International Journal of Advanced Trends in Engineering and Technology (IJATET)
Impact Factor: 5.665, ISSN (Online): 2456 - 4664
(www.dvpublication.com) Volume 2, Issue 2, 2017

BALANCED AWARE FIREFLY OPTIMIZATION BASED COST-EFFECTIVE PRIVACY PRESERVING APPROACH OF INTERMEDIATE DATA SETS OVER CLOUD COMPUTING

J. Sasidevi*, Dr. R. Sugumar & P. Shanmuaga Priya***

* Research Scholar, Department of Computer Science and Engineering, St. Peter's University, Chennai, Tamilnadu

** Associate Professor, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai, Tamilnadu

Cite This Article: J. Sasidevi, Dr. R. Sugumar & P. Shanmuaga Priya, "Balanced Aware Firefly Optimization Based Cost-Effective Privacy Preserving Approach of Intermediate Data Sets Over Cloud Computing", International Journal of Advanced Trends in Engineering and Technology, Volume 2, Issue 2, Page Number 215-223, 2017.

Abstract:

Cloud computing is an embryonic archetype with remarkable impetus; however its exclusive facets intensify safety and privacy confronts. In the previous method, the privacy of intermediate data set problems is dealt with which is concentrated to regain privacy sensitive information. Alternatively the previous system contains problem with time and cost intricacy. As well it contains issue with dealing privacy conscious well-organized scheduling of intermediate data sets in cloud by considering privacy preserving. In order to surmount the above stated problems, in the existing system, enhanced balanced scheduling methodology is presented to get better the cost complexity and privacy preservation. Balanced aware FireFly Optimization (BFFO) is used for proficient privacy conscious data set scheduling. This technique is utilized to discover the resolution that carries out best on poise amongst a set of resolutions with similar execution time. Consequently the research system gives superior privacy preservation and enhanced scheduling cost more willingly than the previous method. The encryption technique is used to guarantee the security and end users decrypted the real information with improved privacy. The experimentation outcome show that the presented method confirms superior privacy, lesser cost, lesser time complexity and proficient storage metrics utilizing BFFO methodology compared to the previous Cost based Heuristic (C_HEU) algorithm.

Key Words: Privacy Preserving, Intermediate Data Set, Firefly Algorithm & Cloud Computing

1. Introduction:

Cloud computing is a type for enabling pertinent, on-demand network access to a general pool of configurable calculating resources which could be quickly stipulated and discharged with minimum running attempt or service provider communication [1]. In order to conserve privacy it is a complicated crisis in developing protected cloud storage system. In cloud utilizes the additional amount of intermediate data sets produced; however till now protecting the privacy of intermediate datasets turns out to be a tricky accomplishment issue since adversaries might get better privacy secret information by examining several intermediate datasets. Encrypting each and every datasets in cloud is typically accepted in previous techniques related to this confront.

Encrypting each and every intermediate datasets are either regimented or profitable since it will consume additional time and expensive for data demanding applications to encrypt and decrypt datasets frequently whilst carrying out any operation on the data set [2]. Privacy is a significant problem for cloud computing, in relation to lawful acquiescence and user belief and requires to be treated at each stage of design. The methodology gives the significance of guarding person's privacy in cloud computing and gives certain privacy preserving methods utilized in cloud computing services [3] [4]. It says to that it is extremely significant to consider privacy whilst developing cloud services, in case these engage the collection, processing or sharing of individual data. According to this study, major subject considered is of preserving privacy of data. This system explains privacy of data however doesn't let indexed search and doesn't veil user's identity.

In [5] a methodical design, Privacy-MaxEnt, to include environment information in privacy quantification. This technique is reliant on the utmost entropy principle. This research care for each and every conditional probabilities P (SA | QI) as indefinite variables; it care for the background acquaintance like the restraints of these variables; additionally, it as well devise restrictions from the published data. The objective turns out to be discovering a resolution to those variables (the probabilities) which fulfill all these restraints. Even though numerous solutions might subsist the main unbiased estimate of P (SA | QI) is the one which attains the utmost entropy.

With no failure of generality, the idea of intermediate data set in this denotes intermediate and resulting data sets [6] [7]. On the other hand, the storage of intermediate data expands attack surfaces with the intention that privacy requisites of data holders are at possibility of being desecrated. Typically, data sets in cloud are processed by many parties, nevertheless barely ever managed by real data set holders. This enables an opponent to gather intermediate data sets together and menace privacy-sensitive information from them, taking significant financial loss or ruthless social reputation mutilation to data owners. However, small consideration is remunerated to such a cloud-specific privacy problem.

2. Related Work:

In [8] Lin and Tzeng (2010) stated this issue utilizing erasure codes. Proxy re-encryption methods are utilized. These methods assist public key encryption, for sustaining data forwarding steadily. According to the methods users could give re-encryption keys to server and afterward facilitates server to share data. The storage server re-encrypts the data that is encrypted previously. In [9] Tang et al (2008) proposed a time based proxy re-encryption method. This key contains more efficacies while matched up with other keys associated to security.

In [10] Ateniese et al (2008) for integrity verification numerous methods came into subsistence wherever focal point was on data store, retrieval and forwarding in safe manner. Later on, the idea of provable data possession (PDP) methods is opened up. Messages are utilized in plain text in these methods. On the other hand, it is concentrated on the cloud storage security concerning data storage retrieval and forwarding. This research progresses their method by developing a timestamp based method which guarantees ideal collaboration amongst the servers. Therefore the issue of communication discrepancies amongst the servers is mentioned.

In [11] Benjamin et al (2007) thought about the issue of guarantying a person's anonymity whilst issuing person specific data for classification analysis. It indicated that the earlier best possible k-anonymization dependent upon a closed form cost metric doesn't specify the classification requisite. Our technique is dependent upon two notifications particular to classification: Information particular to persons are likely to be over fitting, therefore of small efficacy, to classification; although a masking operation removes certain helpful classification formations, another formations in the data come into sight to help. Consequently, not each and every data items are uniformly helpful for classification and less helpful data items give the room for anonymizing the data devoid of negotiating the utility. With these notifications, we proposed a top-down technique to iteratively distill the data from a common state into a special state, directed by increasing the exchange amid information and anonymity. This top-down technique provides a natural and competent formation for dealing with categorical and continuous attributes and numerous anonymity requisites. Experimentations proved that our technique successfully conserves information utility and person's privacy and scales well for large data sets.

In [12] Wang et al (2008) illustrate a common methodology to carry out inference analysis crosswise numerous published datasets in PPDR. We devise the issue as a well studied utmost entropy estimation issue, and utilize standard non-linear programming tool to resolve it. Experimentation outcomes express the efficacy of this technique. The methods we suggest for privacy quantification is common. We could also build up practical PPDR techniques to make sure that privacy requisites are fulfilled at every releasing point. In addition, we could incorporate background knowledge (the previous information which opponents may be familiar with the original data) in privacy analysis of PPDR.

In [13] Schwarzbach et al (2016) propose a technique for privacy preserving cloud based collaborative business process management incorporating structural design and a privacy modeling technique. We assess the structural design and the proposed privacy modeling technique by a use-case from the logistics area and prove their achievability. These activities and policies are associated with the existing services of the service repository. In case there are no services accessible matching up the users privacy policies, the user is noticed and could alter his privacy policy consequently or has to guarantee that suitable services are incorporated into the platform. Subsequent to modeling the business process is being qualified by the certification module.

In [14] Arora et al (2013) examination is carried out on Firefly algorithm regarding its active behavior and convergence as significant facets. Algorithm is examined utilizing tools from discrete time dynamic system and this investigation offers qualitative instructions for common (random) algorithm parameter selection. Simulation experimentations are carried out with three parameter sets, three parameter set of fireflies in field and five benchmark functions. The rapidity of convergence robustness tradeoff was conversed. Superior outcomes are attained in little amount of fireflies in field and lesser value of α and γ. Additional investigation is required to elucidate consequence of randomness and their consequence on convergence. Enhanced parameter sets almost certainly wait for detection in the outlined algorithm convergence domain.

3. Proposed Methodology:

In the proposed method, the BFFO technique is used to progress the privacy cost deduction with competent task scheduling. The on the whole presented system is demonstrated in the fig 1.

3.1 Problem Analysis:

Data provenance is applied to direct intermediate data sets in our study. Provenance is normally described as the source, basis or history of origin of certain objects and data that could be considered as the information upon how data were produced [15]. Reproducibility of data provenance could aid to rejuvenate a data set from its adjacent accessible predecessor data sets more willingly than from scrape. It presumes in this that the information documented in data provenance is influenced to develop the generation associations of data sets.

It describes numerous fundamental notations below. Consider to be a privacy sensitive original data set. It utilizes D <d1; d2; . . . ; dng to signify a collection of intermediate data sets produced from do where n is the number of intermediate data sets. Remind that the idea of intermediate data in this denotes intermediate and resulting data.

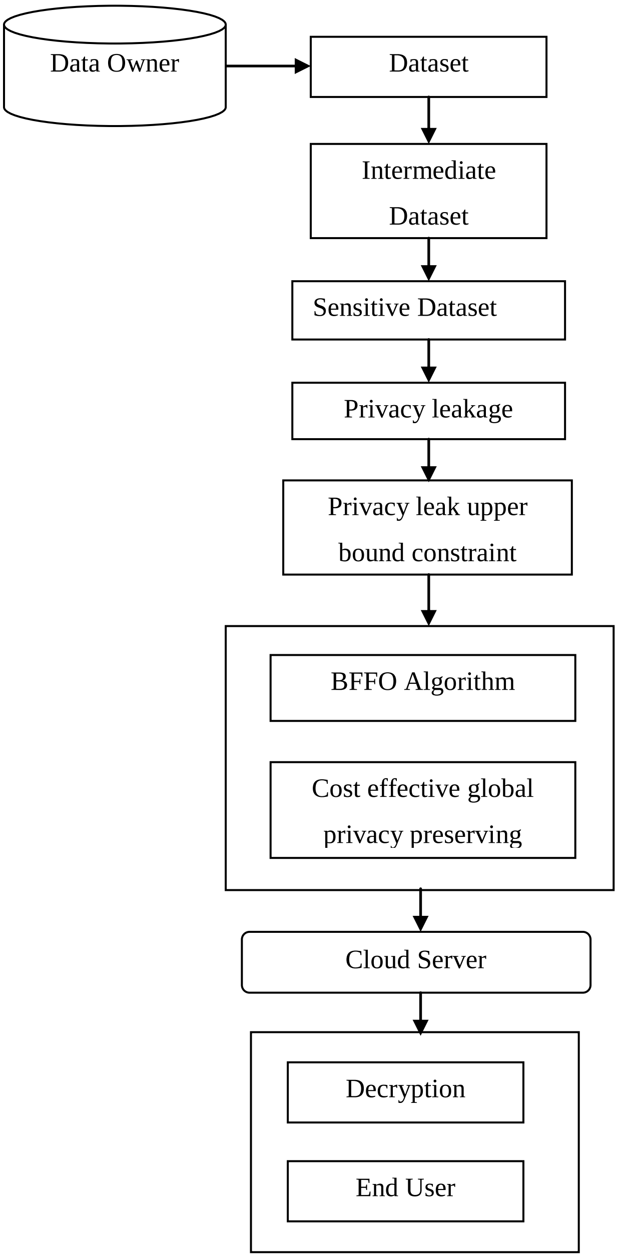

Figure 1: Overall block diagram of the proposed system

Directed Acyclic Graph (DAG) is utilized to confine the topological formation of generation associations amongst these data sets. A DAG denoting the generation associations of intermediate data sets D from do is described as a Sensitive Intermediate data set Graph, indicated as SIG.

Privacy-preserving cost of intermediate data sets branches from recurrent en/decryption with charged cloud services. Cloud service venders set up a variety of pricing models to help the pay-as-you-go model, for instance Amazon Web Services pricing model. Sensibly, encryption or decryption requires calculation power, data storage, and other cloud services. In order to evade pricing information and concentrate on the conversation of our nucleus notions, it merges the costs of a variety of services needed by encryption or decryption into one. This joint price is represented as PR. PR point to the overhead of encryption or decryption on per GB data per execution.

An attribute vector is used to structure a number of significant properties of the data set di. The vector is represented as hSi; Flagi; fi; PLii. The word Si denotes the size of di. The word Flagi, a dichotomy label, denotes whether di is unseen. The word fi denotes the frequency of accessing or processing di. If di is tagged as secreted, it would be en/decrypted all time while accessed or processed. Therefore, the larger fi is, the more cost would be acquired if di is unseen. Typically, fi is anticipated from the data provenance. The word PLi is the privacy leakage via di, as well as is calculated.

3.2 Privacy Preservation for Intermediate Data Set:

The worth of the joint privacy leakage deserved by several data sets in $D = \{d_1, d_2, \ldots, d_n\}$, $n \in N$ is described by

$$PL_m(D) \triangleq H(S,Q) - H_D(S,Q) \qquad (1)$$

$H(S,Q)$ as well as $H_D(S,Q)$ are the entropy of $< S,Q)$ previous to and subsequent to data sets in D are noticed, correspondingly. $H(S,Q) = \log(|QI|) . |SD|). H_D(S,Q)$ Could be computed once $P(S,Q)$ is anticipated subsequent to data sets in D are noticed. Specified the association amid ε and $PL_m(D^{une})$ in PLC, ε ranges in the interval $[max_{1 \leq i \leq n}\{PL_s(d_i)$, log(|QI|.|SA|)]

It is concentrated on to draw from an upper bound of $PL_m(D^{une})$ which could be effortlessly calculated. Instinctively, in case an upper bound $BPL_m(D^{une})$ is identified, a healthier privacy leakage limit $B(PL_m(D^{une})) \leq \varepsilon$ could be a satisfactory condition of the PLC. Consequently, $PL_m(D^{une})$ would never go beyond the threshold ε in case $B(PL_m(D^{une})) \leq \varepsilon$ holds.

Consider d_u as well as d_v be two data sets whose privacy leakage are $PL_s(d_u)$ and $PL_s(d_v)$ correspondingly. The combined privacy leakage rooted by them mutually is$PL_m(d_u, d_v)$. This property of combined privacy leakage could be expanded to numerous unencrypted data sets in D^{une}

$$PL_m(D^{une}) \leq \sum_{d_i \in D^{une}} PL_s(d_i)$$

An upper bound restraint based technique to choose the essential subset of intermediary data sets which requires to be encrypted for reducing privacy-preserving cost. In order to convince the PLC, we de-compose the PLC recursively into diverse layers in an SIT. Subsequently, the issue mentioned in (3) could be noticed through dealing with a sequence of small-scale optimization issues. Consider the privacy leakage threshold needed in the layer L_i be ε_i, $1 \leq i \leq H$. The privacy leakage deserved by UD_i in the resolution π_i could by no means be superior than ε_i, i.e., $PL_m(UD_i) \leq \varepsilon_i$. The threshold ε_i could be considered as the privacy leakage threshold of the remainder part of an SIT subsequent to the layer L_{i-1}. In relation to the fundamental thought of method, the privacy leakage restraint $PL_m(UD_i) \leq \varepsilon_i$ is replaced by one among its satisfactory conditions.

In relation to (7), the PLC could be surrogated by a group of privacy leakage restraints, called as PLC1:

$$\sum_{d \in UD_i} PL_s(d) \leq \varepsilon_i, \quad 1 \leq i \leq H \qquad (2)$$

The above threshold ε_i, $1 \leq i \leq H$ is computed by

$$\begin{cases} \varepsilon_i = \varepsilon_{i-1} - \sum_{d \in UD_i} PL_s(d) \\ \qquad \varepsilon_1 = \varepsilon \end{cases} \qquad (3)$$

A local encryption solution in the layer Li is sufficient in case it fulfills thePLC_1. The group of reasonable solutions in L_i is represented as$\Lambda_i^f \triangleq \{\pi_{ij} | \pi_{ij} \epsilon \Lambda_i$, here j is the amount of reasonable solutions. Correspondingly, a reasonable global encryption solution could be represented as$\pi_f^k \triangleq < \pi_{1_{j1}}, \ldots \ldots \pi_{H_{jH}} >$, here

$$\pi_{ij} \in \Lambda_i^f, 1 \leq i \leq H, 1 \leq k \leq \prod_{i=1}^{H} |\Lambda_i|$$

Specified a possible global solution π_f^k for an SIT, condense the SIT into a "compressed" tree layer by layer from L_1 to L_H represented as $CT(\pi_f^k)$,here H is the height of the DG in SIT. The structure of $CT(\pi_f^k)$ is attained by means of three steps.

Initially, the data sets in EDi are "compressed" into one encrypted node. In relation to the EDT property, these compressed nodes in conjunction with the real data set come into sight to be a string comprise the length being H. Next, every offspring data sets of the data sets in UDi are skipped. This would not have an effect on the privacy preserving in relation to the RPC property. Thirdly, the data sets in UDi are condensed into one node.

Frequently, above one reasonable global encryption solution subsists underneath the PLC1 restraints, since there are numerous another local solutions in every layer. Additionally, every intermediary data set contains different size and frequency of usage, bringing about diverse on the whole cost with diverse solutions. Consequently, it is preferred to discover a reasonable solution with the least privacy-preserving cost underneath privacy leakage restraints. Observe that the least solution stated in this is fairly pseudo minimum since an upper bound of combined privacy leakage is simply a rough calculation of its accurate value. However a solution could be accurately smallest in the logic of the PLC1 restraints. It originates the recursive minimal cost formula in this manner.

The minimum cost for privacy preserving of the data sets subsequent to L_{i-1} underneath the privacy leakage threshold ε_i is denoted as$CM_i(\varepsilon_i)$, $1 \leq i \leq H$. Specified a reasonable local encryption solution $\pi_i =< ED_i, UD_i >$ in L_i the rate deserved by the encrypted data sets in Li is represented as $C_i(\pi_i)$

$$C_i(\pi_i) \triangleq \sum_{d_k \in ED_i} S_K.PR.f_k, \quad 1 \leq i \leq H \qquad (4)$$

Then $CM_i(\varepsilon_i)$ is compute by the recursive formula

$$\begin{cases} CM_i = \min_{\pi_{ij} \in \Lambda_f^k} (\sum_{d_k \in ED_i} S_k . PR . f_k) \\ \quad + CM_{i+1}(\varepsilon_i - \sum_{d_k \in ED_i} PL_s(d_k)) \} \\ \quad CM_{H+1}(\varepsilon_{H+1}) = 0 \end{cases} \qquad (5)$$

3.3 Balanced aware Fire Fly Optimization (BFFO):

Feature Selection Using MFA:

The idea of enhanced algorithm is to develop the best possible cost reduction process. The firefly technique is generated leisurely process for high dimensional dataset and not proficient in identifying the universal optimal solution. For this reason, Improved Firefly Algorithm (IFA) is improved utilizing Gaussian Firefly Algorithm (GFA) to amplify the best possible solutions. [16]. This Gaussian algorithm uses three behaviours to get better performance of FA.

- ✓ The primary behaviour is an adaptive step length which modifies random step length by the time
- ✓ The second is individual behaviour or directed movement which manages random movement to toward universal best
- ✓ The final behaviour is a social behaviour which modify the location of every firefly dependent upon a Gaussian Distribution (GD)

Adaptive Step Length: According to standard Firefly Algorithm (FA), firefly movement step length is a preset value. Each firefly is in motion with a preset length in every iterations. By reason of the preset step length, the method would ignore enhanced local search abilities and sometimes it ensnares into numerous local optimums. It is enhanced that FA searches the space internationally in first iterations and in the final iterations it utilize the specific place to take out enhanced solutions

Directed Movement: In standard Firefly Algorithm (FA), firefly movement is dependent upon light intensity and contrasting it amid each pair of fireflies. Consequently for any two fireflies, the fewer bright one would go in the direction of the brighter one. In case no one is brighter compared to a specific firefly, it would move arbitrarily. In the proposed method this random movement is focussed, and that firefly goes in the direction of the greatest solution with improved cost in that iteration. The firefly movement is fascinated to the finest solution which is very gorgeous (brighter). This leads to that, if there was no local finest in every firefly's adjacent, it goes in the direction of the finest solution and creates enhanced position for every firefly for the subsequent iteration and they would get more close to the universal best.

Social Behaviour: Random walk is a random process that comprises a sequence of successive random steps. At this point the step size or length in a random walk could be permanent or changing. The step length conform the Gaussian distribution in the random walk. By reason of these benefits, this work utilizes GFA for clustering Anonymized data.

Feature Selection Using Gaussian Firefly Algorithm (GFA):

The most important three features to perform the classification process for features are specified below:

- ✓ The whole fireflies are unisex. Accordingly one attribute data matrix (firefly) would be drawn to other attributes data matrix (fireflies) not taking their sex [20].
- ✓ Pleasant appearance is comparative to their vividness. Consequently, for any two flashing attributes data matrix (fireflies), a lesser number of bright one would go towards the brighter one. In case it has no brighter one compared to a specific firefly, it would travel randomly
- ✓ The vividness of an attribute in the data matrix (firefly) inclined by the landscape of the resemblance and variation is regarded as objective function. For a maximization setback, the brightness could essentially be comparative to the value of the objective function

The steps in Gaussian Firefly Algorithm (GFA) are defined in table 1. The parameter values of the FA are specified in Table 1.

Table 1: Parameters of Firefly Algorithm (FA)

Parameters	Values	Description
α	0.2	Alpha
β_0	0.3	Beta$_0$
γ	0.2	Gamma
Iterations	20	Generations

Random walk is a random process that comprises taking a sequence of successive random steps. Currently, the step size or length in a random walk could be preset. While the step length goes after the Gaussian Distribution (GD), the random walk turns out to be the Brownian motion. With a view to moving every attribute data matrix values, it utilizes random walk notions to move every agents based upon a GD which is defined in the specified equation (3):

$$p = f(dm|\mu, \delta) = \frac{1}{\delta\sqrt{2\pi}} e^{-(dm-\mu)^2/2\delta^2} \qquad (6)$$

Here x denotes an error amid most admirable resolution and fitness value of numerous data matrix (firefly) i is denoted in equation (4).

$$x = f(g_{best}) - f(dm_i) \qquad (7)$$

In equation (3), μ denotes Mean and δ represents Standard Deviation (SD). On account of utilizing standard Normal Distribution (ND), it is set to $\mu = 0$ and $\delta = 1$. After that a random number would be drawn from this GD which is connected to every firefly (data matrix) probability (p). Social behavior of fireflies is presented by,

$$dm_i = dm_i + \alpha * (1 - p) * rand() \qquad (8)$$

Here α in equation (5) denotes firefly parameter that is fine tuned by adaptive parameter technique [17]. On the other hand, in case the novel place of firefly i root to improved fitness for that exclusive firefly, it would move to that novel place. In this presented GFA, data clustering methodology based upon the random walk, every firefly from single data holder matrix i movement is drawn to an optimum solution which is very striking (brighter) based upon the location for every firefly for subsequent iteration and they discover very near to universal best as specified in equation (4). The fitness value is regarded as accuracy here as well as the value is exposed the best possible finding of intrusion characteristics. The best possible characteristics are utilized to find out the accuracy amid all characteristics. The intrusion detection accuracy is improved by raising the brightness values (6). This accuracy value is utilized to recognize the intrusion class for the specified KDD cup dataset successfully.

$$Accuracy = \frac{(True\ positive\ + true\ negaive\ )}{TP + TN + FP + FN} \qquad (9)$$

In equation (6), denotes the characteristics as standard or intrusion attacks. TP calculates the optimistic outcomes in opposition to both standard and intrusion characteristics. TN is described as negative outcome devoid of standard and intrusion class of every characteristic. FN is negative outcomes in opposition to both characteristics and FP is positive outcomes devoid of both normal and intrusion characteristics.

```
✓ Initialize algorithm parameters: Objective function of f(x), from, here $x = (x_1, \ldots\ldots, x_d)^T$
✓ Produce initial population of fireflies or $x_i$ $(i = 1, 2, \ldots, n)$
✓ Typify light intensity of $I_i$ at $x_i$ through $f(x_i)$ from equation(p)
✓ Whilst (t < MaxGen)
    • For i = 1 to n (all n fireflies);
    • For j = 1 to n (all n fireflies)
    • If $(I_j > I_i)$ move firefly i in the direction of j; end if
    • Prettiness alters with distance 'r' through Exp [− r2]
    • Approximate novel solutions and adjust light intensity;
    • End for j;
    • End for i;
✓ Grade the fireflies and recognize the present best cluster
✓ Illustrate normal distribution
✓ For k = 1, ... n all n fireflies
✓ Acquire a random data records from Gaussian distribution implement (o) for selected data matrix value
✓ Guesstimate novel solution (new solution(k))
✓ If new (new solution(k < solution(i) ) && (new solution(k) < last solution(k))
✓ Shift features in the direction of present best
✓ End if
✓ End for k;
✓ End while;
✓ Post process outcomes and visualization
✓ End procedure;
```

The research technique is utilized to choose the data sets with less cost in conjunction with superior privacy to encrypt. Dependent upon this, $g(SN_i)$ *is defined as* $g(SN_i) \triangleq C_{cur} / (\varepsilon - \varepsilon_{i+1})$, *where* C_{cur} is the privacy preserving cost which is deserved up to now, ε is the preliminary privacy leakage threshold, and ε_{i+1} is the privacy leakage threshold for the layers subsequent to L_i. Particularly, C_{cur} is computed by $C_{cur} = \sum_{d_j \in U_{k=1}^{i} ED_k} (S_j.PR.f_j)$. The lesser Ccur is, the lesser total privacy-preserving cost would be. The privacy-preserving solution and consequent cost are originated from the goal state.

In order to evade the size of the priority queue enlarge noticeably, the method just preserves the state nodes with top K uppermost heuristic values. While identifying to append child search nodes in layerLiþ1 into the priority queue, the technique produces a local encryption solution from CDEi at first. The method most likely undergoes poor competence since it has to confirm each and every combination of data sets in CDEi.

3.4 Privacy Using Encryption Method:

It is significant to guard both the data being dispersed and the identities of users let to access data. In the previous method, encrypted file systems are unsuccessful to guard the privacy of users. User privacy is negotiated since the fundamental encryption methodologies reveal the identities of a ciphertext's receivers.

Numerous such systems just give away the identities of the users in the type of labels appended to the ciphertext. In addition, those systems which try to evade disclosing the receiver's identity are susceptible to having their user's privacy negotiated by a novel selected-ciphertext attack which presented.

This research method, private broadcast encryption, facilitates the competent encryption of a message to numerous receivers devoid of exposing the identities of the receivers of the message, to other receivers. It fulfills a sturdy description of receiver privacy in the facade of vigorous attacks. It as well attains decryption in a constant number of cryptographic operations, doing comparably to present systems which do not present user privacy.

4. Experimental Result:

U-Cloud is a cloud computing environment at the University of Technology Sydney (UTS). The system general idea of U-Cloud is represented in Fig. 4. The calculating services of this system are placed amongst numerous labs at UTS. Resting on hardware and Linux operating system, we set up KVM virtualization software that virtualizes the infrastructure and gives united computing and storage resources. In order to create virtualized data centres, set up OpenStack open-source cloud environment [18] for global management, resource scheduling and interaction with users. Additionally, Hadoop [19] is set up dependent upon the cloud constructed through Open Stack to assist enormous data processing. The experimentations are carried out in this cloud environment.

Original dataset

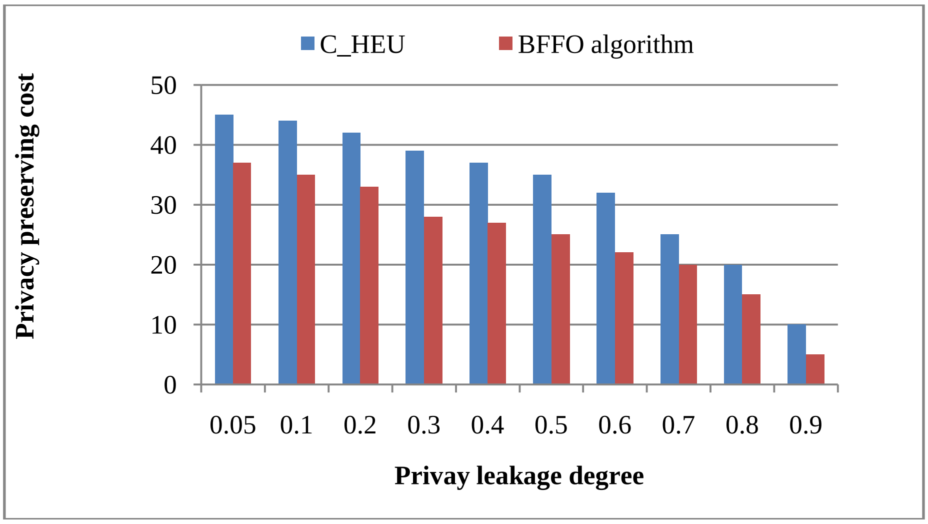

Figure 2: Privacy preserving cost vs privacy leakage degree

ε_d =0.01

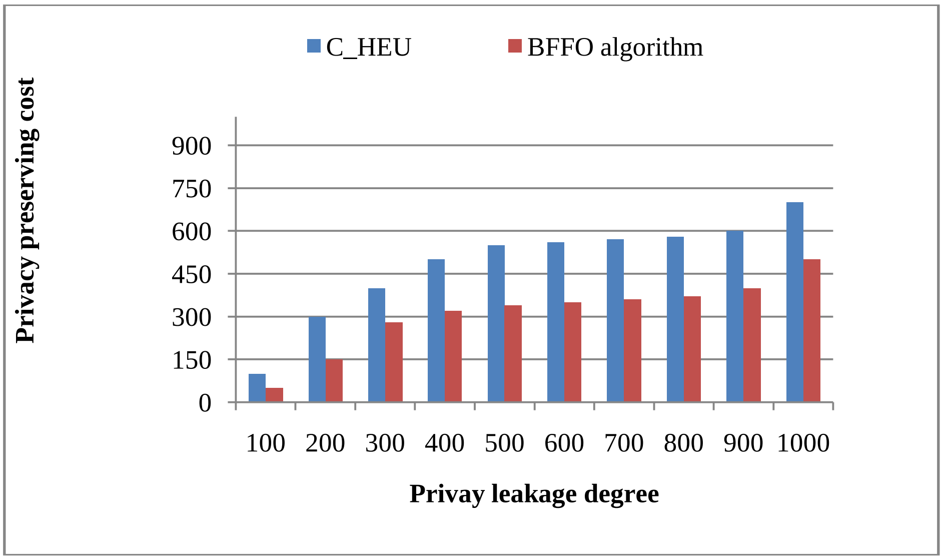

Figure 3: Privacy preserving cost vs privacy leakage degree

ε_d =0.05

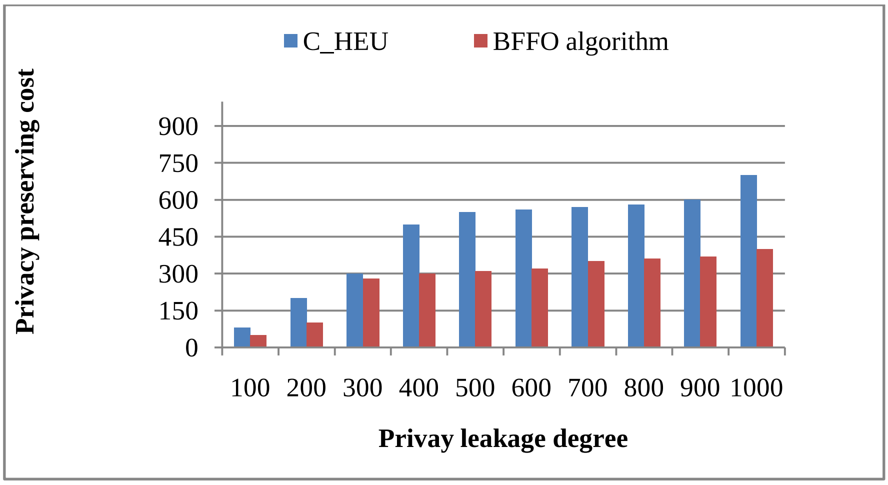

Figure 4: Privacy preserving cost vs privacy leakage degree

ε_d =0.1

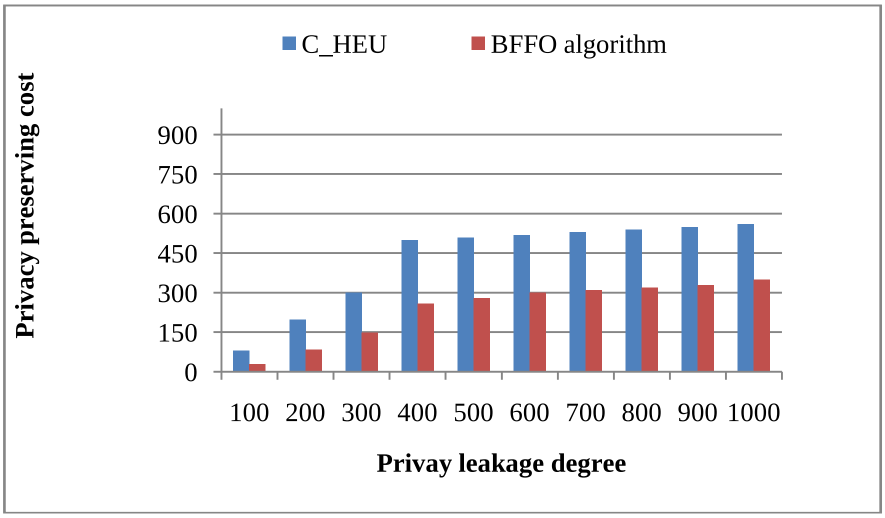

Figure 5: Privacy preserving cost vs privacy leakage degree

ε_d =0.2

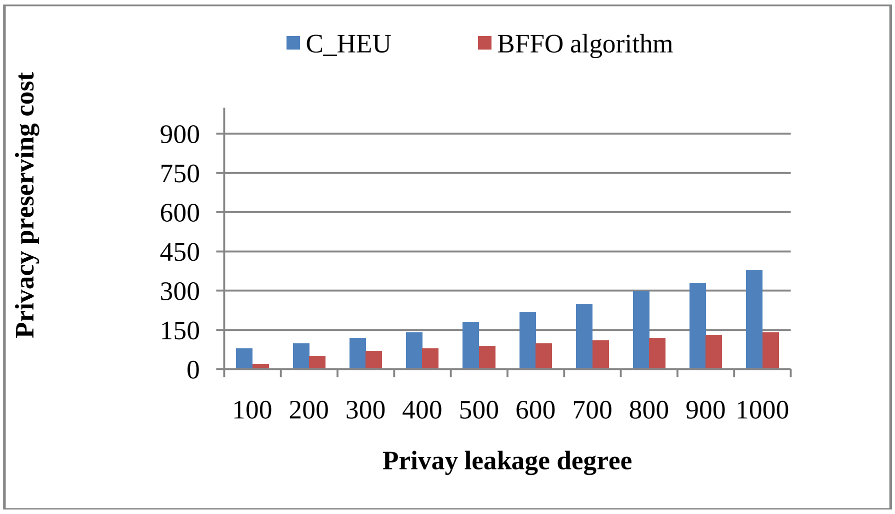

Figure 6: Privacy preserving cost vs privacy leakage degree

From the fig 2, the presented BFFO technique proves the lesser privacy preserving cost compared to the diverse privacy leakage degree. According to the fig 3, diverse privacy leakage degree and privacy cost metric is calculated for $\varepsilon_d = 0.01$. while the amount of intermediary data sets is being larger, more data sets are needed to be encrypted. The privacy price is decreased considerably utilizing the BFFO method. The fig 4, 5 and 6 are plotted$\varepsilon_d = 0.05$, $\varepsilon_d = 0.1$ as well as $\varepsilon_d = 0.2$ correspondingly that proves the lesser privacy cost for diverse privacy leakage degree significantly. Therefore the outcome states that the research system is superior in terms of less cost and improved privacy.

5. Conclusion:

Cloud technology assists the legitimate cloud users to entrée bounty of resources which are moved and gathered in cloud. In order to conserve the privacy on intermediary dataset the privacy method with optimization technique is presented in this work. The research concentrated on the privacy aware competent scheduling of intermediary data sets in cloud computing environment. In this study, BFFO method is presented to amplify the privacy preservation and task optimization process very efficiently. The presented BFFO methodology is mostly concentrated to get better the privacy cost deduction and protection on the specified intermediary datasets. It discovers the finest solutions with less computational complexity in cloud. The outcome shows that the presented BFFO has greater performance more willingly than the previous C_HEU system in terms of less privacy cost and improved privacy.

6. References:

1. R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, “Cloud Computing and Emerging It Platforms: Vision, Hype, and Reality for Delivering Computing as the Fifth Utility, “Future Generation Computer Systems, vol. 25, no. 6, pp. 599-616, 2009
2. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data,” Proc. IEEE INFOCOM ’11, pp. 829-837, 2011.
3. Wang, Jian, et al. "Providing privacy preserving in cloud computing." Human System Interactions (HSI), 2010 3rd Conference on. IEEE, 2010.
4. K. Zhang, X. Zhou, Y. Chen, X. Wang, Y. Ruan, “Sedic: Privacy Aware Data intensive Computing on Hybrid Clouds,” Proc. 18th ACM Conf. Computer and Comm. Security (CCS”11), pp 515-526, 2011.
5. Du, Wenliang, Zhouxuan Teng, and Zutao Zhu. "Privacy-maxent: integrating background knowledge in privacy quantification." Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008.
6. D. Yuan, Y. Yang, X. Liu, and J. Chen, “On-Demand Minimum Cost Benchmarking for Intermediate Data Set Storage in Scientific Cloud Workflow Systems,” J. Parallel Distributed Computing, vol. 71, no. 2, pp. 316- 332, 2011.
7. Yuan, Dong, et al. "A data dependency based strategy for intermediate data storage in scientific cloud workflow systems." Concurrency and Computation: Practice and Experience 24.9 (2012): 956-976.
8. Lin and Tzeng, “A Secure Decentralized Erasure Code for Distributed Network Storage,” IEEE Trans. Parallel and Distributed Systems, vol. 21, no. 11, pp. 1586-1594, Nov. 2010.
9. Tang, Qiang. "Type-based proxy re-encryption and its construction."International Conference on Cryptology in India. Springer Berlin Heidelberg, 2008.
10. Ateniese, Giuseppe, et al. "Scalable and efficient provable data possession."Proceedings of the 4th international conference on Security and privacy in communication netowrks. ACM, 2008.
11. Benjamin C. M. Fung, Ke Wang, and Philip S. Yu, Fellow “Anonymizing Classification Data for Privacy Preservation” IEEE transactions on knowledge and engineering 2007.
12. Wang, Guan, et al. "Inference analysis in privacy-preserving data re-publishing." Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on. IEEE, 2008.
13. Schwarzbach, Björn, et al. "Cloud Based Privacy Preserving Collaborative Business Process Management." Computer and Information Technology (CIT), 2016 IEEE International Conference on. IEEE, 2016.
14. Arora, Sankalap, and Satvir Singh. "The firefly optimization algorithm: convergence analysis and parameter selection." International Journal of Computer Applications 69.3 (2013)
15. K.-K. Muniswamy-Reddy, P. Macko, and M. Seltzer, “Provenance for the Cloud,” Proc. Eighth USENIX Conf. File and Storage Technologies (FAST ’10), pp. 197-210, 2010.
16. Farahani, Sh M., et al. "A Gaussian firefly algorithm." International Journal of Machine Learning and Computing 1.5 (2011): 448.
17. Nayak, Janmenjoy, Bighnaraj Naik, and H. S. Behera. "A novel nature inspired firefly algorithm with higher order neural network: Performance analysis." Engineering Science and Technology, an International Journal (2015)
18. Open Stack, http://openstack.org/, July 2012
19. Hadoop, http://hadoop.apache.org, June 2012.