Securing Big Data Over Cloud Using Classification and Encryption Techniques

Gitanjali[1], Dr. Kamlesh[2]

[1]*CSE Department, LPU, Jalandhar, India,* [2]*Associate Professor, CSE Department, LPU, Jalandhar, India*

Abstract- Cloud computing provides the different types of services to the client over a network which is delivered by the third party and it reduce the burden from the user end. But security and privacy are the biggest issues in cloud computing. Now days researchers have devoted their work toward security over the cloud. This paper proposed the way to secure the sensitive big data over the cloud. This technique consist of three steps- Data Classification, Encryption and Cloud Utilization. The data classification is used to provide the effective level of security and secure the sensitive data.

This paper focuses on the use of various classification and encryption approaches and provides comparative study of major classification and encryption approaches. This paper concentrates on the problem of cloud operators security issues and attempts to avoid cloud users' data release from cloud servers. Our proposed mechanism aims to classify all sensitive and non sensitive data and encrypt all sensitive data and distributive store the data to the different cloud servers without causing big overheads and latency.

Keywords- Security, Privacy, Big data, Encryption, Data classification

I. INTRODUCTION

Cloud Computing is the boom in the field of medium world. Earlier user use to create application on the local server but if the local system crashes, the entire system and the application crashes automatically. To overcome this problem and to store data online in bulk cloud computing was brought into action. Branded Companies such as Google, Microsoft, Amazon and Face book have their own clouds [1]. Many organizations are shifting to cloud because of minimum investment, low cost and ubiquitous access services. Cloud Computing provide services such as Application-as-a services (SaaS), Platform as a carrier (PaaS) and Infrastructure as a carrier (IaaS) services.

Many cloud vendors provide attractive storage service offerings and scalable cloud-based storage spaces for users, such as Amazon, Dropbox, Google Drive, and Microsoft's One Drive [2]. However, the security issue caused by the operations on cloud side is still an obstacle of using Cloud Services. Many cloud users concern about their sensitive data to which the cloud operators have the access. Security threats are stumbling block in the success route of cloud computing.

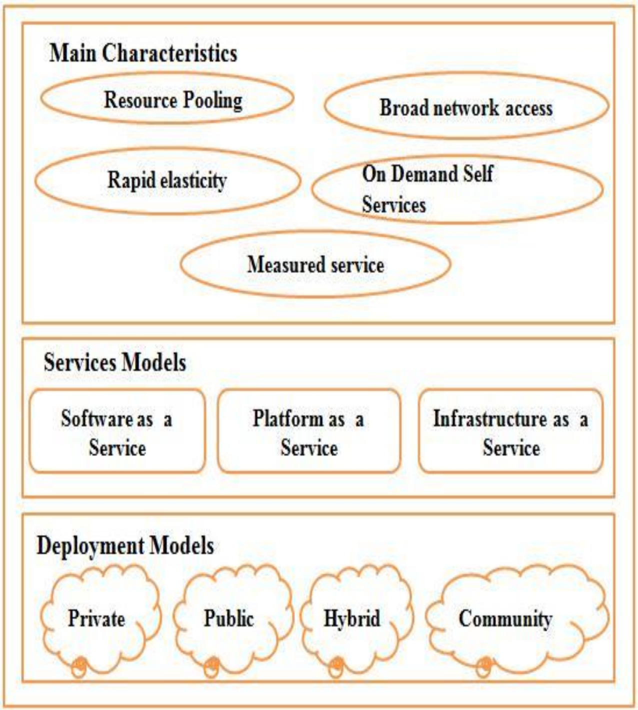


Fig.1: Model of NIST definition of cloud computing

In cloud computing user have no idea related to the physical location of data because it saved on the remote servers in which always a risk of confidentiality leakage. This work presents a security framework for security on cloud. This framework provides a collaboration of cloud providers service providers and consumers for effective security management.

Big data in the cloud computing is secured by using the intelligent cryptography approach. This approach divides the file and stores it on the distributed cloud servers. In this work another approach is also used which determine the data packet need to split for short the operation time. This approach provides the effective computation time with good security services [4].

II. RELATED WORK

A. Classification Based Approach for security

Faraz et al. [5] proposed the Variable Data Classification Index (VDCI) approach which is a variable data based on main three parameters i.e. availability, integrity and confidentiality. The value for this index is calculated without using the value of data owner or system admin. This is calculated using the history of stored data.

Munwar et al. [6] worked on the data confidentiality and data retrieval problem in the cloud computing. These issues are solved by using classification of data and cloud model. The problem is solving by using hybrid multi cloud model with data classification. This model is worked on the basis of multiple cloud, classification and different numbers of clusters.

Tawalbeh L et al. proposed a model which is based on the classification and provides secure cloud computing. This model reduces the overhead and processing time which is included in the security mechanism. It defines the security at different level with variable key sizes. The proposed model is tested with different security mechanism and it gives effective outcomes with high efficiency in proposed work [7].

Diwan V et.al. [8] proposed different cryptographic algorithms that have been compared and taken into account to ensure data confidentiality. In this different cryptographic algorithms are compared by considering different parameters like block size, key length type and characteristics. He provided the idea of a different cryptographic algorithm that can be used to ensure data security in the cloud.

Shaikh, Rizwana et.al [9] proposed classification method which works on the basis of different parameters. These parameters define the different dimensions. The data security can be provided according to the level and required protection. The proposed method solves the issue of data leakage and privacy protection.

B. KNN Based Security Approaches

Munwar Ali Zardari et al. [10] presented the K-nearest neighbor classifier for providing data confidentiality in the cloud based data. The approach is applied on the virtual cloud and it classifies the data according to the security needs of it. KNN classifier classifies the data into two classes that are sensitive and non-sensitive data. This classification of data mentions which data needs to be more security. The security to the data is provided by using RSA algorithm by encryption process. This work is done on the Cloudsim simulator and gives the effective results by deciding which data needs security.

C. Cryptography Based Security Approaches.

Sandip K. Sood et al. [11] introduced a combined approach which provides the data security in cloud computing. In this work different techniques are combined together to provide the effective security from the sender to receiver ends. The security of data provided to the user on the basis of data confidentiality, integrity and availability. The data security is provided by secure socket layer using encryption mechanism and integrity is provided by using MAC (Media Access Control). The security is enhanced by using the login id and password method to all the users.

Sengupta et al. [12] designed a security system for cloud computing by using cryptography. The cryptography in this work is done by using hybrid ceaser cipher encryption method. It provides security to the cloud on client, server and network location. This method provides the effective security from the hackers.

D. RSA Algorithm Based Approach

Somani U et.al proposed RSA algorithm which is utilized to guarantee the confidentiality part of security while Digital marks were utilized to improve greater security by confirming it through Digital Signatures. The approach utilized carryout encryption in five stages. In initial step, key is produced. In second step, advanced marking is performed and in stage 3 and stage 4 encryption and decoding is done. In last advance Signature confirmation is performed [13].

E. AES Algorithm Based Approach

Rewagad P et al. proposed a design to secure confidentiality of information put away in cloud by influencing utilization of computerized mark and Diffie Hellman to key trade with Advanced Encryption Standard encryption (AES) algorithm. Regardless of whether the key in transmission is hacked, the office of Diffie Hellman key trade make it pointless in light of the fact that key in travel is of no utilization without client's private key, which is given just to the true blue client. This three way instrument proposed design makes it extreme for programmers to break the security framework, in this manner ensuring information put away in cloud [14].

Prabhakar and Joseph et al proposed an information encryption procedure in light of the AES algorithm. In the cloud condition AES approach shields the information from start to finish for the whole life cycle. This encryption procedure utilizes AES-256 algorithm for encryption and Secure Socket Layer for ensuring the information records amid exchange to the cloud. The information proprietor or framework chairman scrambles information utilizing AES algorithm and then utilizing SSL for transfer information record security to the cloud. The proposed method guarantees that give finish security to information amid all stages and it's isolated into two stages. To start with stage manages information encryption and transfer information safely in the cloud and next stage manages information recovery which incorporates confirmation of clients and information decoding. In first stage information encryption is finished by AES - 256 encryption. In second stage client should be verified, the client sends the username and secret word to the cloud. At the point when cloud gets the demand from the client at that point confirms the client's subtle elements, if client is substantial at that point begin the procedure of information recovery [15].

F. Blow Fish Based Approach

Khatri N et al. [16] proposed blow fish algorithm to provide the security to data. This is a symmetric algorithm worked similar as DES algorithm. This is also working on block code approach with size 64 bits. It defines the two boxes that are s-box and P-box. This work is based on the variable length of the block cipher.

III. PROPOSED WORK

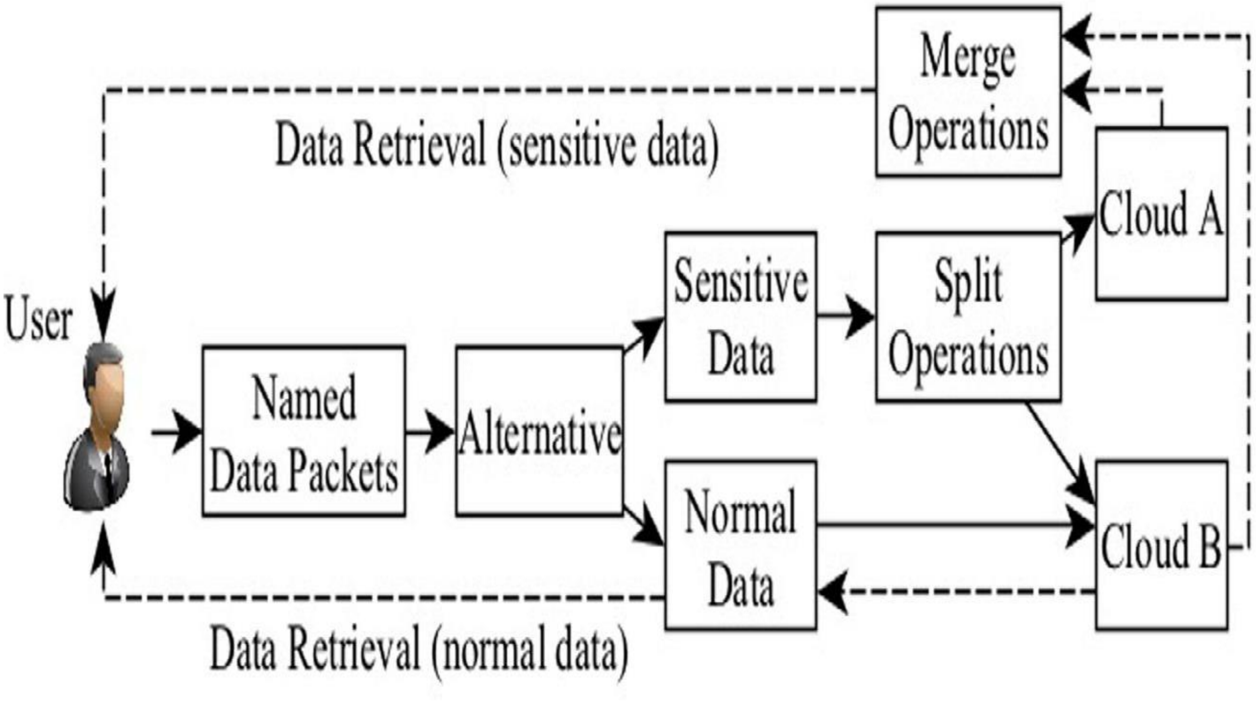
The research involves exploring various data classification algorithm such as KNN, Naïve Bayes and improved Naïve Bayes algorithm and analyzes their performance.

The present work is based on the secure data classification model which is based on the sensitivity level of the data and classifies according to this level. This approach encrypts the sensitive data only and stores it on different cloud and another cloud is used to store the non-sensitive data for the efficient utilization of data.

The goal of the proposed work is to provide the better results than the existing algorithms by using parameters accuracy, time and also enhance the confidentiality and integrity of the cloud data.

The below given parameters are used to analyze the performance of the proposed system:

1. Time taken for classifying the data
2. Accuracy of the classified data
3. True Positive rate
4. Encryption Time
5. Decryption Time



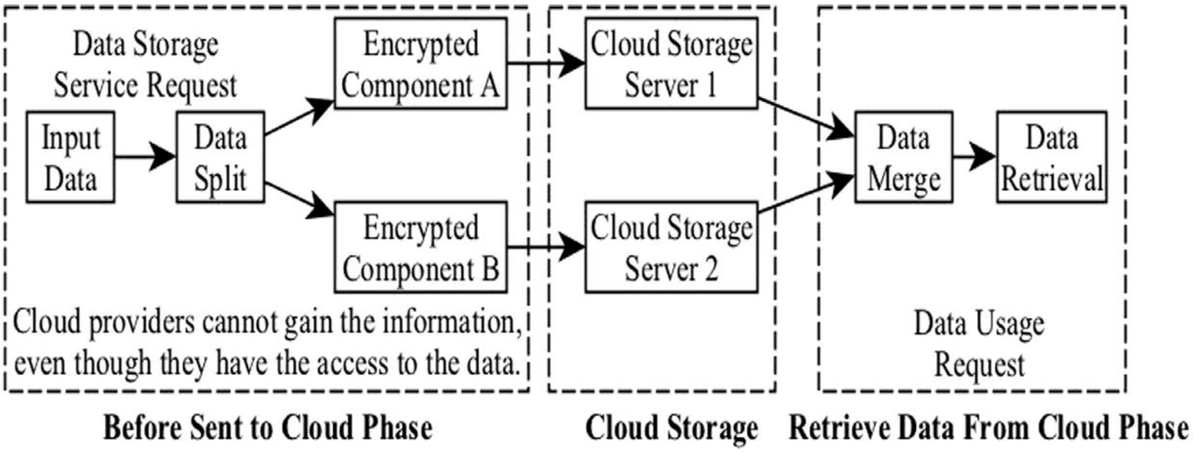

Data Storage Service Request
Input Data
Data Split
Encrypted Component A
Encrypted Component B
Cloud providers cannot gain the information, even though they have the access to the data.
Cloud Storage Server 1
Cloud Storage Server 2
Data Merge
Data Retrieval
Data Usage Request
Before Sent to Cloud Phase
Cloud Storage
Retrieve Data From Cloud Phase

Fig.3: Encryption decryption Process on cloud [4]

Comparison of data encryption approaches which is previously used.

Author's Name	**Techniques used for Encryption**	**Pros**	**Cons**
Prabhakar et al	AES	This approach prevents the data from brute force attack and provides effective security to data.	It does not effect more on privacy and efficiency of data.
Nandita et al	Caesar cipher and Vigenere cipher	This approach provides effective security by using the concept of triple encryption.	The drawback of this approach is it takes high computation time because it is based on the concept of homomorphc encryption.
Feng Zhao et al.	Homomorphism Encryption Method	It provides the effective data security to the data on cloud and also the data stored on the cloud.	The disadvantage in this work is homomorphism encryption which takes high computation time.
Xin Dongy et al.	Hierarchical Identity Based Encryption (HIBE)	It provides effective data security with low overhead and communication.	The major drawback of this work is user accountability and resistance of collusion.

Table 1: Comparison of data encryption approaches which is previously used

IV. CONCLUSIONS

In this research, a data privacy technique in the cloud environment is proposed. The goal of the research was to implement information security prerequisites that divide data into sensitive and non-sensitive data using an enhanced machine learning algorithm. The fundamental contribution of this security model is the confidentiality of data and the classification of data using an automatic learning classification approach. The classified confidential information is then encrypted using different cryptographic techniques such as blowfish and is stored in the cloud server. The proposed system was simulated in a cloud simulation environment designed using a cloudsim simulator. The results show that the proposed technique is more relevant than storing data without deciding on data security needs. In addition, the results show that the improved naive Bayesian technique works better than the K-NN classification technique in terms of accuracy, classification time and TP and encryption and decryption times also shows that security is more enhanced in the proposed work.

In the future, other security requirements may be taken into account in making the classification decision using an automatic learning algorithm. In addition, to improve security at the authentication level, image sequencing passwords based on different themes will be used to prevent unauthorized access to the cloud environment. Authentication security can be extended to a multi-level authentication scheme so that each user has different access permissions and roles. The availability of encrypted data can also be improved in the future.

V. REFERENCES

[1]. Gitanjali, Sukhjit Singh Sehra and Jaiteg Singh. Article: Policy Specification in Role based Access Control on Clouds. International Journal of Computer Applications 75(1):39-43, August 2013

[2]. Munwar ali zardari, Low Tang Jung, Nordin Zakaria," K-NN Classifier for Data Confidentiality in Cloud Computing", IEEE, pp.1-6, 2014.

[3]. Almorsy, M., Grundy, J., & Ibrahim, A. S., "Collaboration- Based Cloud Computing Security Management Framework" IEEE conference of cloud computing, Washington (DC), pp. 364-371,2011.

[4]. Li, Yibin & Gai, Keke & Qiu, Longfei & Qiu, Meikang & Zhao, Hui. (2016). Intelligent Cryptography Approach for Secure Distributed Big Data Storage in Cloud Computing. Information Sciences. 387. 10.1016/j.ins.2016.09.005.

[5]. Faraz Fatemi Moghaddam, Moslem Yezdanpanah ,Touraj Khodadadi 2014 VDCI: Variable Data Classification Index to Ensure Data Protection in Cloud Computing Environments, IEEE Conference on Systems, Process and Control (ICSPC 2014), pp.53-57 2

[6]. Munwar Ali Zardari, Low Tang Jung, Nordin Zakaria, 2013 Hybrid Multi-cloud Data Security (HMCDS) Model and Data Classification IEEE Advanced Computer Science Applications and Technologies (ACSAT), pp. 166-171 2

[7]. Lo'aiTawalbeh NS, Raad S. Al-Qassas and Fahd AlDosari,"A Secure Cloud Computing Model based on Data Classification". InFirst International Workshop On Mobile Cloud Computing Systems, Management and Security (MCSMS-2015) 2015 (Vol. 52, pp. 1153-1158).

[8]. S Diwan V, Malhotra S, Jain R. Cloud security solutions: Comparison among various cryptographic algorithms. IJARCSSE, April. 2014 Apr.

[9]. Shaikh, Rizwana, and M. Sasikumar. "Data Classification for achieving Security in cloud computing." *Procedia Computer Science* 45 (Elsevier-2015): 493-498.

[10]. Munwar Ali Zardari, Low Tang Jung, Nordin Zakaria, 2014 K-NN Classifier for Data Confidentiality in Cloud Computing, IEEE Computer and Information Sciences (ICCOINS), pp. 1 – 6

[11]. S. K. Sood, 2012 A combined approaches to ensure data security in cloud computing, ACM, Journal of Network and Computer Applications, vol. 35, no. 6, pp. 1831–1838.

[12]. Nandita Sengupta, Jeffrey Holmes 2013, Designing of Cryptography Based Security System for Cloud Computing, IEEE International Conference on Cloud & Ubiquitous Computing & Emerging Technologies pp. 52-57

[13]. Somani U, Lakhani K, Mundra M. Implementing digital signature with RSA encryption algorithm to enhance the Data Security of cloud in Cloud Computing. InParallel Distributed and Grid Computing (PDGC), 2010 1st International Conference on 2010 Oct 28 (pp. 211-216). IEEE.

[14]. Rewagad P, Pawar Y. Use of digital signature with Diffie Hellman key exchange and AES encryption algorithm to enhance data security in cloud computing. InCommunication Systems and Network Technologies (CSNT), 2013 International Conference on 2013 Apr 6 (pp. 437-439). IEEE.

[15]. D.M. Prabhakar, K.S. Joseph, 2013, A new approach for providing data security and secure data transfer in cloud computing, International Journal of Computer Trends and Technology (IJCTT) pp 1202-120

[16]. Ms. Neha Khatri Valmik and Prof. V. K Kshirsagar, "Blowfish Algorithm", IOSR Journal of Computer Engineering (IOSR-JCE), Volume 16, Issue 2, Ver. X (Mar-Apr. 2014), PP: 80-83, www.iosrjournals.org, eISSN:2278-0661, ISSN:2278-8727.