad (6)$$

where $\tau_{before}(u,S)$ denotes the pheromone amount on the edge from city *u* to *S*.

For this equation, the pheromone density on the structured tour is greater than the initial pheromone; the rule diminishes the density of the selected edge to reduce the repeated visiting, on the other hand it increases the pheromone density. This step is used to prevent the ant population from selecting same paths and make sure that other possible paths have a better opportunity to be selected.

6) Fitness function: The fitness function is been employed to estimate every tour (solution) produced by the ants. The fitness function represents the weighted sum of three side effects. It is described as follows:

$$fitness = W_1 * \beta + W_2 * \gamma + W_3 * \alpha \quad (7)$$

where

β Denotes the number of F-T-H item sets

γ Represents the number of N-T-H item sets

α Is the number of N-T-G item sets.

W_1, W_2, W_3 is the weights of the each side effect that is adjusted by user.

Basically, all wish to hide much sensitive data rather than the non-sensitive data that are missed or some artificial data can be introduced. So W_1 set higher than the other two weighted values. The depth of the database has an influence on the fitness tour. To get the best balance between the side effects W_1 set as higher in dense database because the data distribution is condensed in dense database. Conversely, data distribution is scarce in sparse database so W_1 set as lower than other two values.

7. Update Pheromone: To find the best tour (solutions) from all executed iteration or determine the best solution from current iteration, the Gravitational Search Algorithm (GSA) [16] is applied. The steps of GSA are given below:

- Identify the search space
- Initialize randomly- solutions are assumed as agents
- Evaluate fitness of agents- minimization of all solutions
- Update the best and worst solution- This values calculated by using following equations:

$$best(T) = \min_{j \in \{1,\ldots,n\}} fit_j(T) \quad (8)$$

$$worst(T) = \max_{j \in \{1,\ldots,n\}} fit_j(T) \quad (9)$$

- Calculate force in different directions

$$f_i^d(T) = \sum_{j \in KBEST, j \neq i} rand_j f_{ij}^d(T) \quad (10)$$

Set of first K agents with best fitness value is denoted as *KBEST*.

- Computation of acceleration and mass- This values can be evaluated as follows: acceleration,

$$A_i^d(T) = \frac{f_i^d(T)}{m_{ii}(T)} \quad (11)$$

The internal mass can be calculated as,

$$M_i(T) = \frac{fit_i(T) - worst_i(T)}{best(T) - worst(T)} \quad (12)$$

$$m_i(T) = \frac{M_i(T)}{\sum_{j=1}^{n} M_i(T)} \quad (13)$$

- Update velocity and position- This can be determined using following equations,

$$V_i^d(T+1) = rand_i * V_i^d(T) + A_i^d(T) \quad (14)$$

$$X_i^d(T+1) = X_i^d(T) + V_i^d(T+1) \quad (15)$$

Where, $rand_i$ is the uniform random variable in the interval [0,1]

- Meet the criterion return the best solution
- Otherwise repeat the steps

The pseudo code of the ACOGSA is given as below:

D. Algorithm

Algorithm: ACOGSA

Input: Set of projected transactions (d^*), set of sensitive item sets to be hidden (*si*), frequent Item sets in d^* (*fi*), minimum support threshold (δ), represents the number of ants used in every iteration.

Output : A new database (d') is formed.

Step 1: Initialize a projected database

Step 2: Construct an ant routing graph

Step3: Compute the maximum number of selected transaction

Step 4: While ants did not built their solution

do

select an ant did not complete its tour

set best solution of this tour denoted by $g_b = \varphi$

Step 5: if the amount of selected transactions = M then

an ant complete the tour earlier, thus the solution generated by an ant is denoted by g_b, then update the global best solution.

continue

end if

Step 6: Calculate the heuristic function

Step 7: Using pseudo random proportional rule

Step 8: Use pheromone local updating rule

Step 9: Put current possible solution as c_s

Step 10: Evaluate the fitness value $f(c_s)$ for the current possible solution

Step 12: if $f(c_s) > f(g_b)$ then

$g_b \leftarrow c_s$

end if

Step 13: Otherwise go to the step 5

Step 14: end while

Step 15: Set g=g+1

Step 16: if $g = G$ then

return G_b

end if

Step 17: Process the pheromone updating using GSA
Step 18: Repeat the steps from line 4 for next iteration

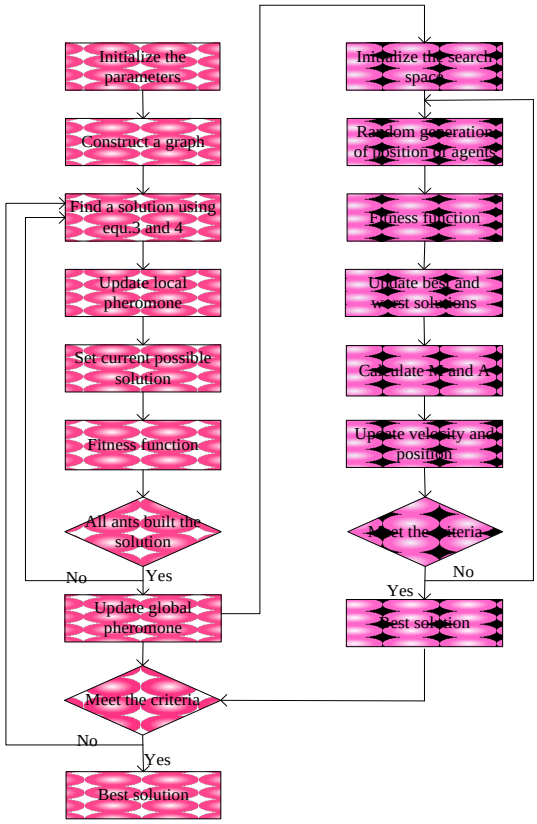

Fig. 3. Flowchart of ACOGSA approach

Therefore, the system gives the best solution for transaction deletion and the sanitized database is produced. The proposed hybrid system structure is organized in the following figure 3. Section 6 explains the execution of proposed method and the results are compared with the state-of-art methods.

IV. RESULTS AND DISCUSSION

The experimental section is done by using the proposed method ACOGSA approach. We have performed the experiment on a Personal computer with Intel® core(TM) i3-7100, Dual core CPU, Dual core 3.90 GHz processor, 4GB RAM, 64 bit windows, 7 Operating System and for implementation the JAVA software tool is used. The experimental results have been compared with the Ant colony system-based algorithm (ACS2DT) [17], Particle swarm optimization-based algorithm (PSO2DT) [18] techniques.

A. Evaluation Metrics

The performance evaluation is done for the sensitive percentage and minimum support threshold with the runtime, F-T-H (Fail to be hidden), N-T-H (Not to be hidden), and DS (Database similarity).

1) Fail To be Hidden (F-T-H): F-T-H is defined as the number of sensitive item sets from the original database to the number of sensitive item sets that are presented in the sanitized database. It is defined in mathematically as

$$F-T-H = \frac{|SI^*|}{|SI|} \quad (16)$$

Where the sensitive item sets in the original database is denoted as SI^* and the sensitive item sets still appeared in the sanitized database is given as SI.

2) Not To be Hidden (N-T-H): N-T-H is defined as the number of non-sensitive frequent item sets that are masked during the sanitization process. It is given as,

$$N-T-H = \frac{|FI-SI-FI^*|}{|FI-SI|} \quad (17)$$

It can be expressed as $DS = \frac{|D^*|}{|D|}$

B. Performance Evaluation for Sensitive Percentage

The comparison of the proposed with the existing schemes is analyzed. The runtime of the GS-ACO algorithm is shown in the fig 4. While the runtime gets increased the sensitive percentage of our proposed gets better value in the runtime of 4900 sec.

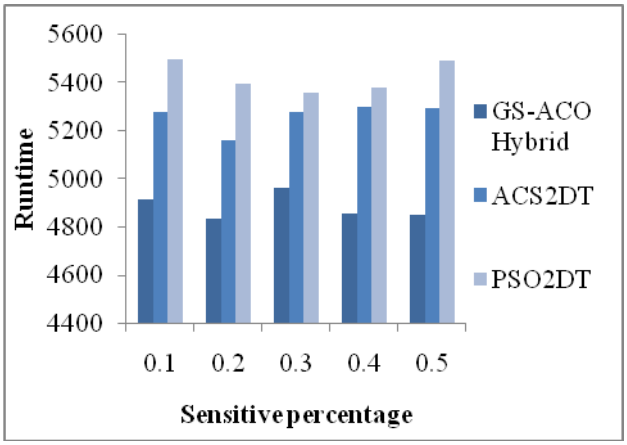

Fig. 4 Comparison of Runtime with Sensitive percentage

From the figure 4 the sensitive percentage is compared for the proposed method with the ACS2DT [17], PSO2DT [18]. The sensitive percentage is maximized for the proposed method and the other two methods are having high sensitive. For the runtime increases the sensitive percentage is minimized. The overall percentage of the proposed method is 6.28% better of ACSDT and 10.32% better of PSO2DT.

From the figure 5 the sensitive percentage is compared for the proposed method with the ACS2DT [17], PSO2DT [18]. The sensitive percentage is maximized for the proposed method and the other two methods are having high sensitive. For the F-T-H increases the sensitive percentage is minimized. The percentage of the proposed is 8.41% for the 0.1 sensitive percentage, 7.08% for the 0.2 sensitive percentage, 8.17% for the 0.3 sensitive percentage, 7.16% for the 0.4 sensitive percentage, 9.57% for 0.5 sensitive percentage better of the ACSDT.

While comparing to the PSO2DT the proposed gets better values for the 0.1 sensitive percentage is 6.88%, for 0.2 the sensitive percentage is 4.8%, for 0.3 the sensitive percentage is 8.82%, for 0.4 the sensitive percentage is 8.16%, for 0.5 the sensitive percentage is 9.11%.

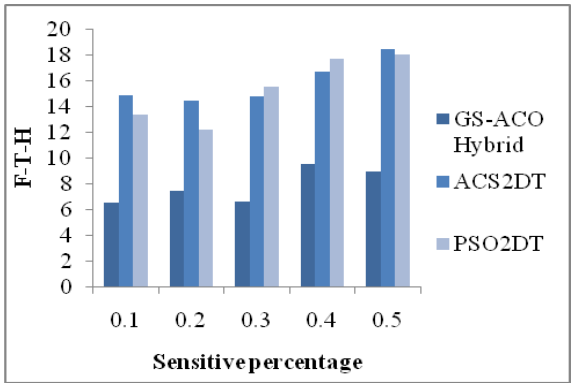

Fig. 5. Comparison of F-T-H with Sensitive percentage

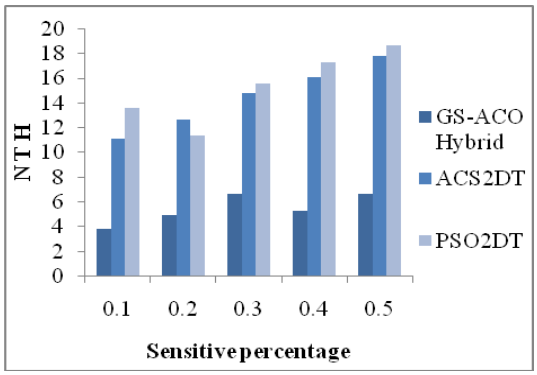

Fig. 6. Comparison of N-T-H with Sensitive percentage

From the figure 6 the sensitive percentage is compared for the proposed method with the ACS2DT [17], PSO2DT [18]. The sensitive percentage is maximized for the proposed method and the other two methods are having high sensitive. For the N-T-H increases the sensitive percentage is minimized. The proposed method gets better of ACS2DT in the percentage of 7.29%, 7.67%, 8.17%, 10.71%, and 11.15% for all the sensitive values. While comparing with the PSO2DT the proposed gets 9.72%, 6.45%, 8.92%, 11.93%, and 12.01% for all the sensitive values.

From the figure 7 the sensitive percentage is compared for the proposed method with the ACS2DT [17], PSO2DT [18]. The sensitive percentage is increased for the proposed method and the other two methods. For the DS increases the sensitive percentage is also increased. While comparing with the ACS2DT the proposed method gets better as 1.58%, 1.43%, 1.5%, 1.71%, and 1.25% for all the sensitive percentages and comparing with the PSO2DT the proposed results are improved in percentage of 2.39%, 2.09%, 1.99%, 1.84%, and 1.54% for all the sensitive values.

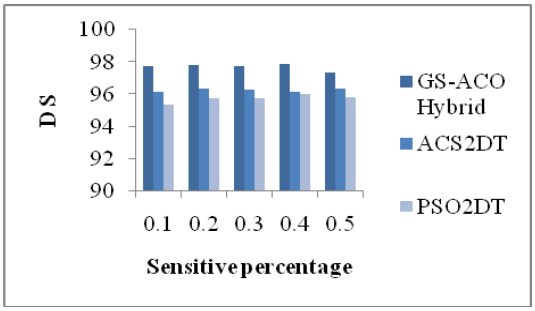

Fig. 7. Comparison of DS with Sensitive percentage

C. Performance Evaluation for Minimum Support Threshold

The comparison of the proposed with the existing schemes is analyzed. The runtime of the GS-ACO algorithm is shown in the fig 8. While the runtime gets increased the sensitive percentage of our proposed gets better value in the runtime of 4950 sec.

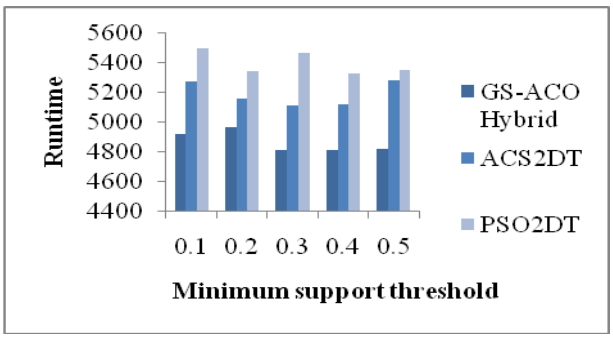

Fig. 8. Comparison of Runtime with Minimum Support threshold

The figure 8 shows the minimum support threshold is compared for the proposed method with the ACS2DT [17], PSO2DT [18]. The minimum support threshold is maximized for the proposed method and the other two methods have been increased. For the runtime increases the minimum support threshold is minimized. The proposed gets the better value when comparing with the ACS2DT as 7.1%, 3.8%, 5.9%, 6.1%, and 8.8% for all the sensitive values. While comparing with PSO2DT the proposed values gets as 10%, 7.1%, 12%, 9.8%, and 9.9% better values in all sensitive percentages.

From the figure 9 the minimum support threshold is compared for the proposed method with the ACS2DT [17], PSO2DT [18]. The minimum support threshold is maximized for the proposed method and the other two methods have been increased. For the F-T-H increases the minimum support threshold is minimized. The proposed values is better in the ACO2DT the values are 8.41%, 8.84%, 7.92%, 11.93%, 8.87% for all the minimum threshold values. While comparing with the PSO2DT, the proposed method shows better of 6.8%, 5.25%, 6.62%, 9.9%, and 8.83% to all the minimum threshold values.

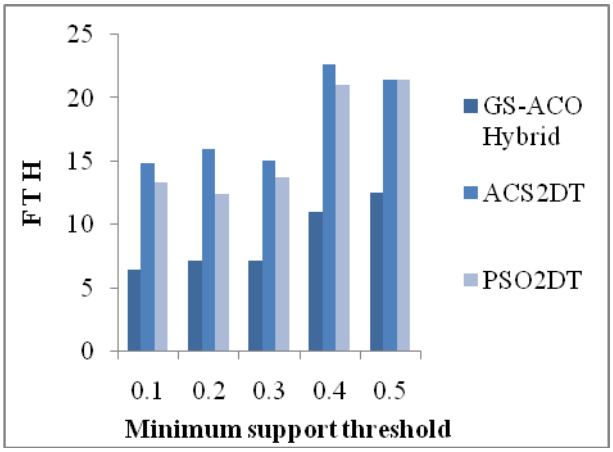

Fig. 9. Comparison of F-T-H with Minimum Support threshold

From the figure 10 the minimum support threshold compared for the proposed method with the ACS2DT [17], PSO2DT [18]. The minimum support threshold is maximized for the proposed method and the other two methods have been increased. If N-T-H increases, the minimum support threshold decreases. The proposed method gets better of ACS2DT in the percentage of 7.29%, 8.69%, 8.18%, 10.02%, and 8.54% for all the minimum support threshold values. While comparing with the PSO2DT the proposed gets 9.72%, 9.35%, 9.9%, 8.72%, and 8.91% for all the minimum support threshold values.

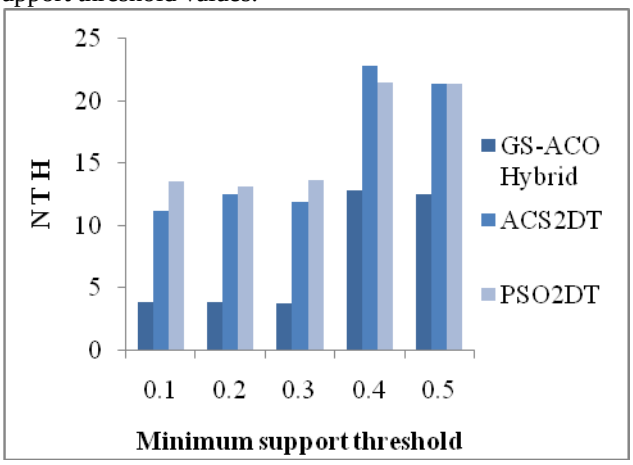

Fig. 10. Comparison of N-T-H with Minimum Support threshold

From the figure 11 the minimum support threshold is compared for the proposed method with the ACS2DT [17], PSO2DT [18]. The minimum support threshold is maximized for the proposed method and the other two methods have been increased. For the DS increases the minimum support threshold is minimized. The proposed method gets better of ACS2DT in the percentage of 1.58%, 0.86%, 1.4%, 1.1%, and 0.87% for all the minimum support threshold values. While comparing with the PSO2DT the proposed gets 2.39%, 1.71%, 2.13%, 1.55% and 2.08% for all the minimum support threshold values.

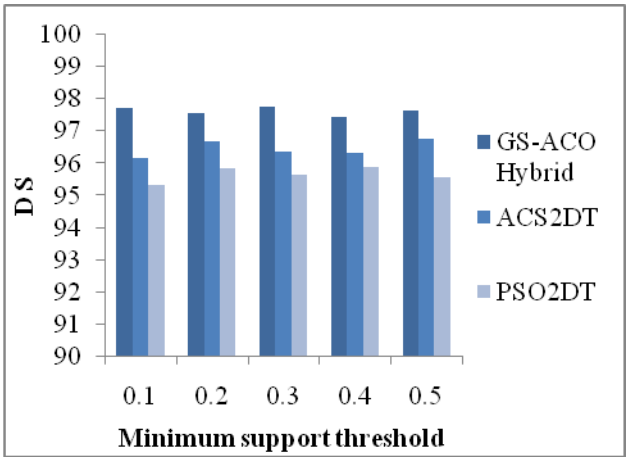

Fig. 11. Comparison of DS with Minimum Support threshold

VI. CONCLUSION

In this paper, a hybrid technique is been introduced in the hybrid cloud for privacy preservation in the healthcare data. The proposed hybrid method is based on the ACO and GSA. The objective of our work is enhancing the privacy policy and reduces the side effects in the existing systems. Using the combination of both techniques our system generates the sanitized database via transaction deletion. Comparing to existing methods our proposed work has better privacy and security. The experimental results are compared for the proposed with the existing methods as ACS2DT and PSO2DT. In the method, the runtime and minimum support threshold is been performed with the runtime, F-T-H, N-T-H, and DS.

REFERENCES

1. A. Amiri, "Dare to share: Protecting sensitive knowledge with data sanitization," Decision Support Systems, vol. 43, no. 1, 2007, pp. 181-191.
2. A. George, and A. Sumathi, "Dyadic product and crow lion algorithm based on coefficient generation for privacy protection on cloud," Cluster Computing, 2018.
3. H. Kaur, N. Kumar, and S. Batra, "An efficient multi-party scheme for privacy preserving collaborative filtering for healthcare recommender system," Future Generation Computer Systems, vol. 86, 2018, pp. 297-307.
4. X. Zhang, W. Dou, J. Pei, S. Nepal, C. Yang, C. Liu, and J. Chen, "Proximity-Aware Local-Recoding Anonymization with Map Reduced for Scalable Big Data Privacy Preservation in Cloud," IEEE Transactions on Computers, vol. 64, no. 8, 2015, pp. 2293-2307.
5. Z. Cai, Z. He, X. Guan, and Y. Li, "Collective Data-Sanitization for Preventing Sensitive Information Inference Attacks in Social Networks," IEEE Transactions on Dependable and Secure Computing, 2016, pp. 1-1.
6. A. Telikani, and A. Shahbahrami, "Optimizing association rule hiding using the combination of border and heuristic approaches," Applied Intelligence, vol. 47, no. 2, 2017, pp. 544-557.
7. H. Wimmer, V. Yoon, and V. Sugumaran, "A multi-agent system to support evidence based medicine and clinical decision making via data sharing and data privacy," Decision Support Systems, vol. 88, 2016, pp. 51-66.
8. J. Lin, Q. Liu, P. Fournier-Viger, T. Hong, M. Voznak, and J. Zhan, "A sanitization approach for hiding sensitive item sets based on particle swarm optimization," Engineering Applications of Artificial Intelligence, vol. 53, 2016, pp.1-18.

9. J. Wu, J. Zhan, and J. Lin, "Ant Colony System Sanitization Approach to Hiding Sensitive Item sets," IEEE Access, 5, 2017, pp.10024-10039.
10. A. Majeed, "Attribute-centric anonymization scheme for improving user privacy and utility of publishing e-health data," Journal of King Saud University - Computer and Information Sciences, 2018.
11. P. Cheng, J. Roddick, S. Chu, and C. Lin, "Privacy preservation through a greedy, distortion-based rule-hiding method," Applied Intelligence, vol. 44, no. 2, 2015, pp. 295-306.
12. G. Kalyani, M. Rao, and B. Janakiramaiah, "Privacy-Preserving Classification Rule Mining for Balancing Data Utility and Knowledge Privacy Using Adapted Binary Firefly Algorithm," Arabian Journal for Science and Engineering, 2017.
13. C. W. Lin, T. P. Hong, K. T. Yang, and S. L. Wang, "The GA-based algorithms for optimizing hiding sensitive item sets through transaction deletion," Applied Intelligence, vol. 42(2), 2015 , pp. 210– 230.
14. F. Prasser, F. Kohlmayer, H. Spengler, and K. Kuhn, "A Scalable and Pragmatic Method for the Safe Sharing of High-Quality Health Data," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 2, 2018, pp. 611-622.
15. R. Sabin Begum, and R. Sugumar, "Novel entropy-based approach for cost-effective privacy preservation of intermediate datasets in cloud," Cluster Computing, 2017.
16. E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "GSA: A Gravitational Search Algorithm," Information Sciences, vol. 179, no. 13, 2009, pp. 2232-2248.
17. W. Jimmy, M. Tai, J. Zhan, and J. Lin. "Ant colony system sanitization approach to hiding sensitive item sets." IEEE Access, vol. 5, 2017, pp. 10024-10039.
18. Lin J, C. Wei, et al. "A sanitization approach for hiding sensitive item sets based on particle swarm optimization," Engineering Applications of Artificial Intelligence, vol. 53, 2016, pp. 1-18.

AUTHORS PROFILE

Mr. Sridhar Reddy Vulapula, M.Tech in Software Engineering from Jawaharlal Nehru Technological University and B.Tech degree in CSIT from Jawaharlal Nehru Technological University, and Research Scholar in CSE Department from K L University, Vaddeswaram . He has 12 years experience in teaching ,published peer-reviewed papers in Accredited and impact factor journals like scopus. Life member in professional bodies like CSTA.

Dr. Srinivas Malladi, Doctorate in Computer Science & Engineering from Koneru Lakshmaiah (KL) University and Master's degree in CSE from Nagarjua University, India. working with K L University, Vaddeswaram as a professor in Dept of CSE . He has 20+ year's experience in teaching ,published more than fifteen peer-reviewed papers in Accredited and impact factor journals like scopus. Life member in professional bodies like CSI,ISCA & ISTE