THE ROLE OF ENCRYPTION IN CLOUD-BASED BIG DATA SECURITY

Aggala Chiranjeevi
Research Scholar
Dept of Computer science & Engineering
Shri JJT University-Rajasthan.

Dr. Prasadu Peddi
Research Supervisor
Dept of Computer science & Engineering
Shri JJT University-Rajasthan.

Dr. Suneel Pappala
Co-Supervisor
Associate Professor
Dept of Computer science & Engineering
Lords Institute of Engineering And Technology, Hyderabad.

ABSTRACT

The proliferation of cloud computing and the exponential growth of big data have revolutionized the way organizations store, process, and analyze data. However, the benefits of cloud-based big data come with significant security challenges, as sensitive information is stored and processed in remote data centers operated by third-party providers. This abstract provides an overview of the crucial role encryption plays in enhancing the security of cloud-based big data environments. Cloud-based big data environments involve the storage and processing of vast datasets across distributed infrastructure, often accessible from various locations and devices. This diversity and accessibility introduce potential vulnerabilities that can be exploited by malicious actors. To mitigate these risks, encryption is a fundamental security measure that safeguards data both in transit and at rest.

Keywords: *Cloud, big data environments, encryption, security.*

Introduction

Big data refers to large volume of data in everyday lives. The data generation rate is growing rapidly. So, it is becoming extremely difficult to handle it using traditional methods or systems. The amount of data generated by social networking sites, sensor networks, internet, health care applications, and many other companies is drastically increasing day by day. All the huge amount of data generated from different source in multiple formats with very high speed is referred as big data. Meanwhile, all data generated may be structured form (relational data), semistructured form (XML data) and un-structured form which is not managed by traditional databases. Big data is heterogeneous data. Social media like face book, twitter, Google, generates huge amount of data daily, which is complex and unstructured in nature to handle by traditional databases. Identifying the source of problems will result in more efficient use of big data. This paper examines and classifies studies on security and privacy breaches and solutions in big data. These perspectives would lead to an understanding of important research areas and the development of new methods. Security issues are the merging concepts in big data. In information security a lot of encryption algorithm is widely used. Asymmetric key known to be public key and symmetric key is known to be a secret key encryption is the two classifications of the encrypted algorithm.

LITERATURE REVIEW

Amr M. Sauber et al (2021) The main goal of any data storage model on the cloud is accessing data in an easy way without risking its security. A security consideration is a major aspect in any cloud data storage model to provide safety and efficiency. In this paper, we propose a secure data protection model over the cloud. The proposed model presents a solution to some security issues of cloud such as data protection from any violations and protection from a fake authorized identity user, which adversely affects the security of the cloud. This paper includes multiple issues and challenges with cloud computing that impairs security and privacy of data. It presents the threats and attacks that affect data residing in the cloud. Our proposed model provides the benefits and effectiveness of security in cloud computing such as enhancement of the encryption of data in the cloud. It provides security and scalability of data sharing for users on the cloud computing. Our model achieves the security functions over cloud computing such as identification and authentication, authorization, and encryption.

M. Geethanjali (2019) In recent years, big data have been hot research topic. The interesting amount of big data also increases the chance of breaching the privacy of individuals. Due to a rapid growth and spread of network services, mobile devices, and online users on the internet leading to a remarkable increase in the amount of data. However, it is not only very difficult to store and analyse them with traditional applications. But also it has challenging data privacy and security problems. This paper shows the fundamental concept of big data, concerns on big data, technologies used and presents comparative view of big data privacy and security approaches in literature.

Advanced Encryption Standard (AES) For Secure Cloud storage

The security method for cloud storage relies on the Advanced Encryption Standard (AES), which was first released in 2001 by NIST. In theory, this symmetric block cipher might replace the widely used Data Encryption Standard (DDES). It is a method of encrypting information. Cloud-based data storage was preceded by encryption per the Advanced Encryption Standard (AES). While in possession of the main element used for security, the data owner arranges the appropriate data files in the human cloud and executes the decryption procedure when data must be recovered. Data that is being backed up across many servers with the assistance of a reliable auditor is encrypted using an AES algorithm at either the client or enterprise level to prevent unauthorized access. After the encrypted data has been transmitted to a reliable auditor, it will be split into several smaller files using a file splitting method.

Data Owner → Encrypted Data Using AES → Public Cloud

This leads us to the deciphering procedure. The necessary information may be easily obtained by downloading it from the cloud and making use of the DO.

Public Cloud → Encrypted Data Using AES→Data Owner→Decrypted Data

A 128-bit key has been used. After the data had been encrypted and sent. The information could be kept in a few different locations. It was taken by unknown individuals once it was within the system. The cloud service provider can't force deletion of this data since it isn't stored in a centralized location. The same

holds true for the massive amounts of data stored on the cloud.

PROPOSED FRAMEWORK

The planned structure is described in further detail in this section so that those in need of human recovery support may make the most of it. However, this even timing of the facts may be used to verify the client's operation. The data operator, who is often a company that generates massive volumes of data, generates and uploads the data to the consumers' cloud storage.

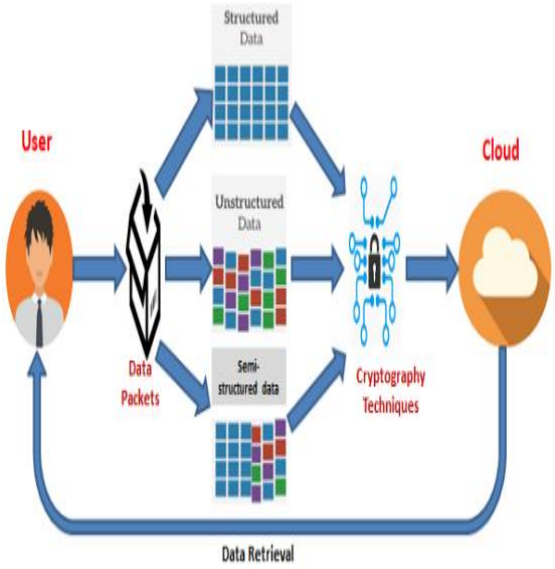


Figure: Secure blur storage and recovery using future structure

Finding an innate strategy that will allow you to successfully finish a real even version made to address the use problem is your current mission. The usefulness of the frame has been much enhanced in this updated version. Consequently, the suggested design has a means of connecting to a preexisting public cloud. The potential benefits of the proposed framework are shown in above Figure. In a number of different formats, well-known truths will be shown. The suggested paradigm may be able to deal with this knowledge heterogeneity since all three types are so common. If you can handle the variety in your data's origins and formats, you'll have a leg up on the competitors. Therefore, it is structured and searchable in terms of these three main categories. The design works well for storing and recovering massive amounts of data.

Algorithm: Protected varied Cloud Data Encryption and Decryption

$$selection\ of\ Data\ (Type)$$
$$if(Type = 1)$$
$$\{$$
$$Structered\ data$$
$$Data\ From\ SQL\ \ type\ data\ bases$$
$$AES\ Encrypt()$$
$$Send\ to\ cloud$$
$$\}$$
$$if(type == 2\ or\ type == 3)$$
$$\{$$
$$Unstructured\ or\ semi\ trusted\ data$$
$$Data\ from\ HBase\ or\ Casandra\ or\ MongoDB\ etc.$$
$$HomomorphicEncrypt()$$
$$\}$$

The next set of objectives places an equal priority on a person's physical well-being and their care. It's possible that the record format is the unifying factor. Given that semi-structured data is often handled similarly to a large, unstructured data set, the relevant computations for this kind of data are provided below.

Algorithm 3.2 AES based encryption algorithm

Algorithm AESEncrypt ()

{

You must infer the essential arrangement yourself.

Put out a whole chunk of information (plain text) on the state display.

To this make-believe rural gang, insert the spherical secret.

You have nine rounds to run your country.

Playing outside is the most common and last round of the nation.

Replay the preceding country's screen in the open air while the cipher text is deciphered.

}

The data does support the signs provided in both forms, but it does so in an overly elastic manner of data recovery. provides entry to a huge number of individual components, as well as spacing variations The framework performs any essential post-construction steps.

Execution Performance Evaluation

In this section, we discuss the time difference between the proposed and current methods and their execution. Information may be encrypted and decrypted using the provided email addresses, as was previously stated.

Table 1: Proposed and Existing Encryption, Decryption performance

Data Size (MB)	Execution Time (sec)			
	Encryption Existing	Encryption Proposed	Decryption Existing	Decryption Proposed
10	0.4057	0.3989	0.3945	0.2921
50	2.2237	2.1956	1.4879	1.3925
100	2.7937	2.5968	2.5472	2.0156
500	13.6537	12.9896	9.6734	9.0132

Table 1 displays the results of an examination into the encryption and decryption times required for both new and old methods. Figure 2 shows a clear number on a flat axis while the implementation time is represented in minutes. The results show just two distinct patterns.

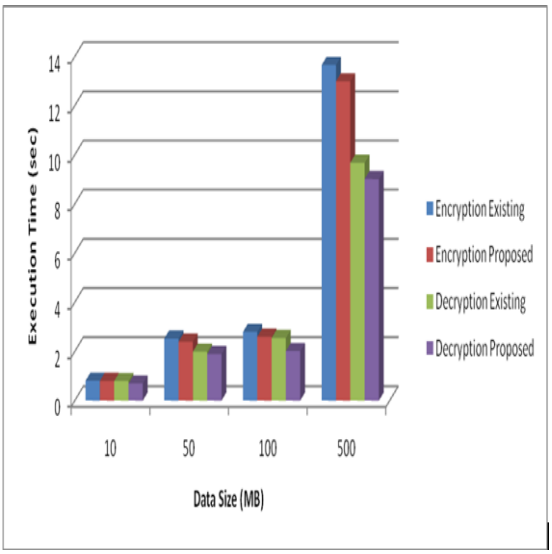


Figure 1: Performance period comparison

Magnitudes and timeframe for implementation. Data that might have an effect on actual practice. Even if the efficiency has been improved. A second trend was uncovered that contrasted with the previous trend, demonstrating the connection between the sizes of data and the timing of data operations in both planned and current systems. In terms of encryption and security, the proposed system performs better.

Comparison of Total Upload Time

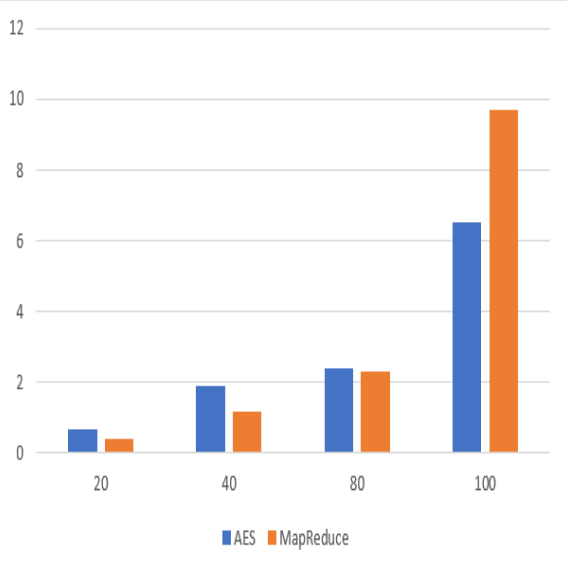
The projected total time, the existing processes, and the whole upload phase (the time it takes to load and then remove information immediately into blur) are all outlined in this section. Since their knowledge of information is measured in MBs, the total duration of the storage will be measured in seconds.

Table 3.2 shows that for both planned and existing operations, the whole duration of the upload phase is shown automatically. The data size is assumed to equal the total amount of time spent observing the data.

Table 2 Comparison of Data Size and Total Upload Time

Block Size (MB)	Total Upload Time (sec)	
	AES	MapReduce
20	0.6762	0.3767
40	1.9151	1.1578
80	2.3851	2.3258
100	6.5283	9.6881

Graph clearly displays the time needed for the absolute sum of information redirected to blur along the vertical axis, while the info measurement is shown along the horizontal axis. With a larger data set, the total upload time might be lengthened.



Graph 2 Block size vs. total upload time

This proposed current strategy also includes a time of uploading that is as important. Incorporating people's predicted security methods into the proposed platform is shown.

Conclusion

Encryption serves as the primary defense mechanism to ensure the confidentiality of big data in the cloud. By encrypting data before it is transmitted or stored in the cloud, organizations can protect it from unauthorized access and eavesdropping. Encryption acts as a last line of defense in the event of a data breach. Even if unauthorized parties gain access to encrypted data, they are unable to decipher it without the encryption keys, making the stolen data useless to them. Encryption is often a requirement to comply with data protection regulations and industry standards. It helps organizations demonstrate their commitment to data security and privacy, potentially avoiding legal penalties and reputational damage.

REFERENCES

1. *M. Geethanjali (2019) Big Data Security A Review of Encryption Techniques in Big Data, International Journal of Computational Intelligence and Informatics, ISSN: 2349-6363, Vol. 9: No. 3.*
2. *Amr M. Sauber, Passent M. El-Kafrawy, Amr F. Shawish, Mohamed A. Amin, Ismail M. Hagag, (2021) "A New Secure Model for Data Protection over Cloud Computing", Computational Intelligence and Neuroscience, vol. 2021, Article ID 8113253, 11 pages. https://doi.org/10.1155/2021/8113253.*
3. *Ray, I, Belyaev, K, Strizhov, M, Mulamba, D & Rajaram, M (2013), "Secure logging as service-delegating log management to the cloud", vol. 7, no. 2, pp. 323-334.*
4. *Rongmao Chen, Yi Mu, Guomin Yang, Fuchun Guo &Xiaofen Wang (2016), "Dual-Server Public-Key Encryption with Keyword Search for Secure Cloud Storage", vol. 11, no. 4, pp. 789-798.*
5. *Shulan Wang, Junwei Zhou, Joseph K Liu, Jianping Yu, Jianyong Chen, and WeixinXie (2016), "An Efficient File Hierarchy Attribute-Based Encryption Scheme in Cloud Computing", vol. 11, no. 6, pp. 1265-1277.*
6. *Kai Zhang, Ying Song, Haifeng Fong, and Yu-Zhong Sun (2015), "Trusted Connection System based on the Virtual Machine Architecture", ISSN: 978-1-4244-5540.*
7. *Kai Hwang & Sameer Kulkarni, Yue HU (2009), "Cloud Security with Virtualized*

Defense and Reputation-based Trust Management",

8. *Karim Chine "Scientific Computing Environments in the Age of Virtualization Towards a universal Platform for the Cloud", 978-1-4244-445.*