International Journal of Computer Science Trends and Technology (IJCST) – Volume 7 Issue 4, Jul-Aug 2020

RESEARCH ARTICLE OPEN ACCESS

BigData Security and Data Encryption in Cloud Computing

Naseemuddin Mohammad
IT Project Manager, Software Engineering Wipro Limited Hyderabad, India
Karuturi S R V Satish
Research Scholar, Computer Science and Engineering Mewar University Rajasthan, India

ABSTRACT
The cloud computing and the collaborative source security system for big data security are explained in detail in this paper. This research proposes a collaborative encryption technique framework to satisfy the needs of quicker encryption in the future. The whole security during cloud computing cannot be provided by a standard information security solution. The approach outlined in this study uses a distributed, parallel encryption system to reap the benefits of homomorphic encryption. It is laborious to use the encryption feature when communicating with an object via mobile. Because of the quantity of the huge data, the encryption and decryption process is slowed down by the security of the data. Big data cloud computing security cannot be achieved with a single encryption method based on a single source.

Each cloud has the ability to cooperate with other cloud servers and has its own security features. As a result, distributed and parallel encryption capabilities are available at every cloud's doorstep without interfering with the encryption process's order. The most important resources become the distant resources, and these resources may be allocated and managed across all clouds. The majority of the time, the network and other resources are available when using cloud computing. It is challenging to provide information security when resources are unavailable for encryption and decoding. With the help of the collaborative encryption approach, many clouds can operate concurrently with dispersed processing. The homomorphic encryption enhances the security mechanism.
Keywords – Data Security, Data Encryption, BigData, Cloud Computing

I. INTRODUCTION

Cloud computing has developed over the past few years from a potential commercial idea to one of the IT industry's fastest-growing divisions. Now, businesses struggling with the recession are beginning to realize that they can significantly increase their infrastructure resources or quickly obtain best-of-breed business apps by simply leveraging the cloud, all at a very low cost. However, worries about the security of the cloud are starting to surface as more and more data about people and businesses is stored there. The security concerns, specifications, and difficulties that cloud service providers (CSP) encounter during cloud engineering are covered in this paper. For the technical and corporate communities, recommended security standards and management approaches are proposed to address them.

A collection of online resources and services is known as cloud computing. Globally dispersed data centers are used to provide cloud services. Through the internet, cloud computing makes virtual resources available to its users. Google applications, which are offered by Google and Microsoft SharePoint, are a general example of cloud services. Severe security problems are also raised by the "cloud computing" industry's explosive expansion. The issue of security has persisted for Open Systems and the internet; cloud computing is particularly vulnerable in this regard. The sole obstacle to cloud computing's widespread acceptance is its lack of security. Numerous security concerns, such as protecting data and monitoring cloud usage by cloud computing providers, surround cloud computing.

The widespread recognition Much as cloud computing has brought forth countless benefits, it has also increased security vulnerabilities. Cloud computing's explosion has presented both service providers and customers with numerous security challenges. How can cloud computing customers know that there are no security or availability problems with their data? Everyone asks: Is their personal data safe? In order to inform suppliers and end users alike about the major security risks connected to cloud computing, this study attempts to pinpoint the most susceptible security risks in the cloud. Through our effort, researchers and security experts will be able to critically analyze the many security models and technologies that have been offered and learn about the issues raised by vendors and users.

Cloud computing is essential to mobile computing because it provides autonomous, ready-to-use mobile resources—like networks, cooperating servers, virtual data storage, tools, and applications—on demand and within a shared pool1. The cloud is a framework for mobility that makes various services and resources available through various cloud models. The cloud computing and data storage scenario is shown in Figure 1. Among the services offered by various clouds are:-

1.1 On-demand Self Service

Automatic services available from the cloud to anyone connected to the cloud.

1.2 Heterogeneous Platforms

Network access can possible to any type of devices having any platform without cloud-to-cloud surfing2.

1.3 Resource Sharing

Every cloud has a pool of resources that are scheduled and accessible to connected users when they travel3. Physical resources in the cloud behave virtually and are not

dependent on location4. These resources are accessible on all clouds without requiring the user to define their locations, allowing them to be transferred from one to another. These resources consist of CPUs, virtual storage, and network bandwidth[5].

1.4 Cloud Collaboration

To offer the linked users uninterrupted services, many clouds are cooperating. Moving from one place to another has no effect on the device's mobile computing.7-8. The cloud services are interoperable and accessible to users who are switching between clouds. Different types of clouds are available to the users such as

- **Private Cloud:** This enterprise cloud provides on-demand internal access to resources and services for users within the designated campus area. Community cloud: This is a backbone network of clouds with restricted private policies that are shared by several enterprises.

- **Public Cloud:** All of the resources and functionalities provided to a connected user are publicly accessible through the public cloud. There are instances where a public organization owns it, and customers have specifically requested cloud services.

- **Hybrid cloud:** Multiple cloud types cooperate to offer services to users who are traveling from one place to another without requiring them to switch between clouds.

II. BIG DATA

Big data is a new technology that is becoming even more significant than cloud computing in comparison. The fields of data storage and virtual smart computing are revolutionized by cloud computing.

Big data is a vast collection of information that may be stored in several places and is made up of several data types with distinct data structures9. A vast amount of storage space is needed because of the daily collection and dissemination of information from faraway regions. It can be challenging to organize, store, analyze, and retrieve the data because of its unusual monolithic structure10–12.The world is currently transitioning to a digital age. All paper documents with bar codes will be converted to digital format.

Traditional paper book libraries are replaced by digital libraries. It is simpler and easier to store huge amounts of material digitally in a paperless format as opposed to hard copy, such what is kept in paper libraries13. However, a significant number of digital manipulating resources are needed due to the ease with which the enormous amount of information may be stored digitally14. Even while it addresses the issue of the storage space needed for paper libraries, it can also offer additional advantages such making digital content easily accessible and manipulable.

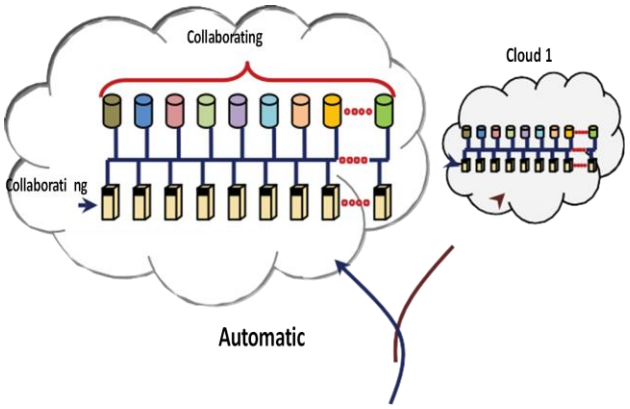

Fig.1. Automatic Cloud Collaboration

2.1 Big Data Security Challenge

Big data is a vast collection of digital information that is kept on several servers and is accessible via the cloud15. Smaller data can be encrypted and decrypted more readily during storage and transport. Larger data sets are more susceptible to security issues, hardware overload, resource management and operating system issues, data collecting, tasks associated with data analysis and processing, indexing, cataloging, searching, data mining, and dissemination16. While data in text format is sometimes easier to handle, its enormous size makes it more difficult to keep and manage in the case of audio, image, and video data. Even with supercomputers or cluster machines, processing data that has been collected by any organization in the few terabytes or petabytes size is more difficult on a single machine17–19. Following Figure 2, which shows the massive data storage and necessary cloud resources, is the security scenario.

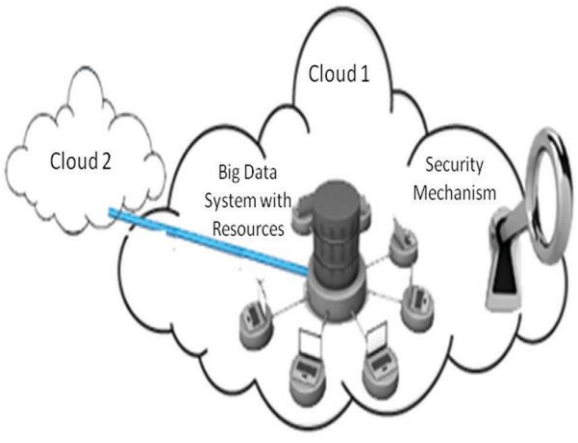

Fig.2. Cloud Data Security Mechanism

2.1.1 Encrypted Storage

All individuals or organizations attempting to store sensitive data on cloud storage. The cloud gathers vast amounts of data from a variety of sources, whether it be plain, private information or secure, encrypted data20. To encrypt the data, either the client computer or the cloud server can be used, or both can be used. It becomes more difficult for a single cloud server to encrypt and store bigger data21.

2.1.2 Encrypted Workload

The quantity of data handled throughout the encryption process is known as the workload. The amount of labor required for encryption increases with the size of the entering data. The encryption process can be managed by the client or server. In the event that the client executes the encryption and frees up the server, the transfer of massive encrypted data22 would negatively impact network traffic. It is preferable to encrypt data at the source, or the client computer, for various security reasons23. It can fix issues with key maintenance, transfer, encryption, and decryption.

2.1.3 Decryption

The client computer can handle the decryption portion if the cloud server stores the encrypted data that the clients have sent. Transferring the encrypted data to the client computer is the only task.

2.1.4 Encrypted Storage

Larger data sets require more processing power and network bandwidth for indexing, sorting, and searching, which makes retrieval time-consuming.

2.1.5 Failure and Recovery

When working with transactions, the Database Management System (DBMS) makes use of the ACID qualities, such as Atomicity, Consistency, Isolation, and Durability. It guarantees the successful completion of the transaction processing. The log-based recovery solution is available in case of failure.

The application of ACID principles in conjunction with the log-based idea can facilitate recovery in big data cloud computing. This study looks into the security problems that big data and cloud computing are now facing and suggests a framework to improve security. The core of big data security can be effectively managed by a collaborative encryption system, wherein the entities involved in the security process can be cloud servers cooperating with client security mechanisms. Instead of being handled by a single framework.

III. EXISTING BIG DATA SECURITY FRAMEWORK

The following are the BigData security challenges.

3.1 Secure Parallel and Distributed Processing

The client's large data set is split up into equal-sized portions for distributed, parallel processing. The security at each computer must be maintained at a separate level while the data is split up for encryption and then collected again24–26.

3.2 Secured Data Storage and Retrieval

However, as data storage grows exponentially across clouds, keeping massive amounts of data becomes increasingly difficult in terms of scalability and availability.

3.3 Source Input Validation

Numerous sources contribute large amounts of data to the storage. How can the data coming from the reliable source be guaranteed? Controlling the data storage from the legitimate data sources is a significant difficulty in this case, because input validation is required.

3.4 Active Monitoring

One of the biggest challenges is real-time, active security monitoring using large data. The volume of data is increasing, necessitating the use of multiple cloud servers to monitor the data flow in real time for storage purposes.

3.5 Privacy Preserving

The more the data, the more cloud servers it needs to be stored on. One important factor in privacy leaks is the possibility of the transaction log being stored on the same servers. Restricting the private storage and retrieval of confidential data via large data cloud storage systems is a serious problem. The privacy is leaked by real-time activity.

3.6 Secure Communication

Data stored by the client in the cloud must be encrypted and securely stored before being sent via an unsecure communication channel.

3.7 Access Control

To ensure the security of the data, access to any other reliable sources cannot be blocked. Malicious users are prevented from accessing the original source by recognizing its authentication. By preserving the metadata of the user and their access, this can be made possible.

IV. PROPOSED SOLUTION TO BIG DATA SECURITY

Larger data means more security challenges. A constraint for big data and cloud computing technologies is weaker security. The big data platform is examined in the section that follows.

4.1 Big Data Platform

Not all data is helpful, and scientists are working to extract the relevant information from the vast amount of data. Big Data requires an infrastructure that can move, store, and integrate massive amounts of data more quickly and accurately than traditional infrastructure can. The process involves using a sophisticated structured database management system to convert unstructured big data into a structured format. The necessary Big Data security platform is shown in Figure 3.

Storage: To store the enormous amounts of data, large storages are needed. Greater storage is needed as the user base grows. Since no single point of storage retains all of the data, parallel collab- rating storages are necessary for large storage.

Communication and Distribution: A considerable amount of data must be distributed quickly and nimbly between servers and from client to Big Data cloud storage.
Structuring unstructured data: More hardware, processing, and networking resources are needed to convert the unstructured incoming data into a structured format.

Unstructured data take up needless storage space whereas structured data is valuable. Metadata management: As the data is bigger, the metadata also becomes bigger. The unstructured data cannot yield the right metadata. To search the useful data across the metadata, the data required be structured.

Management of metadata: When data grows, so does the metadata. The proper metadata cannot be extracted from the unstructured material. The data must be formatted in order to search through the metadata and find relevant information.

4.2 Big Data Processing

Processing such large amounts of data is more expensive. Processing greater data requires more expensive networking, software, hardware, and storage. **Sharing:** It takes time to share data from client to server, server to server, and cloud to cloud. The exchange of data poses a security risk. A primary objective for all organizations, individuals, and the Big Data cloud itself is maintaining the confidentiality of sensitive data. Transition: Converting unstructured data into structured data to ease the process of indexing, sorting, and identifying the data's valuable information.

Retrieval: To obtain the valuable information from the massive data storage, a sophisticated query framework is necessary27. Data that is well-structured and ordered makes searching easier and conserves cloud resources. Retrieval, communication, and sharing speeds decrease with user count.
Views and query processing: Complicated query processing effectively handles massive amounts of data, resulting in valuable client views. The valuable data from the massive amount of data is produced by distributed query processing with colla- bration functions from numerous servers across different clouds and their storages.

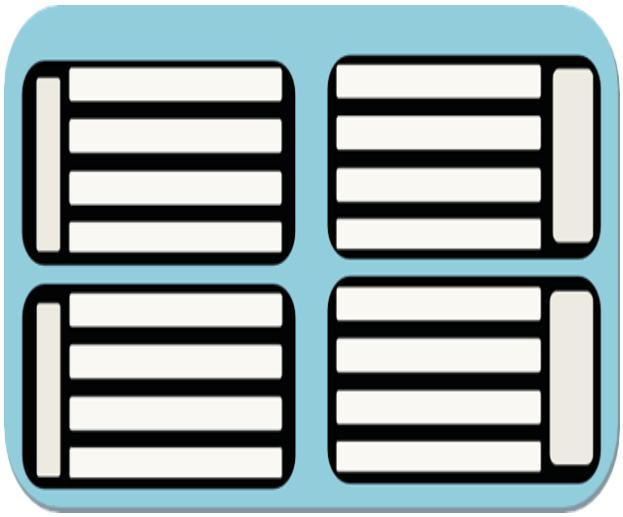

Fig.3. Big Data security platform.

Security:

User/Administrator Authentication: By keeping the metadata for them, registered users, guests, and administrators must authenticate themselves. For every cloud big data, a software or hardware-based front end processor controls the security.
Secure Front End Processors for Encrypted Search: One of the most important components of a Big Data security solution is encrypted search. Front-end processors, whether hardware or software-based, are crucial to preserving cloud and big data security.
Type-dependent security: Various data types call for various security measures. Popular cryptographic techniques like AES are used to secure text data. While several steganographic techniques are used to ensure the security of the audio, pictures, and video.

Failure/Leakage Management: In the event of a failure, log-based recovery must be preserved. Big Data leakage management benefits from the adoption of the Two-Phase-Locking protocol.

4.3 Encryption:

Big Data information security is mostly dependent on encryption. To protect the data, a number of encryption methods are available, including TDES and AES. **Real-time big data encryption**: For tiny local data, offline encryption works well. Nonetheless, real-time encryption is required to secure the vast amounts of online data. Only by encrypting and recording the data's source can this be accomplished. The key management issues are resolved when the client encrypts the data and transfers it via the cloud.

Parallel/Distributed Encryption: This technique effectively manages large amounts of data related to security in the cloud. Together, several servers in several clouds carry out the encryption.

Encryption Key Management: The maintenance of the encryption key is greatly aided by a reliable third party. The key does not need to be maintained on the server if the user encrypts the data at the source. The digital certificate technique is helpful if servers handle the encryption portion.

Homomorphic Encryption: This more recent completely homomorphic encryption method supports the encrypted search by combining several encrypted data segments. In multiparty encryption, homomorphic encryption helps to maintain the encrypted data at the merging location 28, 29. When using homomorphic encryption, where clients from many locations encrypt their own data and store it in a single cloud, collaborative encryption is useful. Without knowing the meaning of the ciphertext, the encrypted query can be searched on that server.

V. CONCLUSION

It takes some time to put the security framework for big data in cloud computing into practice. For complete security, the current cryptosystem is insufficiently useful. The encrypted data saved on cloud big data server storages cannot be searched at all if conventional encryption techniques are used. One amazing method that can assist the encrypted search is the fully homographic encryption technology. The cooperating encryption method covered in this work can be used to implement this. Another main topic of this article is real-time encryption of data moving via cloud networks. Collaboration between cloud-to-cloud servers makes this possible. A few years later, significant technological advancements cause standard encryption methods to no longer meet security requirements. This work presents a distributed, parallel encryption solution that lessens the need for secure storage of sensitive data. Encrypting the various data components at different places and combining them at one place is helpful.

REFERENCES

[1] Ranjan R. Streaming big data processing in datacenter clouds. Proceedings of IEEE Cloud Computing. 2014 May; 1(1):78–6.

[2] Bagheri R, Jahanshahi M. Scheduling workflow applications on the heterogeneous cloud resources. Indian Journal of Science and Technology. 2015 Jun; 8(12):1–8. DOI: 10.17485/ijst/2015/v8i12/57984.

[3] Bagheri R, Jahanshahi M. Scheduling workflow applica- tions on the heterogeneous cloud resources. Indian Journal of Science and Technology. 2015 Jun; 8(12):1–8. DOI: 10.17485/ijst/2015/v8i12/57984.

[4] Zhao F, Li C, Liu CF. A cloud computing security solu- tion based on fully homomorphic encryption. Proceedings of 16th International Conference on Advanced Communication Technology (ICACT); 2014 Feb. p. 485–4.

[5] Jasmine RM, Nishibha GM. Public cloud secure group sharing and accessing in cloud computing. Indian Journal of Science and Technology. 2015 Jul; 8(15):1–7. DOI: 10.17485/ijst/2015/v8i15/75177.

[6] Jeuk S, Szefer J, Zhou S. Towards cloud, service and ten- ant classification for cloud computing. Proceedings of 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid); 2014 May. p. 792–10.

[7] Murthy PK. Top ten challenges in Big Data security and privacy. Proceedings of IEEE International Test Conference (ITC). 2014 Oct. p.1.

[8] Kirubakaramoorthi R, Arivazhagan D, Helen D. Analysis of cloud computing technology. Indian Journal of Science and Technology. 2015 Sep; 8(21):1–3. DOI: 10.17485/ijst/2015/ v8i21/79144.

[9] Pal AS, Pattnaik BP. Classification of virtualization envi- ronment for cloud computing. Indian Journal of Science and Technology. 2013 Jan; 6(1):127–33. DOI: 10.17485/ ijst/2013/v6i1/30572.

[10] Xu L, Jiang C, Wang J, Yuan J, Ren Y. Information security in big data: privacy and data mining. IEEE Access. 2014 Oct; 2:1149–28.

[11] Lee JY. A study on the use of secure data in cloud storage for collaboration. Indian Journal of Science and Technology. 2015 Mar; 8(S5):33–6. DOI: 10.17485/ijst/2015/ v8iS5/61462.

[12] Rajathi A, Saravanan N. A survey on secure storage in cloud computing. Indian Journal of Science and Technology. 2013 Apr; 6(4):4398–401. DOI: 10.17485/ijst/2013/v6i4/31871.

[13] K. Ramesh et al., "Intrusion Determent using Dempster-Shafer Theory in MANET Routing", *(IJCSIT) International Journal of Computer Science and Information Technologies*, vol. 6, no. 1, pp. 37-41, 2015.

[14] Ahmed ST, Loguinov D. On the performance of mapreduce: a stochastic approach. Proceedings of IEEE International Conference on Big Data (Big Data); 2014 Oct. p. 49–54.

[15] M Swamy Das et al., "REVIEW OF CLOUD COMPUTING AND DATA SECURITY", (IJAEMA) The International journal of analytical and experimental modal analysis, vol. 10, no. 3, pp. 123-130, 2018.

[16] Matturdi B, Xianwei Z, Shuai L, Fuhong L. Big data secu- rity and privacy: a review. China Communications. 2014 Supplement; 11(14):135–11.

[17] Kalpana V, Meena V. Study on data storage correctness methods in mobile cloud computing. Indian Journal of Science and Technology. 2015 Mar; 8(6):495–500. DOI: 10.17485/ijst/2015/v8i6/70094.

[18] Marchal S, Jiang X, State R, Engel T. A big data architecture for large scale security monitoring. Proceedings of IEEE International Con-gress on Big Data (BigData Congress); 2014 Jun 27–Jul 2. p. 56–8.

[19] Dong X, Li R, He H, Zhou W, Xue Z, Wu H. Secure sensitive data sharing on a big data platform. Tsinghua Science and Technology. 2015 Feb; 20(1):72–9.

[20] Bosch C, Peter A, Leenders B, Lim HW, Tang Q, Wang H, Hartel P, Jonker W. Distributed searchable symmetric encryption. Proceed-ings of Twelfth Annual International Conference on Privacy, Security and Trust (PST); 2014 Jul.

[21] Tan Z, Nagar UT, He X, Nanda P, Liu RP, Wang S, Hu J. Enhancing big data security with collaborative intrusion detection. IEEE Cloud Computing. 2014 Sep; 1(3):27–7.

[22] Satish, Karuturi S R V, and M Swamy Das. "Multi-Tier Authentication Scheme to Enhance Security in Cloud Computing." IJRAR (International Journal of Research and Analytical Reviews) 6, no. 2 (2019): 1-8, 2019.