World Journal of Engineering Research and Technology
WJERT
www.wjert.org
SJIF Impact Factor: 5.924

TEXT CIPHER MULTI-SHARING CONTROL FOR BIG DATA STORAGE WITH PRIVACY-PRESERVINGCIPHER TEXT

**[1]Dr. Rajesh K. S., [2*]Dr. R. Balakrishna and [3]Dr. S. Shamshekhar Patil

[1]Associate Professor, Department of CSE, RRCE, Bangalore.
[2]Dean & Professor, Department of CSE, RRCE, Bangalore.
[3]Associate Professor, Dept of CSE, Dr. AIT, Bangalore.

Article Received on 21/01/2022 **Article Revised on 11/02/2022** **Article Accepted on 01/03/2022**

***Corresponding Author**
Dr. R. Balakrishna
Dean & Professor,
Department of CSE, RRCE,
Bangalore.

ABSTRACT

In today's world, consumers and businesses prefer to upload their files and information to Hadoop because of its large storage capacity and processing power. The confidentiality of the data is a fundamental security restriction for large-scale data storage. In order to meet the requirement, some encryption technologies are employed. Utilizing PKE, the sender of data can encrypt it so that only the valid recipient has access to it; this is done by using a cryptographic key as the receiver's public key (Public Key Encryption). However, this may not meet the needs of all customers in terms of massive data storage location. We've used a technique called multiple hop identity-based re- encryptions to protect the data in this case. Individuality is used to construct the keys, andidentity-based encryption is used to encrypt the individuality. This uses Message-Digest5 and DESwith password-based encryption as well as Message-Digest5.

KEYWORDS: Privacy, multiple hop, conditional sharing, bid data, Password based Encryption,MD5 and DES.

I. INTRODUCTION

Big data also refers to large amounts of records or data. With the term "BigData," a collection or group of data that is both large and growing at an exponential rate is described. Data is so large and difficult to manage that no out-of-date data management technologies are able to store or process it efficiently. One of the biggest issues with big data is storing and

analysing vast amounts of continually expanding, heterogeneous data, and then determining how to best safeguard that data. But how much data we have has nothing to do with Big Data's use.

There are three ways to implement Big Data

Structured: The term "structured" data refers to data that can be organised, stored, and retrieved according to a predetermined framework.

We're now experiencing issues as data quantities grow to enormous proportions; average data sizes were in the order of many zettabytes. A good example of structured data is a database table called 'Student'.

Unstructured: Any information or data with unidentified system is called as unstructured data. Size being massive, un-structured data creates several challenges like processing and developing worth out of it. An example of unstructured data isEmail.

Semi-structured: Semi-structured data contains both of the above stated types of data. It refers to data presented in a JSON (JavaScript Object Notation) file and web application information or data.

Characteristics of Big Data are as follow: **Volume:** There is a huge volume of information or data that is available, and the volume remains toincrease. Since the vast data cannot be stored on out-dated systems, we use distributed system where each parts of data are stored on differentlocations and united together by the software.Finally, big data size can be increased to infinitenumber.

Variety: Variety means many categories of data can be used. We just don’t have only structured data that is suitable into data table. Nowadays data’s are unstructured. Innovative and new big data tools now allow organized and inefficient data to be collected, put in storage, and used concurrently.

Velocity: Velocity refers to the rate at which new data is generated. Using Big Data Velocity, it is possible to compress the pace of data flow from a variety of sources including business developments, application records and networks, as well as social media websites, mobile devices, and sensors. There is a never-ending and massive influx of data.

In addition to these industries, big data has applications in finance as well as government and health care.

Hadoop is an OSS framework that permits to process and store BigData. Scalability is Hadoop main strength. It improves from working on a single point in a network to thousands points in a network without any issue in a unified manner.Google's **Doug Cutting** and his group fellows developed an OSS project called as Hadoop which permits to handle the very huge volume of data. Applications run by using Hadoopon the source of Map Reduce where the data is managed and complete entire arithmetical study on huge volume of data.

Hadoop is a concept or framework that is written in Java programming language that permits handling of huge data or dataset in a distributed manner among group of computers using easyprogramming types.

Hadoop can work on any mounted file_systemthat is distributed such as HFTP FS, local FS, and S3FS etc. The file system most commonly used byHadoop is HDFS.

HDFS is for storing large data files that are running on groups of hardware. When the data go beyond the storage capacity on a single machine,it is essential to divide it through number of discrete machineries. A storing mechanism amongmultiple machines network is managed by a file system called as distributed file_system. HDFS utilizes master-slave architecture. HDFS is kind ofthat software.

Hadoop framework includes four modules

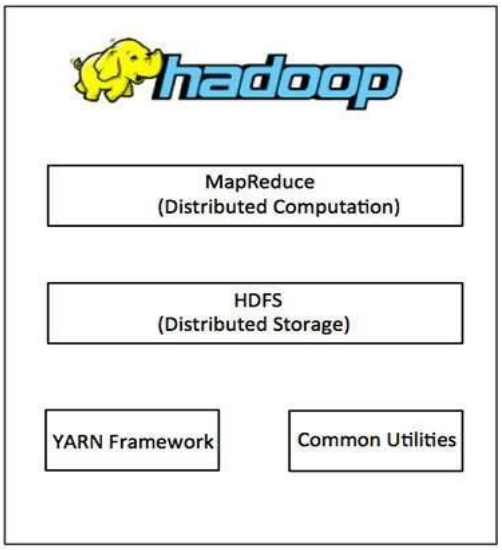

Figure 1: Hadoop Framework Components.

Hadoop_Distributed_File_System(HDFS)

HDFS is the distributed storing system that is designed to afford good-performance approach to data through multiple points in network in a cluster. More than terabytes of data are stored in HDFS and being used by applications. **Map Reduce:** This programming model and processing approach is referred to as Map Reduce, and it is used when the various components of a system areplaced on multiple computers that are networked using Java. The Map Reduce algorithm is made up of two parts: a task called Map, and a function called Reduce. To begin, Map takes a set of data and turns it into a new set of data in which certain items are broken down into tuples or rows (key/value pairs)... A reduced collection of rowsor tuples is created by concatenating the input from a map with the result from reduce. Map Reduce, as its name implies, is a data reduction technique., **Map**: This task is executed first. **Reduce:** Using the output of map then Reduce jobis performed.

YARN: Yarn was developed in the Hadoop_2.x.Yet_Another_Resource_Negotiator islarge-scale, software of a collection of operational points in network for BigData applications. YARN was considered to be the Map Reduce_2 or next generation MapReduce. It has grouping platform that helps for managing resources and scheduling tasks. Management of resources for particular applications was developed and set up by YARN.

II. Literature Survey

In the software development process, literature surveys play a critical role. Organization measures to prevent unauthorised access make up the specific topic of networking safety. Large systems with internal constraints are classified using the Data Loss Prevention (DLP) approach.

A network intrusion prevention system (IPS) monitors network traffic flow to detect and prevent malware from entering the system. **Symmetric_Key_Cryptography** In addition, there is a way in which the sender and receiver share a common key. In June 1976, it was the soleway to encrypt data.

Advanced_Encryption_Standard- AES is created on a design standard known as a SP-network, and is effective in both hardware and software.

Data_Encryption_Standard- DESis asymmetric_key encryption. DES is an application of a Feistel Cryptograph.

Whitney Diffie and Martin Hellman proposed a public key cypher encryption in 1976, which uses two keys, one public and the other private, that aredifferent, but arithmetically connected with oneanother.

Other public key system (RSA) was created in1978 by Adi.S, Ronald.R, and Len.A.

J.H.Ellis invented asymmetric key cryptography in 1997.

According to research done in 2009 by Giuseppe A., Karyn B et al, it is possible to transform a cypher text using one unique key into an image or code of the same data under a different key, whichis known as a proxy re-encryption (PRE).

An atomic proxy re-encryption application was proposed in 2005 by three authors: Giuseppe.A, Karyn.B, and Susan.H. It uses a semi-trusted third party (a proxy) to turn Alice's cypher text into Rachel's encrypted text without looking at the plain text underneath.

In 2004, Dan.B and Xavier.projected Identity_Based_Encryption(IBE) schemes that are selective_identity secure IBE. Alittle weaker security system than the typical security systemfor IBE is Selective_identity secure IBE.

In 2005, Dan.B, Xavier.B, and E.J. Gohsuggested and projected a system called as **Hierarchical_Identity_ Based_Encryption** (HIBE) where the decryption needs only two computations of bilinear record and encrypted text consisting of three group of elements, neverthelessof the hierarchy deepness. Encryption is as real asin another HIBE system.

In 2009, Jan C, Markulf. K, Alfredo. R and Caroline. Projected an encryption scheme that permits users to privately examine encoded data in a public_key situation by keywords and decrypt the searched results.

III. System Architecture

It is the theoretical process by which the behaviour, structure and other perspectives of the system are described. Explaining architecture entails giving a formal demonstration of the system and providing information about its structure and behaviour. Externally apparent system properties, as well as relationships between them, can be included in a system's architecture. That plan can then be used to create goods and systems that function in concert to put the entire system in place, all of which work together.

Using Systems Architecture, you may define complicated systems with confidence and design them to their full potential, such as:

A company engaged in manufacturing (the originalmeaning of Systems Architecture)

A business that deals with information technology (Enterprise Architecture)

A group or a company (OrganizationalArchitecture)

A corporation (Business Architecture)

The implementation of a plan (ProjectArchitecture).

The system architecture is depicted in thefollowing diagram.

Figure 1.3 shows the system architecture divided into six modules, each of which is responsible fora specific set of tasks.

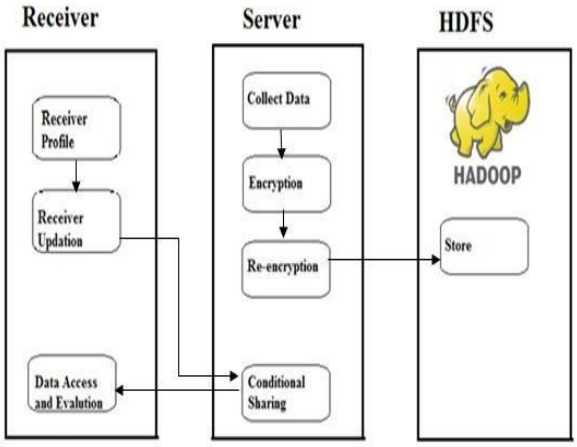

Figure 2: System Architecture.

Data Collection: Gathering all the relevant information is step one. To begin, preprocessing the dataset is required. Preprocessing consists solely of data purification. Before data can be changed or transformed in any way, it must be cleaned and selected. HDFS then receives and stores the data.

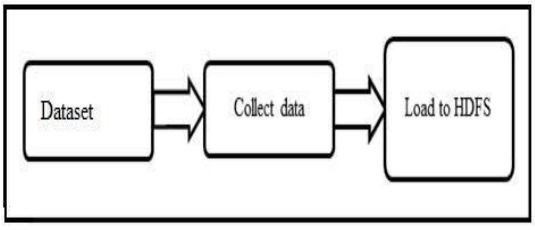

Figure 3: Data Collection.

Data Encryption: After collecting the dataset, weneed to cryptic the data. In this module, we use theEncryption method called Identity basedEncryption method (using Message-Digest5

andDES with encryption based on password). Theobtained encrypted data is again encrypted using AES and the obtained encrypted data is stored on to HDFS.

Password_Based_Encryption (Encryption)

Plain_text+pswd+new_salt PBEcipherPBE cipher Ciphertext

New_salt+base 64 encoded salt + CiphertextCiphertext= salt+ciphertext

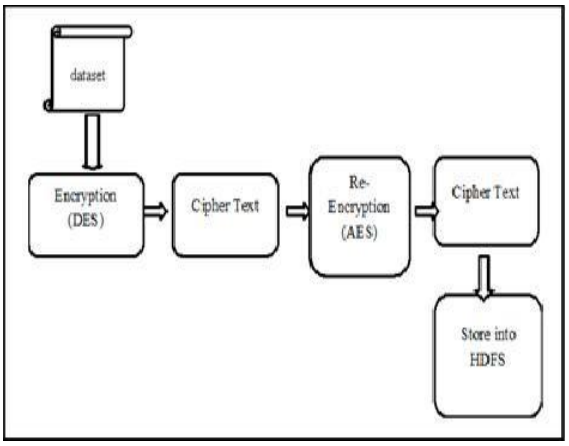

Figure 4: Data Encyrption.

Password_Based_Encryption (Decryption) Salt+Ciphertext + base64 decode Ciphertext, salt

Ciphertext+salt+pswd PBEcipherPBEcipherPlain_text

Receiver Profile: In this module, after encrypte and anonymize, data_set records has been sent to the multiple receivers. For the obtained ciphertext,the recipient of the cipher text can be updated in multiple times. We refer this process as "multiple-hop".

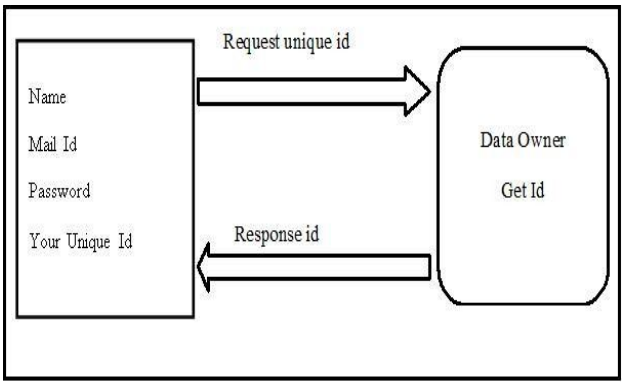

Figure 5: Receiver Profile.

Receiver Update: In this module, the user requests the owner for the category where the owner sends the appropriate response and displaysthe requested category on the text field.

Conditional Sharing: If the predefined conditions are met, a fine-grained cipher text can be shared with other users via this module. We have to validate the users and senders in this conditional sharing. A "all-or-nothing" sharing method is provided by this work, which restricts the flexibility of data sharing.

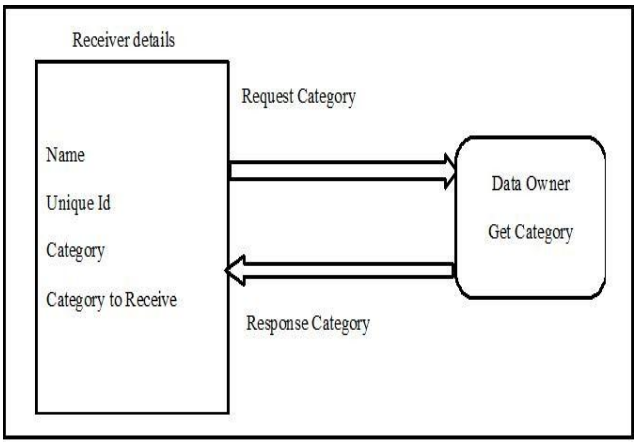

Figure 6: Multiple receiver update.

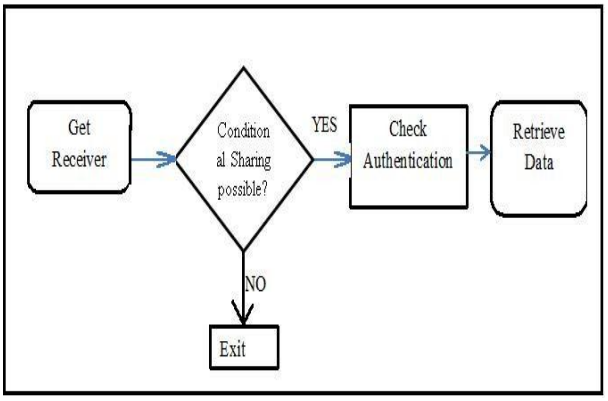

Figure 7: Conditional Sharing.

Data Access and Evaluation: In this module, based on the conditional sharing, here we decrypt and access the information's. Then, we estimate the performance of present system and projected system using the CCA-secure MH-IBPRE mechanism by using the parameters such as Time Efficiency and Privacy. Here we calculate the decryption time and encryption time of thealgorithm and finally we calculate the whole time for executing the algorithm.

V. RESULT

We compare our findings to those of the currently utilised system. In comparison to our approach, the existing system has a higher time efficiency. The security of our system is superior to that ofthe competition's.

VI. CONCLUSION

The ProxyR e-encryption with Anonymity and Conditionally Shared Ciphertext for Big Data achieves qualities such as multisharing and conditional sharing while also providing anonymity. An encrypted document can be transferred safely and conditionally several times with anonymous between recipients and senders using the Proxy Re-encryption mechanism described in this work.

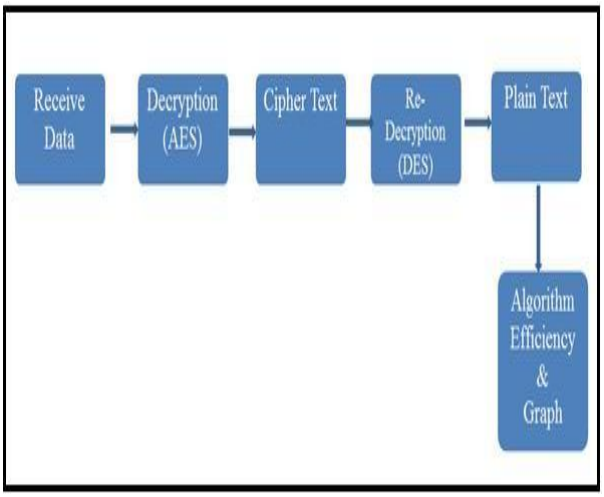

Figure 8: Data Access and Evaluation.

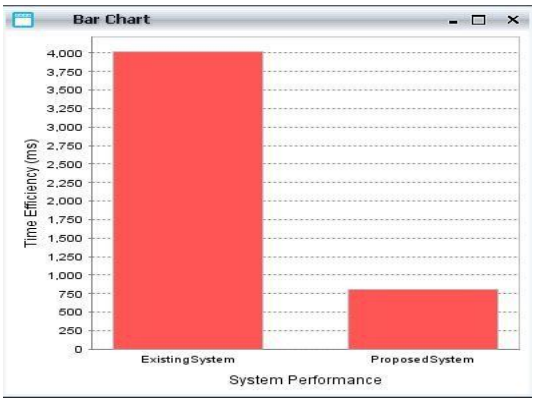

Figure 5.1: Bar Chart.

VII. Enhancement

Threshold re-encryption system is a method that offers great security. Using this method, you'll have dispersed storage servers and key servers. Because keeping encrypted keys on the same key- server as the users' private keys is a dangerous practise, we employ separate key-servers for that purpose. There are numerous security measures in place to keep these crucial servers safe. Each of the dispersed servers completes their tasks on its own. Data confidentiality and data robustness are key features of the cloud storage system. It is possible

to transfer data from one user's server storage to another user's server storage withouthaving to get the data back. The Re encryption system permits advancing operations over coded and encrypted messages as well as coding operations over encrypted messages, which is the fundamental methodological addition.

REFERENCES

1. G. Ateniese, K. Benson, and S. Hohenberger, "Key-private proxy re-encryption", in Topics in cryptology-CT-RSA (Lecture Notes in Computer Science), vol. 5473. Berlin, Germany: Springer- Verlag, 2009; 279-294.
2. D. Boneh and X. Boyen, "Efficient selective- ID secure identity-based encryption without random oracles," in Advances in Cryptology– EUROCRYPT (Lecture Notes in Computer Science), vol. 3027. Berlin, Germany: Springer- Verlag, 2004; 223–238.
3. D. Boneh, X. Boyen, and E.-J. Goh, "Hierarchical identity based encryption with constant size ciphertext," in Advances in Cryptology– EUROCRYPT (LectureNotes in Computer Science), vol. 3494. Berlin, Germany: Springer-Verlag, 2005; 440–456.
4. M. Bellare and S. Shoup, Two-tier signatures,stronglyunforgeable signatures, and Fiat–Shamir without random oracles," in *Public KeyCryptography* (Lecture Notes in Computer Science), vol. 4450. Berlin, Germany: Springer- Verlag, 2007; 201–216.