IJETMR
International Journal of Engineering Technologies and Management Research
A Knowledge Repository

DATA STORING AND RETRIEVAL METHOD IN BIG DATA USING FUZZY BASED SCALABLE CLUSTERING ALGORITHMS

S. Muthuraj Kumar [*1]
[*1] Department of Computer Technology, Madras Institute of Technology (MIT) Campus, Anna University, Chrompet, Chennai - 600044, India

Abstract:

A massive volume of digital data holding valuable information, called Big Data, is produced each generation. To process and excavate such valuable information, clustering is commonly used as a data investigation technique. A huge amount of Big Data diagnostics contexts have been established to measure the clustering procedures used for big data analysis. There exists one and only framework called Fuzzy based mechanism which actually fits in for iterative method by associate in storage divisions and accessible. The proposed algorithm is motivated towards the design and implementation of fuzzy based clustering algorithms on big data, which could be present for clustering huge datasets due to their low computational necessities. In this paper, we propose Random Data Storing with Optimization Fuzzy Logic algorithm (RDS-FLA) applied on cluster data to handle the tasks that are connected with big data clustering. Experimental trainings data taking place several big datasets have been showed. The performance of RDS-FLA is tried in evaluation with the proposed scalable form of the temporal fuzzy and Random data storing that is implemented on the big data cluster. The computation outcomes are recounted in terms of time and space complexity, run time and measure of clustering quality, showing that RDS-FLA is able to run in much less time without compromising the clustering quality. Thus, the algorithm proposed alleviates the processing time and increases the security of storage data effectively. Advantages such as cost optimization and efficiency in data security can be identified from the experimental results of proposed algorithm.

Keywords: Data Storage; Big Data; Fuzzy Logic; Clustering; Rds-Fla.

Cite This Article: S. Muthuraj kumar. (2019). “DATA STORING AND RETRIEVAL METHOD IN BIG DATA USING FUZZY BASED SCALABLE CLUSTERING ALGORITHMS.” *International Journal of Engineering Technologies and Management Research,* 6(6), 9-17. DOI: https://doi.org/10.29121/ijetmr.v6.i6.2019.389.

1. Introduction

A colossal volume of data becomes unruffled daily which is unpaid towards the collective participation of individuals in the digital space. Such massive amount of data holding valuable material is called Big Data. The situation stands charming more and more general toward mining such big data in order to increase awareness towards the treasured data that can be of more usage in technical and commercial applications. Clustering is a favorable data mining method that stands broadly accepted for data mining treasured data highlighting unlabeled files. Completed in the previous periods, various clustering methods have been established on several models and applications. Among them, partitioned methods are broadly approved owing to their small computational necessities, yet they exists additional appropriate clustering large datasets. The Fuzzy clustering methods efforts to divides the data opinions in the set of fuzzy clusters such that an independent function of a difference size is reduced. Various methods are suggested by the researchers based on partitioned clustering for handling big dataset.

From time to time, feature selection or structure identification in the framework of fuzzy rule mining is completed in an isolated stage, though certain approaches proceeds into versioning the knowledge machines then they regularly eliminate single feature by a period in a stepwise method. Therefore, the designated usual nature of structures cannot remain the greatest since the usual methods are designed for the problematic data at big hand and also since the usual structures are able to act together among themselves, and too, a feature is able to cooperate with the device that is used to solve the problem. In additional verses, the greatest features used for a neural network cannot be avoidably being the top for a fuzzy logic rule created system or for the support vector machines. In code, the usage of a comprehensive search in combination with the machine knowledge tool, which resolves the proposal to the ending system, will know how to solve the problem but is computationally too expensive.

In practice, the form of classification harms, fuzziness is associated through varied features of reasoning action within the human being. The bases of fuzziness are associated to tags communicated in arrangement space, such as glowing for instance, tags of modules occupied keen on account in cataloging measures. While our share of technical growth stake previously be located completed in the space of pattern classification, existing methods of pattern classification keep on lesser to the human classification procedures which achieve exceedingly difficult jobs. Therefore, try to improve a reasonable tool using fuzzy relational calculus logic for demonstrating and representing the reasoning procedure of human thinking for pattern classification.

Popular direction to rebuild big measure and great correctness in fuzzy reasoning maps to data model. The fuzzy modeling methods used to model fuzzy logic relations among changed notions in the procedure of bound for weighted graphs, wherever weights of edges represent the strength and type of relationships between two notions. The impartial knowledge of fuzzy logic is to find correct weights so that the reply of knowledgeable fuzzy is nearby to the experimental data as greatly as probable, which can be showed as an optimization problem. In the meantime the number of weights wants to be resolute growths with the number of notions; most existing fuzzy procedures can only handle problems with lots of nodes. However, the major network they treated has 30 inheritable factors. Therefore, more great education procedures that can handle important harms are necessary.

2. Literature Survey

There has been several research works contributed in the literature on several cloud storage methods and their retrieval procedures [6-11].

Jia Yu et al. [1] have analyzed a model that deploys the arrangement of binary tree along with providing information to the user about the secret keys by using a pre order technique. In order to

ISSN: 2454-1907
DOI: 10.5281/zenodo.3245179

grasp the onward security measures and the verifiability of block less belongings they propose in building an authenticator.

Kan Yang et al. [2] have proposed a data expressive, data efficient and data revocable method for data access control for multiple authority cloud data storage systems. The features are separated by multiple authorities which work integrated.

A revocable multiple right CP-ABE method has been discussed which can be valid it since the basic method to plan the data access control method. Their feature revocation method be able to resource fully attain to get her onward protection and toward the back protection. The examination and replication outcome demonstrate with the purpose of their proposed data access control method is protected in the indiscriminate oracle model.

Poornashree et al. [3] have proposed a verifiable data control method in which the consumers outsources the desired content to the distant data service provider. The service provider handles the data storing and preserving capabilities. Cloud service providers give access on their storage infrastructure to the end consumers by enabling a payment basis. The service providers should ensure whether all the data copies are stored and processed based on the agreement as well as the consumers need to verify if the data member providing the table hold distinct data. Some of the prominent tasks include data security, data dynamics and data storage in multiple clouds. Several works in PDP and its associated extensions are elaborated to achieve these tasks. Various types of PDP methodologies has been investigated and the key idea is by the comparison of optimal methods for achieving effectiveness and highly secure PDP.

Joseph et al. [4] has established a new 2-factor data security defense method for cloud data storage organization which comprises of data associated is allowable to encrypt content with knowledge of distinctness of a consumer only, at the same time as the recipient is necessary to make use of equally his/her top secret key and a defense mechanism in the direction of increase right to use to the facts. The clarification not no more than improve the privacy of the information except in addition recommends the revocability of the mechanism consequently that on one occasion the mechanism is withdraw, the matching cipher-text resolve exists efficient mechanically by the cloud data server exclusive of some observer of the data holder. There are notable number of works that elaborates different data storage mechanisms in cloud [12-20] which considers energy and security features in cluster data.

Wenjing et al. [5] proposed a protected cloud data storage method sustaining isolation protect for unrestricted inspection. The person responsible more extensive the effect allows the third parity auditor to carry out check for many clients at the same time and powerfully. Wide spread defense and concert scrutiny illustrate the proposed method are provably protected and very much capable.

3. Architecture of The Proposed System

Figure 1 contains eight modules namely User Interface, Data Operation Manager, Data Access Control Manager, Temporal Data Manager, Fuzzy Technique, Data Security Constraint Manager, Rule Base, and the Data Storage which depict the proposed architecture. This architecture enables storage of data and retrieving stores data in a more secure method from server.

ISSN: 2454-1907
DOI: 10.5281/zenodo.3245179

Several users are accessing the data storage in this system. The task of identifying the record holder and demanding the user interfaces is being handled by Data server manager. Any data that should be forwarded and provided to the users request creates a prominent level to all other data which is handled by Data Server Manager which settles data issues. Temporal Information Manager manages the activities of moving the data.

Every client is made to verify the constraints such as temporal, identity and their individual status level before getting permission to use the cloud database. The clients request are thus forwarded to the temporal manager which receives the request and identifies the requirements of clients. These requirements can comprise tasks such as storing, retrieving, updating and deleting data in secure manner. The data thus obtained from the client undergoes the map reducing methods like Sorting, Searching and Indexing, Classification, Joining, Term Frequency – Inverse Document Frequency. At last, Temporal Data Manger receives the reduced data.

Fuzzy Technique is used to allow only users for security data access with Boolean numbers. The compact data has are stored safe keeping by Data Security Constraint Manager Module. The Data Security Constraint Manager unit summarizes the user demands and identifies the encryption and decryption user demands. In order to determine the methods used in encryption and decryption values Hill cipher technique is utilized. This technique yields better security level to the data received from users by manipulating Data Access Control Manager.

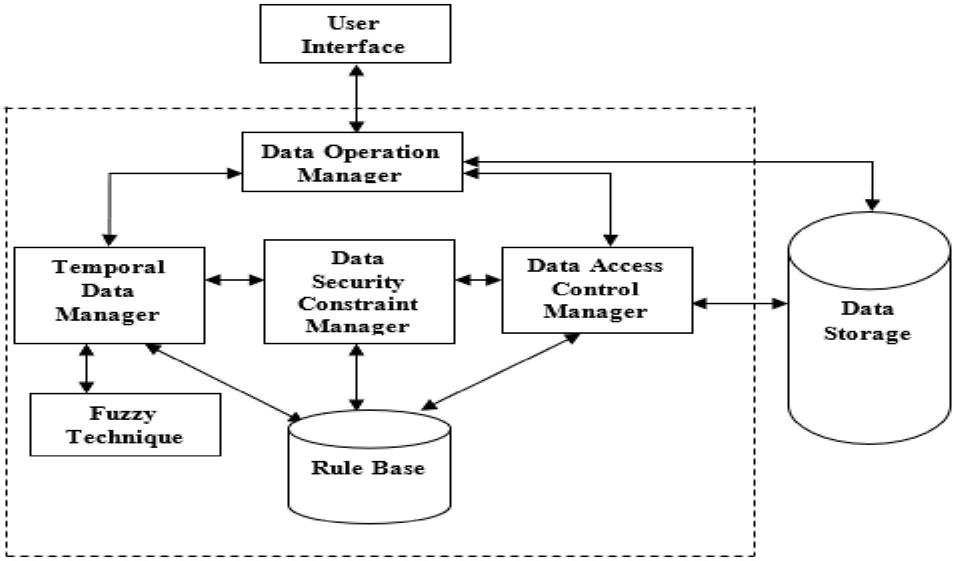

Figure 1: System Architecture

Different data obtained from the users is gathered as knowledge and store in the form of rules and referred as the rule base. The system should ensure ease of access in order to increase its efficiency. So, the frequent users are provided with this ease which yields enhanced security by encrypting user content using Data Access Control Manager.

Data base stores the data requested from the end users. It manages the methods involved in data storage and retrieval which is distributed in several places in huge volumes. Additionally, facilities such as effective storage and retrieval methodologies are provided by data storage manager.

The results section should provide details of all of the experiments that are required to support the conclusions of the paper. The section may be divided into subsections, each with a concise subheading.

It is advised that this section be written in past tense. It is a good idea to rely on charts, graphs, and tables to present the information. This way, the author is not tempted to discuss any conclusions derived from the study. The charts, graphs, and table should be clearly labeled and should include captions that outline the results without drawing any conclusions. A description of statistical tests as it relates to the results should be included.

4. Data Storing and Retrieval Fuzzy Logic Method

In this work, a new algorithm called Data Storing and Retrieval Fuzzy Logic Method Algorithm (DSRFLA) has been proposed for optimizing the overall performance. The overall idea of the proposed algorithm is to schedule and operate periodically by identifying the temporal constraints to increase the security. The proposed Data Storing and Retrieval Fuzzy Logic Method Algorithm consists of three phases namely Literal Fuzzy Minimize, Random Clustering Fuzzy Minimize and Mapping temporal clustering and random clustering joining which performs access control based on temporal limits.

4.1. Literal Fuzzy Minimize (LFM)

N – sized data
$D = (d_1, d_2, \dots\ d_n\}$
Objective function:

$$F_s(X, Y') = \sum_{a=1}^{n} \sum_{b=1}^{m} x_{ab}^{s} \, \|d_a - y'_b\|^2.$$

Input: D, Y, m, s
Output: X, Y’
Random initialize cluster centers $Y = (y_1, y_2, \dots\ y_m\}$.
Compute cluster membership.

$$x_{ab} = \frac{\|d_a - y_b\|^{\frac{-2}{s-1}}}{\sum_{l=1}^{m} \|d_a - y_l\|^{\frac{-2}{s-1}}} \ \forall a, b.$$

Check the constant.

$$\sum_{b=1}^{s} y_{ab} = 1.$$

Compute the cluster centers.

$$y'_b = \frac{\sum_{a=1}^{n} [x_{ab}]^s d_a}{\sum_{a=1}^{n} [x_{ab}]^s} \ , \forall b.$$

If $\|Y' - Y\| < \in then\ stop, else\ go\ to\ 2.$

ISSN: 2454-1907
DOI: 10.5281/zenodo.3245179

4.2. Random Clustering Fuzzy Minimize (RCFM)

```
Input: D, Y, m, s
Output: X, Y'
Load D as $n_m$ sized randomly chosen subsets$D = (d_1, d_2, \dots\ d_n\}$.
Sample $d_1$from D without replacement.
$X, Y' = LFM\ (\mathrm{D}_1, \mathrm{Y}, \mathrm{m}, \mathrm{s})$.
    $for\ g = 2\ to\ k\ do$
$X_g\ , Y_g = LFM\ \left(D_g\ , Y_{g-1}\ , m, s\right)$
$end\ for$.
Compute the partition on full data set.
    $X = \sum_{a=1}^{k} X_r$
Compute cluster center $Y'\ with$ partition on full data set using from LFM algorithm step 4
Compute the objective function using formula.
    $F_s(X, Y') = \sum_{a\ =1}^{n} \sum_{b=1}^{m} x_{ab}^{s}\ \|d_a - \ y'_b\|^2$.
Return $X, Y'$.
```

4.3. Mapping Temporal Clustering and Random Clustering

```
Input: $d_a, Y$
Output: $< b,\ < h_{ab}, d_a >>$
For each $Y_a\ in\ Y\ do$
b = index of cluster center y.
    $h_{ab} d_a = \ d_{ab} *\ d_a$.
    $yield\ \ < b, < h_{ab}, d_a \gg$
    $end\ for$.
```

5. Performance Analysis

In order to evaluate the proposed model, we consider a cloud data known as eucalyptus. It is cloud setup which with holds private cloud data. The system setup involves Intel I55 processor which has 2.4 GHz and possesses 16 GB RAM which runs at 7200 RPM Western Digital 1 TB Serial ATA drive which has inbuilt 16 MB buffer.

ISSN: 2454-1907
DOI: 10.5281/zenodo.3245179

The outcomes obtained from the implemented algorithm are clarified using tables. Table 1. Represent the estimation on the number of needs for the user the Data Storing and Retrieval without Fuzzy Logic Method Algorithm and Data Storing and Retrieval with Fuzzy Logic Method Algorithm in 5 experimentations with different number of client demand. From Table1, it can be perceived that the proposed Data Storing and Retrieval with Fuzzy Logic Method Algorithm does better when compared by without fuzzy algorithm in avoiding the consumers. The proposed system provides an accuracy of 92% in hindrance of repellent users of different agents. It also elaborated the distribution of different methods proposed.

Table 1: Number of User Requirements without Access

Exp. No	**No. of User Requirements Tried**	**No. of Requirements denied by without Fuzzy Logic**	**No. of Requirements denied by with Fuzzy Logic**
Exp. 1	10000	3000	5000
Exp. 2	20000	6000	9000
Exp. 3	30000	10000	13000
Exp. 4	40000	13000	17000
Exp. 5	50000	14000	22000

The data results obtained by permitting several authorized consumers to test the Data Storing and Retrieval Fuzzy Logic Method Algorithm are plotted in Figure 2. By analyzing the figure, the authorization entrée with the proposed fuzzy logic has increased notably when compared with model without fuzzy logic. Further, the repudiated entrees are fewer with 20%.

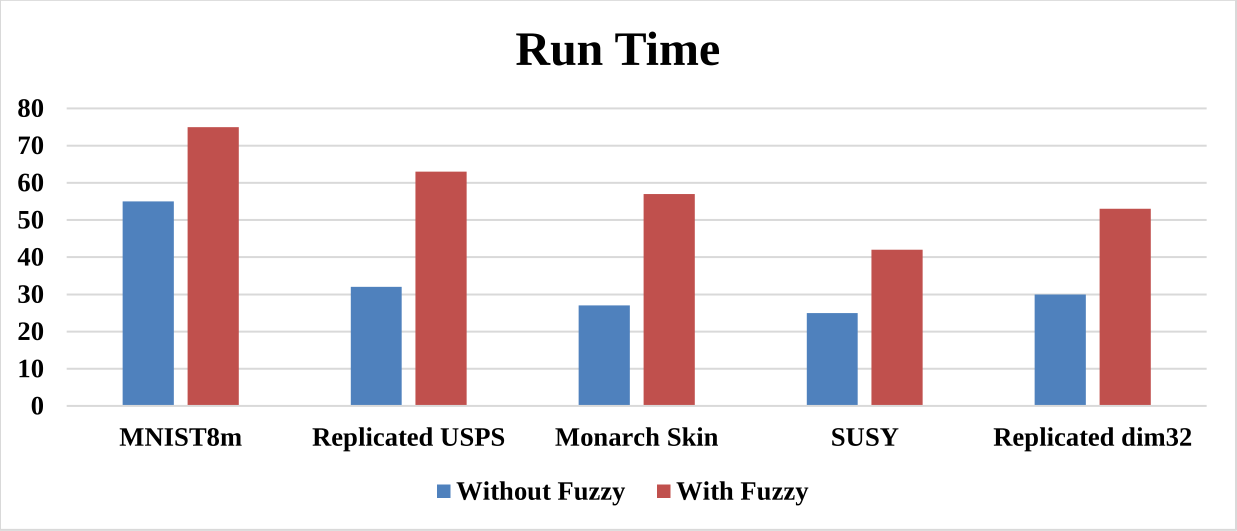

Figure 2: The Run Time Comparison Level

Figure 3. Shows the speedup necessities comparisons are for Data Storing and Retrieval Fuzzy Logic Method Algorithm for the time duration (t_1, t_2). It can be noted from the implementation, the results provided 10% reduction speedup accounting for the encryption oqing to the compression.

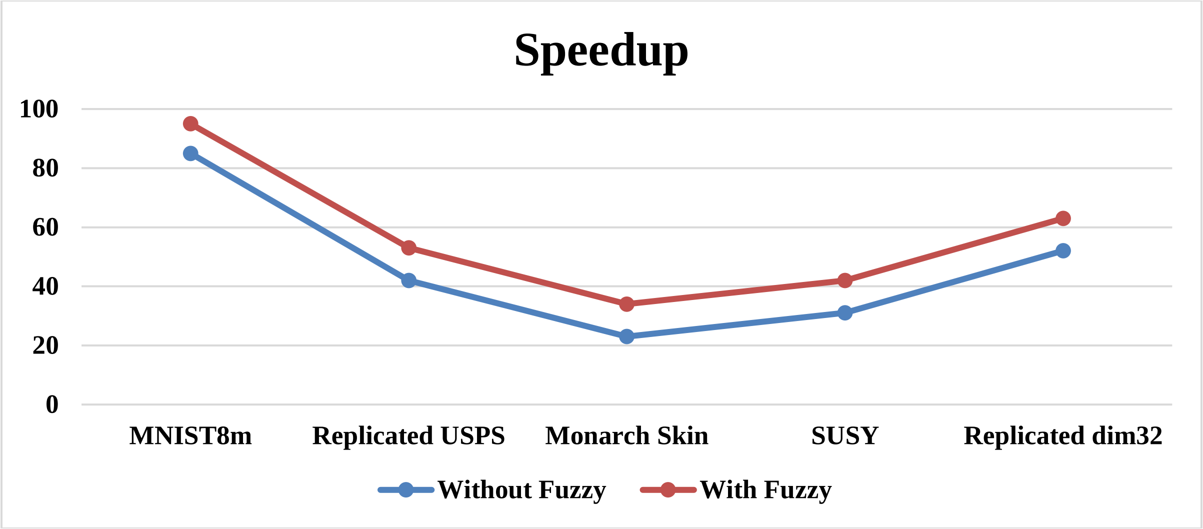

Figure 3: Speedup Analysis for Different Data Sets

People who contributed to the work but do not fit criteria for authorship should be listed in the Acknowledgments, along with their contributions. It is advised that authors ensure that anyone named in the acknowledgments agrees to being so named. Funding sources that have supported the work should also be cited.

6. Conclusion

In this paper, proposed a Data Storing and Retrieval Fuzzy Logic Method Algorithm for using store data in secured manner. The proposed method is implemented using fuzzy methods to arrange data with the experiments related with fuzzy data clustering for control Big Data. The Data Storing and Retrieval Fuzzy Logic Method Algorithm apportioned the big data into several amounts and develop the data facts current inside the chunk in a parallel manner. The additional main typical is that, throughout the implementation of the proposed method, which removes the need of storing the association environment, which types the execution of the proposed algorithm faster by reducing the run-time. This is a decent optimization approach for clustering of Big Data then, the association matrix will be also in an enormous data form to be stored.

References

[1] Jia Yu, KuiRen, Cong Wang and Vijay Varadharajan, “Enabling Cloud Storage Auditing With Key-Exposure Resistance”, IEEE Transactions on Information Forensics and Security, 10, 2015, 1167 – 1179.

[2] Kan Yang, XiaohuaJia, Expressive, Efficient and Revocable Data Access Control for Multi-Authority Cloud Storage, IEEE Transactions in Parallel and Distributed Systems, 25, 2014, 1735 – 1745.

[3] Poornashree B R, S Srividhya “A Survey on Provable Data Possession in Cloud Computing Systems,” International Journal of Engineering Research & Technology, 5, 2016,707 – 709..

[4] Joseph K. Liu, Kaitai Liang, Willy Susilo, Jianghua Liu, and Yang Xiang, "Two-Factor Data Security Protection Mechanism for Cloud Storage System", IEEE Transactions on Computers, 65, 2016, 1992 – 2004.

ISSN: 2454-1907
DOI: 10.5281/zenodo.3245179

[5] Muthurajkumar, S, Vijayalakshmi, M., and Kannan, A., Secured Temporal Log Management Techniques for Cloud, Procedia Computer Science, 46, 2015, 589 – 595.

[6] Ganapathy, S., Kulothungan, K., Muthurajkumar, S., Vijayalakshmi, M., Yogesh, P. and Kannan, A., Intelligent feature selection and classification techniques for intrusion detection in networks: a survey, EURASIP- Journal of Wireless Communications and Networking – Springer Open Journal. 271: 2013, 2013, 1 – 16.

[7] Aparna, K. S. M. V Kumar, "Privacy Preserving and Authorized Data Deduplication in Public Cloud Framework”, International Journal of Advanced Research in Computer Science and Software Engineering, 5, 2015, 772 – 775.

[8] Wenjing Lou, KuiRen , Qian Wang , Sherman S.M. Chow , Cong Wang, "Privacy-Preserving Public Auditing for Secure Cloud Storage" in IEEE Transactions on Computer, 62, 2013, 362 – 375.

[9] Hong Liu, HuanshenNing, QingxuXiong, Luarence T. Yang, "Shared Authority Based Privacy-Preserving Authentication Protocol in Cloud Computing" in IEEE Transactions on Parallel and Distributed Systems, 26, 2015, 241 – 251.

[10] Jia Yu, KuiRen, Cong Wang, "Enabling Cloud Storage Auditing With Verifiable Outsourcing of Key Updates" in IEEE Transactions on Information Forensics and Security, 11, 2016, 1362 - 1375.

[11] SwapnaliMorea, SangitaChaudhari, "Third Party Public Auditing scheme for Cloud Storage", in 7th International Conference on Communication, Computing and Virtualization 2016.

[12] Luca Ferreti, Michele Colajanni, MircoMarchetti, “Distributed, Concurrent and Independent Access to Encrypted Cloud Databases,” IEEE Transactions in Parallel and Distributed Systems, 25, 2014, pp. 437 -446.

[13] Muthurajkumar, S., Vijayalakshmi, M. &Kannan, A. “Secured Data Storage and Retrieval Algorithm Using Map Reduce Techniques and Chaining Encryption in Cloud Databases”, Wireless PersCommun (2017). doi:10.1007/s11277-017-4437-3.

[14] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, “Improved Proxy Re-Encryption Schemes with Applications to Secure Distributed Storage,” Proceeding Network and Distributed Systems Security Symposium (NDSS), 2005, 29-43.

[15] R. Lu, X. Lin, X. Liang, and X. Shen, “Secure Provenance: The Essential of Bread and Butter of Data Forensics in Cloud Computing,” Proceeding ACM Symposium Information, Computer and Comm. Security, 2010, 282-292.

[16] B. Waters, “Ciphertext-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization,” Proc. Int’l Conference Practice and Theory in Public Key Cryptography Conf. Public Key Cryptography, http://eprint.iacr.org/2008/290.pdf, 2008.

[17] V. Goyal, O. Pandey, A. Sahai, and B. Waters, “Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data,” Proceeding ACM Conference Computer and Comm. Security (CCS), 2006, 89-98.

[18] Zhongma Zhu, Rui Jiang, “A Secure Anti-Collusion Data Sharing Scheme for Dynamic Groups in the Cloud”, IEEE Transactions on Parallel and Distributed Systems, 2015, 130-146.

[19] S. Kamara and K. Lauter, “Cryptographic Cloud Storage,” Proceeding International Conference Financial Cryptography and Data Security (FC), 2010, 136-149.

*Corresponding author.
E-mail address: muthurajkumarss@ gmail.com