Indian Journal of Science and Technology, *Vol 9(28), DOI: 10.17485/ijst/2016/v9i28/93817, July 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Conditional Entropy with Swarm Optimization Approach for Privacy Preservation of Datasets in Cloud

R. Sabin Begum[1*] and R. Sugumar[2]

[1]Research and Development Centre, Bharathiar University, Coimbatore - 641046, Tamil Nadu, India; sabinsabiya@gmail.com

[2]Department of Computer and Engineering, Velammal Institute of Technology, Chennai - 601204, Tamil Nadu, India; sugu16@gmail.com

Abstract

Background/Objective: The primary intension is to provide utility trade off and good privacy for intermediate datasets in cloud. **Methods:** An efficient conditional entropy and database difference ratio is employed for the process. Utility is taken care with the employment of conditional entropy with the help of Swarm Optimization (PSO). Privacy handled by database difference ratio. **Findings:** Conditional entropy is found out between the first column and the original database and this is taken as the fitness function in Particle Swarm Optimization (PSO). Database difference is taken between the original database and convoluted database in the consequent module to yield right selection data and tuple information from the intermediate datasets in cloud. **Applications/Improvements:** Improved methods are required to provide privacy and utility for the right selection of datasets. This approach has better results by having higher entropy values and lower dataset difference ratio.

Keywords: Cloud Computing, Conditional Entropy, Database Difference Ratio, Intermediate Datasets, Particle Swarm Optimization (PSO), Privacy Preservation

1. Introduction

The cloud computing has appeared with a big bang as a novel and sophisticated technology in the appealing arena of Information Technology and has become the heart-beat of IT enterprises, with ample potential for advancement in the days ahead[1]. It is endowed with the vision of computing as an efficiency and has the innate skills of creating a sea change in the IT industry as a whole, projecting software as a highly alluring service[2]. It is self-sufficient though it varies by grid and utility computing as a whole[3]. The novel technology offers a means permitting immense controlled sharing and interoperation between resources owned and managed[4]. It may be broadly categorized into three levels such as the infrastructure layer, platform services layer, application layer software[5]. It has affected a paradigm shift in the information technology sector[6], which gives as a service[7]. The novel technology employs three delivery models for offering several kinds of services to the end user such as the SaaS, PaaS and IaaS which elegantly provide the infrastructure resources, application platform and software as services to the client[8].

The safety aspect in any cloud computing infrastructure is the highly significantly element, in view of the fact that authorized access can only be accepted and safe behavior adequacy is highly essential[9]. There are three important features to be taken well-care of in respect of cloud security are detailed below: 1. Cloud security is exactly in the same footing as the internal safety. The safety gadgets are deployed to safeguard the internal network cloud, and the data in the cloud. 2. Certain safety technologies have to be shifted to the cloud for the purpose of achieving enduring economic competitiveness. 3. In a large

*Author for correspondence

majority of the cases it is high time the current safety is augmented with the intention of short-listing a quality cloud service provider which is superior to the current security[10].

The cloud security is identical for both the cloud provider and the cloud consumer, both having due faith in the association as they tend to harmonize each other when it emerges as confidential data at rest and also during transit[11]. The trust is a vital requirement for the consumers and services provider, who take part in a cloud computing scenario. As, cloud computing embraces several local techniques and draws several members from diverse backdrops, it becomes highly intricate[12].

In the domain of cloud computing, a privacy-preserving public auditing technique for data storage security was proficiently proposed[13]. Devoid of the pile of local data storage and preservation, clients were able to uncertainly attain their data and realize the magnificent applications and administrations from an imparted group of configurable figuring assets by the technique for cloud storage.

[14]At the outset, a renovated random generator was envisaged to generate a precise "noise". Subsequently, a bother perturbation algorithm was employed to supplement noise to the primary data. Thus the safety of the data was guaranteed, yet the mean and covariance of data of the administration supplier remained consistent. At this juncture, secret data was offered to regain the primary data from the muddled one. In the long run, they were able to combine the retrievable irritation with the right to gain entrance control technique to ensure that only the authorized clients were capable of retrieving the original data.

Security and privacy in cloud computing was efficiently brought to limelight. At that time, the outsourcing of data and business application to a stranger resulted in several safety and protection challenges which took the shape of a zooming concern. They invested their sweat and blood in investigating the safety and protection hassles in distributed computing, taking into consideration a trait driven system. They were able to sort out five most descriptive safety and protection qualities such as the secrecy, uprightness, accessibility, responsibility and security preservability. They took pains to explain the linkages between them, in addition to the susceptibilities which were probably ill-treated by the ambushers, the risk models as well as the peer safety techniques in a cloud scenario.

2. Proposed Privacy Preservation of Intermediate Datasets in Cloud

Cloud computing has been considered as the most promising and innovative technology of IT enterprise. The major issue in the growth of cloud computing area is security. It also have major risk in privacy areas, integrity and also mostly in users authentication. When handling large data within cloud computing environment, more number of data are generated in each and every step. For such kind of situation, opponents compare various intermediate data sets, to take valuable and sensitive information. In privacy preservation with a large number of intermediate data sets, encryption of sensitive data values can ensure privacy requirements but right selection of data and tuple information is more challenging.

Other constraint while dealing with cloud computing is the privacy and utility tradeoff. It is desirable to have good tradeoff between the two as to have more efficient and stable systems. In this paper, privacy preservation technique for intermediate datasets in cloud computing is designed and proposed keeping in mind to have good privacy and utility tradeoff. Utility is taken care with the employment of conditional entropy. Privacy is handled by the database difference ratio. The schematic diagram of the proposed technique Figure 1.

2.1 Conditional Entropy Method

The input for the proposed approach is an intermediate dataset in cloud. For highlighting the utility constraint,

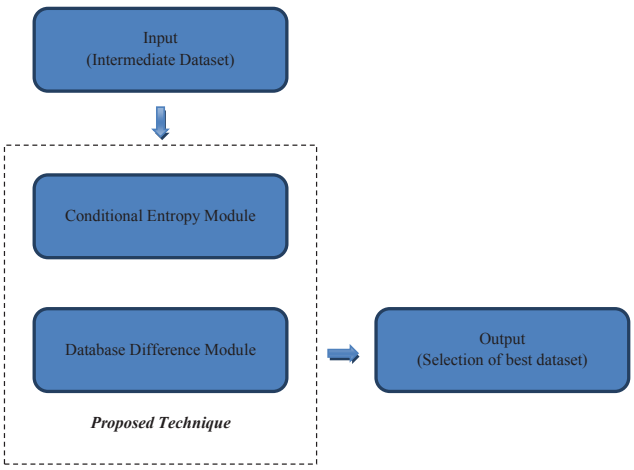

Figure 1. Schematic diagram of the proposed technique.

conditional entropy is employed with the help of PSO based optimization algorithm. In this module, the conditional entropy is found out between the first column and the original class and this is taken as the fitness function in PSO. The process is iterated after the convolution process. The process is explained in Figure 2.

The Particle Swarm Optimization algorithm is a population-based search algorithm. In PSO, member called a particle and the population called a swarm. Initially, random population is generated, each particle fly towards the searching space. Swarm converse best position and velocity to all. Then all particles shifted to the better position. At each and every step the position and velocity is updated and also compared with the fitness function.

2.1.1 PSO can be briefed in steps as

- Initially, random population is generated.
- Each particle, Calculate the fitness value pos and vel).
- Calculate p_{best} and g_{best} .
- Update the particle position.

$$vel_{t+1} = vel_t + \psi_1 (p_{best} - pos) + \psi_2 (g_{best} - pos) \quad (1)$$

$$pos_{t+1} = pos_t + vel_{t+1} \quad (2)$$

After finding the fitness and the PSO operation, the dataset is convoluted. The convolution is carried out column wise. The convolution is given by:

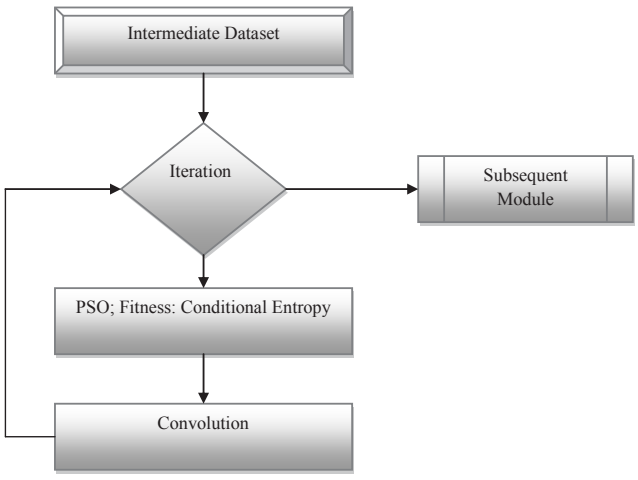

Figure 2. Block diagram of the conditional entropy module.

$$u$$

$$f * g(u) = \int f(x).g(u - x).dx \quad (3)$$

$$x=0$$

After the convolution process, the PSO based optimization making use of conditional entropy as fitness function is carried out.

2.2 Database Difference Method

Database difference ratio is found out in this module so as to improve the privacy tradeoff. Database difference is taken between the original database and convoluted database. Lower database difference would mean higher privacy. The schematic diagram of database difference module in Figure 3.

As from the figure, we see that after convolution step, the database difference ratio is found out between the original database and the convoluted database. Suppose the original database be represented by Z and the convoluted are represented by G , then the database difference ratio D is given by:

$$D = Z / G \quad (4)$$

Suppose the original database be expanded as $Z = \{z_1, z_2, z_3, ..., z_k\}$ and convoluted be expanded by $G = \{g_1, g_2, g_3, ..., g_k\}$, then database difference ratio D is given by:

$$D = (z_1 + z_2 + z_3, ... + z_k)/k \quad (5)$$

$$(g_1 + g_2 + g_3, ... + g_k)/k$$

Suppose after M convolutions, the database difference obtained be represented by $\{D_1, D_2, ..., D_M\}$. Then select the dataset having minimum database difference as the

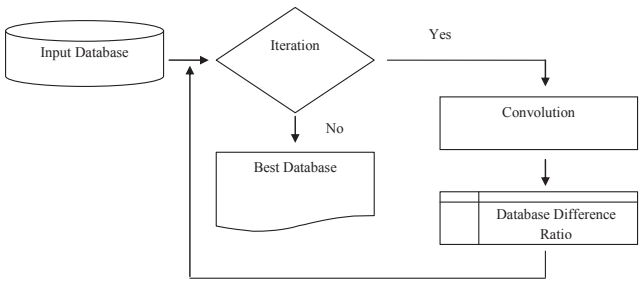

Figure 3. Schematic diagram of the database difference module.

best dataset represented by Z_B . This can be mathematically represented as:

$$Select\ Z_B\ from\ Z\ such\ that\ D_B \le D_i\ for\ 0 < i < M \qquad (6)$$

Hence, the best dataset is selected taking both the utility and privacy constraints.

3. Results

The proposed conditional entropy-based privacy leakage measure for privacy preservation of intermediate datasets in cloud is implemented in the JAVA program and the retrieval process is experimented with the Tamil and Telugu documents.

3.1 Performance Evaluation

The basic idea of our research is to privacy preservation of intermediate dataset in cloud. Here we utilize the four type of UCI machinery dataset. The performance of the proposed approach is carried out by varying the number of iteration. The obtained results are used to measure the entropy and Database Difference Ratio. In our work utility is taken care with the employment of entropy. Privacy is handled by the Database Difference Ratio.

It shows the performance of proposed approach in terms of entropy measures using mushroom dataset. The utility of the intermediate data is measures using the entropy values. When the iteration is 15 we obtain the entropy of 122567 for our proposed work which is 118311.0201 for existing approach. Here, we understand our proposed approach is achieving the maximum utility of intermediate data.

Figure 5 shows the graphical representation of proposed against existing approach in terms of entropy measure using Flags dataset. Here, the iteration value is five we obtain the maximum entropy which is high compared to existing approach.

Similarly, Figures 6 and 7also explain the performance of proposed approach using Census income (KDD) dataset and yeast dataset. Here also our proposed approach achieves the better result compared to existing approaches. Moreover, Database Different Ratio (DBDR) is improving the privacy tradeoff of intermediate dataset. In this work we clearly understand the lower database difference is increase the higher privacy.

Figure 8 shows the graphical representation of the proposed against the existing approach in terms of DBDR using mushroom dataset. Here, we obtain the minimum

- **Entropy based performance analysis**

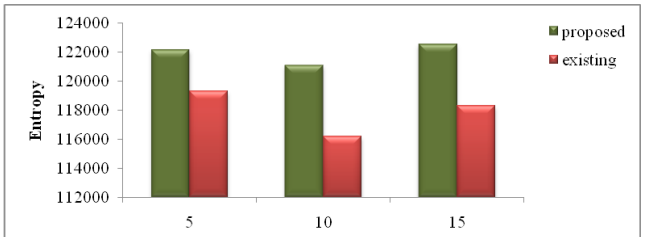

Figure 4. Comparison between the proposed technique and the existing approach in terms of entropy using mushroom dataset.

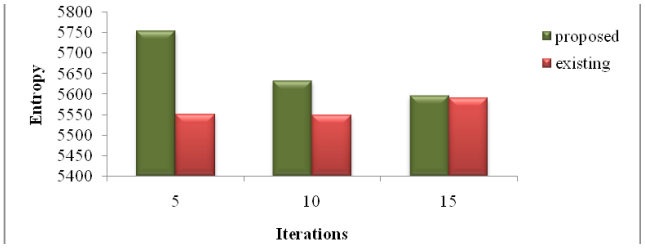

Figure 5. Comparison between the proposed technique and the existing approach in terms of entropy using flags dataset.

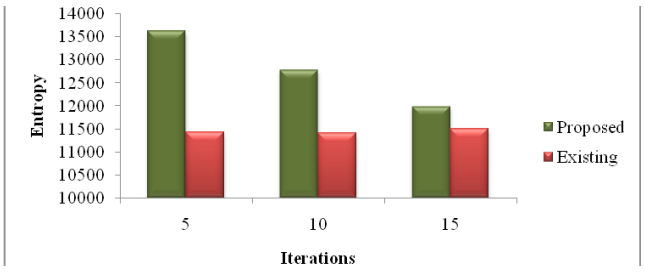

Figure 6. Comparison between the proposed technique and the existing approach in terms of entropy using Census Income (KDD) dataset.

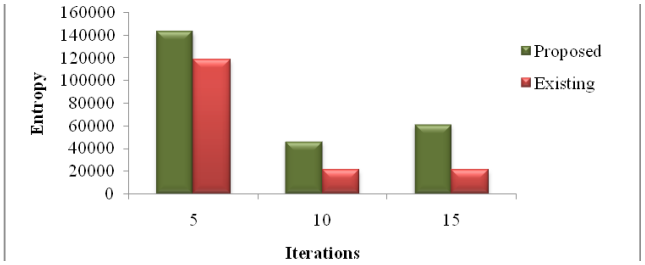

Figure 7. Comparison between the proposed technique and the existing approach in terms of entropy using yeast dataset.

DBDR of 0.52 for our proposed intermediate data privacy preserving which is 0.8 for existing approach.

Figure 9 shows the performance of proposed approach using Flags dataset. Here also we obtain the minimum DBDR value which is indicating our proposed approach having the high privacy.

Similarly, Figures 10 and 11 also explain the performance of proposed approach in terms of DBDR using Census income (KDD) dataset and yeast dataset. From the eight graph we clearly understand our proposed approach achieves the good performance compared to the existing approach.

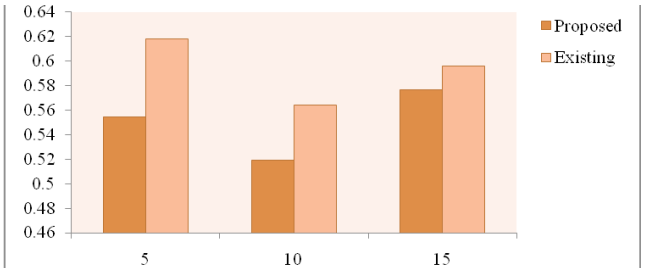

Figure 8. Comparison between the proposed technique and the existing approach in terms of DBDR using mushroom dataset.

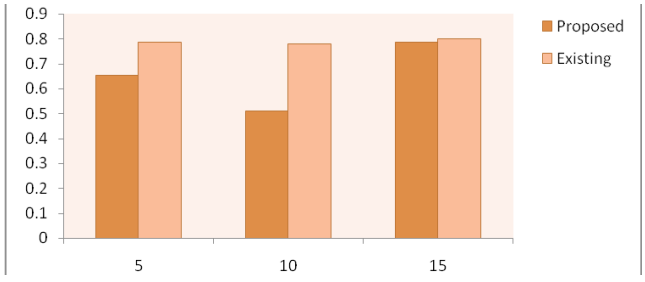

Figure 9. Comparison between the proposed technique and the existing approach in terms of DBDR using flags dataset.

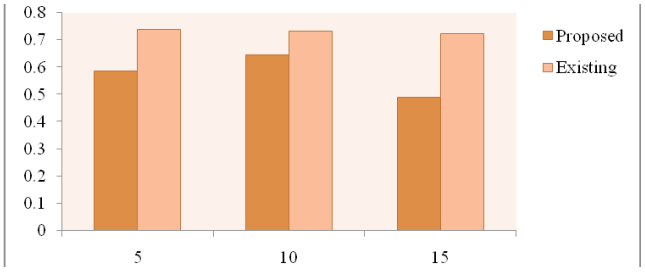

Figure 10. Comparison between the proposed technique and the existing approach in terms of DBDR using Census Income (KDD) dataset.

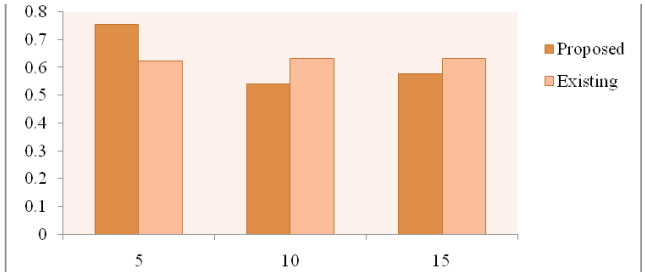

Figure 11. Comparison between the proposed technique and the existing approach in terms of DBDR using yeast dataset.

4. Conclusion

In this paper, a privacy preservation technique for intermediate datasets in cloud computing is designed. Entropy and database difference ratio is taken as the evaluation matrices and is tested using various datasets of mushroom, yeast dataset, flags dataset and Census Income (KDD) dataset. The technique is also compared to the existing technique. The proposed technique achieved better results by having higher entropy values obtained 143621 and lower dataset difference ratio obtained 0.4876.

5. References

1. Han S, Xing J. Ensuring data storage security through a novel third party auditor scheme in cloud computing. Proceeding of IEEE International Conference on Cloud Computing and Intelligence Systems; 2011. p. 264–8.
2. Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson G, Rabkin A, Stoica I, Zaharia M. A view of cloud computing. Communications of the ACM. 2010; 53(4):50–8.
3. Shaikh FC, Haider S. Security threats in cloud computing. Proceeding of IEEE International Conference on Internet Technology and Secured Transactions; 2011. p. 214–9.
4. Shen Z, Li L, Yan F, Wu X. Cloud computing system based on trusted computing platform. Proceeding of IEEE International Conference on Intelligent Computation Technology and Automation. 2010 May; 1:942–5.
5. Zhang X, Lai SQ, Liu NW. Research on cloud computing data security model based on multi-dimension. Proceeding of IEEE International Symposium on Information Technology in Medicine and Education. 2012 Aug; 2:897–900.
6. Kantarcioglu M, Bensoussan A, Hoe SRC. Impact of security risks on cloud computing adoption. Proceeding of IEEE International Conference on Communication, Control and Computing; 2011Sep. p. 670–4.

7. Almorsy M, Grundy J, Ibrahim AS. Collaboration-based cloud computing security management framework. Proceeding of IEEE International Conference on Cloud Computing; 2011 Jul. p. 364–71.
8. Chhabra RK, Sharma S, Verma A, Lala A. Dynamic password authentication: A novel approach for user authentication in cloud computing. Proceeding of IEEE International Conference on Cloud Computing; 2012 Sep. p. 1–5.
9. Tripathi A, Mishra A. Cloud computing security considerations. Proceeding of IEEE International Conference on Signal processing, Communications and Computing; 2011 Sep. p. 1–5.
10. Behl A, Behl K. Security paradigms for cloud computing. Proceeding of IEEE International Conference on Computational Intelligence, Communication Systems and Networks; 2012 Jul. p. 200–5.
11. Khan AR, Othman M, Madani SA, Khan SU. A survey of mobile cloud computing application models. IEEE Communications Surveys and Tutorials. 2014 Feb; 1(1):393–413.
12. Wang C, Chow SSM, Wang Q, Ren K, Lou W. Privacy preserving public auditing for secure cloud storage. IEEE Transaction on Cloud Computing. 2013 Feb; 62(2):362–75.
13. Yang P, Gui X, An J, Yao J, Lin J, Tian F. A retrievable data perturbation method used in privacy-preserving in cloud computing. IEEE on communication. 2014; 11(8):73–84.