IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

A COMPARATIVE STUDY ON DIFFERENT TECHNIQUES USED TO SECURE BIG DATA IN A CLOUD ENVIRONMENT

Renu Yadav (*Research Scholar*), School of Computer Application & Technology,
Career Point University, Kota, Rajasthan

ABSTRACT

Big data is the generated data from different sources sensors, digitizers, scanners, mobile phones, the internet, video, e-mail, and social media. Big data involves appropriate processing capability and analytical capabilities, but conventional methods are unlikely to meet essential expectations. Cloud service providers collaborate on a network that includes computing resources, servers, storage, applications, and a variety of other services. Cloud is a platform for cloud providers to provide services to users and manage data in the data center. With all of these advantages of cloud computing, the protection of huge data remains a serious concern, which also has an adverse influence on cloud migration. Businesses seek robust cyber security methods to protect big data against threats and intrusions while shifting data to the cloud. However, several researchers in the previous study used a variety of security approaches to assure optimum data protection and found that they were effective. To combat security concerns in cloud computing, several security technologies and strategies have been developed and implemented. But still, to enhance data security a security mechanism is always needed to work effectively against unwanted activities. This paper provides a quick comparison of several tested methods for protecting big data in the cloud against security threats during transferring of big data with a comparison that suggested techniques that are recommended to enhance the security of big data.

Keywords: Big data, Cloud computing, Identity, and access management, Encryption and Key Management, Access control, Multifactor Authentication.

1. INTRODUCTION

Sensors, tablet phones, networking sites, as well as a diversity of other internet activities, produce a large quantity of data known as big data. Conventional data solutions strain to organize, preserve, and process big data along with its sophistication. Big data requires the equipment process of maintaining and analyzing using vast information within a short time while somehow guaranteeing cyber security. So instead of implementing a software package on each computer, this technology involves the installation of a single piece of software on each computer that allows users to log in to Web services and host many programs that they require. In this case, another technology cloud computing offers on-demand self-service, rapid elasticity, broad network access, resource pooling, measured services storage, processing, ng, and bandwidth, etc. However, security issues are one of the main big data challenges Organizations store and process large amounts of sensitive and sensitive data in the cloud. All these services made the cloud significant for big data but the security of data is a major restriction to adopting cloud computing. Threats, data breaches, data leakage, denial of service, and other malicious acts that could harm or negatively influence user data are all covered by big data security strategies and methodologies. When confidential data is kept in the cloud, the security of the cloud should be prioritized Privacy and security may be endangered as a consequence of data being stored in several places with multiple legal authorities. To mitigate these dangers, several security mechanisms and authentication procedures have been tested and compared, while data has been outsourced to cloud service providers also required a strong security system is essential to properly defend against attacks on massive data in the cloud[1].

Big Data Overview

Big data is the data produced through different sources sensors, digitizers, scanners, mobile phones, the internet, video, e-mail, and social media. Big data are divided into different categories such as data source, content format, data warehouse, and data processing. Volume, velocity, variety, veracity, and value are all used to represent it. The large amount of data produced every second is represented by volume. The size of a dataset refers to its volume. The speed at which data increases and the exciting area is created is measured by velocity. The flow of data is huge and continuous, with varying rates of streaming data entering and exiting the system. The numerous formats of data, such as text, numbers, videos, and photos, are referred to as variety. Unstructured, semi-structured, and structured data all fall under the category of variety. Veracity concerns the accuracy and reliability of data. It deals with uncertain or imprecise data. It also refers to the noise, biases, and abnormalities in the data. Veracity ensures the quality and accuracy of data. The utility of data is determined by its value, which indicates that some data points, or a combination of values, are more valuable than others. Large amounts of data are meaningless until they provide exact data value, where value pertains to the actual quality of data. It highlights the value of data following analysis.

Cloud Computing and its association with Big Data

Cloud provides services through the Internet, providing convenient and ubiquitous Internet access on-demand is used to group computer resources such as networks, servers, storage, applications, and services. Cloud Computing provides essential services and deployment models i.e. infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS), and public, private community, and hybrid clouds respectively. On the other hand, big data required finding better methods to process and analyze data; it requires infrastructure, more capacity warehouses, and more servers for the rapidly increasing data. In this case, the cloud comes in, and where big data moves to the cloud. Organizations that process large amounts of data will advantage of the cloud, as conventional storage would not be able to understand the needs. Cloud computing can solve all 5V functions, according to big data. Cloud computing features include volume processing with a large pooling of computing resources, data velocity handled by elasticity and on-demand features, variety analysis of data types, resource pooling, self-service, reduced data loyalty, self-service options, improved service, and a Pay-as-you-go pooling model, as well as customer satisfaction. Big data uses cloud computing instead of local storage connected to a computer or internet. Use virtualization technology to develop fast-growing cloud applications to evaluate big data. The cloud not only allows for the processing of large amounts of data but also serves as a service delivery model. Using cloud services, however, eliminates the need for businesses to buy and maintain their computing equipment. There is no need to purchase additional servers, update applications or operating systems, or decommission and discard obsolete hardware or software. Cloud services can provide end-end users with safer and more efficient services since the organization administering and securing these services is likely to have superior abilities and experience than tiny services that can be rented. But because of the cloud's distributed nature, big data organizations are not comfortable outsourcing their data to cloud environments. The security and privacy of their data always have been a concern of the data owners. So, It requires a security approach to prevent threats and challenges from intruders that could strengthen data security in the cloud [2] [3].

2. MOTIVATION

Big data security is a major aspect to be considered in all kinds of information systems, within cloud computing. Due to the exponential growth of data, there is a need to design a faster and strong authentication technique to handle security challenges in the cloud environment. To extract knowledge from an unprecedented number of data, big data necessitates high-performance data processing tools. While cloud computing makes it easier to store, manage, and analyze massive volumes of data, the distributed network it uses creates significant security vulnerabilities. Data security is a major concern in cloud computing, and it encompasses a wide range of features such as confidentiality, integrity, availability, and backup and recovery. Attackers are constantly looking for vulnerabilities to attack the cloud computing environment. Organizations and the users of big data must make sure that the cloud is safe from all the threats and mutual understanding between the users and cloud service providers when it comes to the cloud service. We have seen in most of the previous studies focused only on the discussion of big data opportunities, applications, and issues related to characteristics but the security of data is very less. None of the papers covers specifically big data security in the cloud environment; it only presents the new combination of an existing security topic. The security concerns of big data users and organizations are not properly answered by existing security techniques. Although there are many techniques proposed by researchers that deal with security issues, these works are very limited to addressing the security of data specifically. The research should be more focused on the security of the big data stored in the cloud environment. Cloud service providers need to safeguard the privacy and security of data that they hold on behalf of organizations and users. It is more required to secure the privacy of big data in the cloud. This paper will discuss the different security techniques used while outsourcing big data in the cloud to overcome the challenges of the threat and also compared the techniques that are used in the previous study.

Types of Threats and Challenges

There are many security concerns when outsourcing big data specifically related to the shared, on-demand nature of cloud computing. To identify the top security concerns, CSA surveyed industry experts to compile professional opinions on the security issues in cloud computing. In this latest version of the report, experts have identified the following twelve key cloud security challenges.

- **Breaches of Data:** A data breach is a security alert wherein the personal, individual, or private information data about a corporation or group is retrieved, replicated, or transferred by an unauthorized entity.
- **Weak identity, login, and access management:** Centrally controlled credentials, as well as integrated service, are niceties that present significant threats to cloud protection. When an attacker exploits inadequate authenticity, certifications, and access systems, individuals can access cloud resources and feasibly interpret, alter, or remove confidential material as well as reveal malicious programs into the device.
- **Insecure API:** Weak user interfaces and application programming interfaces expose security flaws in the accessibility, secrecy, and authenticity of cloud infrastructure. Authentication systems themselves could be damaged, or even an API secret revealed unintentionally. Authentication credentials could be used by invaders to obtain access.
- **System and application vulnerabilities:** System and application vulnerabilities are easy to exploit a glitch in systems that intruders have used to penetrate a structure, steal information, and take full control. Vulnerabilities in cloud environments endanger the protection of all systems and services.
- **Account Hijacking:** Whenever a login record is forcibly taken inside a cloud system, it needs to serve as a staging ground for the intruder. The intruder could therefore monitor and manipulate activities, interconnections, and relevant data.

- **Malicious Insiders:** It's a current or former employee, client, or former organization equivalent that has or has allowed network or information access and is now exploiting that access to jeopardize cloud environments' trustworthiness, confidentiality, and integrity.
- **APT (advanced persistent threat):** An APT is a sort of attack where the attacker penetrates networks to gain access to a company's cloud infrastructure and tries to steal personally identifiable information. APTs monitor their victims invisibly for extended periods, frequently changing the security precautions used to confront one another. All such types of attacks get to be growingly difficult to identify as their mitigation strategies achieve better.
- **Data loss:** Data loss is referred to as data damage or uncertainty affected by natural disasters such as hurricanes, as well as simple human discrepancies such as if a cloud supervisor erases details by accident, hard disk malfunction, power failures, or malware attack.
- **Inadequate Due Diligence:** Whenever it arrives to cloud technology, an organization that does not undertake a background investigation and runs to enforce cloud technologies reveals itself to even more marketing, communications technology, legitimate, economic, and compliance risks. Due diligence is the method of reviewing CSPs to make sure that the best standards are maintained. This step comprises evaluating if the cloud provider can provide the required security controls and encounter the consumer's support aspirations.
- **Abuse and Malicious Use of Cloud Services**: One such threat could arise as inefficient cloud security implementations, unregulated cloud storage evaluation, or deceitful compensation for user account service agreement. This reduces the availability, confidentiality, and authenticity of virtual servers for authorized customers. Cloud-based misuse consists mainly of attempting to access Dos attacks or threats, phishing emails, and spoofing initiatives.
- **Denial of Service (DoS):** Denial of Service (DoS) attacks are structured to prevent users from accessing valuable data or applications. A Denial of Service attack affects the availability of a system. A DoS attack has a specific source mechanism where the danger is started, and it is particularly prone to preventative.
- **Vulnerabilities in Shared Technology:** Shared Technology Vulnerabilities: Cloud providers have unlimited resources by sharing infrastructure, platforms, and applications, although this increases security threats. Vulnerabilities in shared technology are hazardous since they can take down the complete cloud system in a moment [4][5].

Tools and Techniques used to Mitigate Security Threats

There are several methods to mitigate each of the threats to big data security in the cloud environment; some of the methods are multi-factor Authentication, encryption, any management system, network layer security, intrusion detection, intrusion prevention system, application programming interface and user interface security, Compliance, Regulations. Security breaches are indeed not unique to cloud computing, but it always surpasses cloud customers' concerns. Cloud providers generally have adequate security for the stuff they're in charge of, but it's up to the end data used to keep their data safe in the cloud. Multifactor authentication and encryption are two key security measures that may assist firms to secure in the cloud and are the finest protection measures to avoid data leaking. Multi-factor authentication is a safe means of obtaining access to sites or services by successfully showing two or more sources of information or credentials. The cryptographic keys in the cryptographic system are managed via encryption and key management. The fundamental between user end and system-level is often referred to it as identity management [6]. The same encryption key is used to encrypt and decrypt data in symmetric encryption (Data-at-Rest). A pair of keys is used for encryption and decryption in cryptographic keys (Data-in-Motion). The two keys are linked together and are generated at the same time. They are called public keys and private keys. Asymmetric keys are also called public keys. Public key infrastructure is a key management system Use hierarchical digital certificates for authentication and public keys for encryption. Network-level security includes handling network protocols and network security, such as distributed nodes, distributed data, and communication between nodes. To ensure the security of packets and ensure that no relevant information can be retrieved even if an unwanted user gains access to network communications, it is required to encrypt entire network connectivity utilizing Transport Layer Security (TSL) and Secure Sockets Layer (SSL). Authentication methods such as Kerberos can also be used to distinguish between authorized and malicious nodes. An intrusion detection and prevention system (IDPS) is a type of software application or device that monitors malicious activity in systems or networks, as well as policy violations. An intrusion Detection System can be used in the virtual machine. An intrusion detection system is a network security management system that collects and analyses data from multiple regions within a network to detect potential security breaches. An Intrusion Detection System is deployed on each host, which monitors the device's transmission and reception of data packets and alerts the administrator if suspicious activity is discovered. Security vulnerabilities are automatically prevented by enforcing security policies, compliances, regulations, and trust. As a result, multi-factor authentication, data security awareness training for employees, access control, an effective incident response strategy, and encryption, as well as good key management, could help to prevent data breaches and other vulnerabilities in the cloud environment[7][8].

Comparative Study of Different Techniques in the Cloud

Author's Name	Research Paper	Techniques	Issues Addressed
J.K. Wang and X. Jia et al.	Data security and authentication in the hybrid cloud computing model.	Authentication interface, single encryption, on CA and PKI model.	The user data authentication interface is protected using a single encryption method and multi-level virtualization technology.
Barsoum and Hasan et al.	Enabling dynamic data and indirect mutual trust for cloud computing storage systems.	Cloud-based storage and mutual trust between data owner and cloud service provider	The technique provides security, integrity, confidentiality, and flexibility.
Lai, R H Deng, C. Guan and J. Weng et al.	Attribute-Based Encryption with Verifiable Outsourced Decryption	Attribute-Based Encryption (ABE) and the decryption method.	A decryption method based on the user-requested features of the outsourced encrypted data.
Sudhansu Ranjan Lenka et al.	Enhancing Data Security in Cloud Computing Using RSA Encryption and MD5 Algorithm	A cloud computing feature that combines RSA encryption and digital signatures within PaaS, SaaS, and IaaS.	It uses the RSA and MD5 algorithms to provide three-way security, authentication, and verification.
Wang et al.	A Trustworthiness evaluation framework in cloud computing for service selection	A framework to rank and measure the trustworthiness of cloud services.	The trustworthiness measuring approach has been used in several inconclusive and inflexible tests.
L. Tawalbeh, N.S. Darwazeh, R.S. Al-Qassas and F.et al.	A secure cloud computing model based on data classification	Based on the type of classified material, encrypt the data using TLS, AES, and SHA security techniques.	It developed the concept of manual, rather than automatic, data classification.
Huo et al.	Fuzzy trust evaluation based on consistency intensity for cloud services	A fuzzy trust evaluation based on consistency intensity for cloud services.	An evaluation model for cloud services that deal with ambiguous data proposes a novel fuzzy assessment approach based on reliability.
Chahal and Singh et al.	Fuzzy rule-based expert system for determining the trustworthiness of cloud service providers	For measuring the trustworthiness of cloud service providers, an expert system based on the fuzzy rule was developed.	It provided security, dependability, and scalability, but it was unreliable, confidential, and unsafe.

Wu et al.	An effective approach for the protection of privacy text data in the Cloud DB	Encryption/decryption Techniques.	In this technique, the user data must first be decrypted, and then the data must be encrypted again after the query procedure is completed.
Chatterjee et. al.	An approach towards achieving transparency in user authentication	Biometrics-based re-authentication system.	This solution outperformed the typical password-based authentication mechanism in terms of security.

Table.1 Comparison of different Techniques used to secure Big data

3. RESULT IN ANALYSIS AND DISCUSSION

From the comparative analysis, we have identified various big data security challenges in the cloud during outsourced data. Phishing assaults, Spoofing attacks, Denial of Service, and intrusion detection are among the vulnerabilities that have been highlighted. Internal and external cyberattacks, such as lack of trust, inadequate SLA, data integrity, and information security, are the other difficulties. Some challenges, like secrecy, consistency, and reliability, are related to cloud properties. We also examine security methods as well as provide thorough explanations of them, such as identity-based access controls. The monitoring system, intrusion prevention, multi-user identity management model, and service-level agreements are some of the terms to describe the management system. Cyberattacks, data loss, and data breaches are all on the rise and a variety of dangers are endangering safety and confidentiality. These mitigation techniques have a strong impact on the performance, security, efficiency, and access control of Cloud Computing. Some defense technologies have generally improved cloud services and used various technologies to protect big data. As a result, a thorough and effective strategy to safeguard big data from security breaches is necessary. The organization and individuals who would employ the cloud service examined each technique to become aware and strengthen data security. To protect a large amount of data in the cloud, there are many methods for data authentication, and there are many. Related studies are associated with secure data within the cloud but the foremost significant techniques among those are access control and encryption[9]. We noticed that the paper has taken only a fast review of the safest approaches to solving data security issues together with the fundamental properties of huge data. The results signify that the majority of researchers show their interest in encryption technology to boost the safety of information in an exceeding cloud computing environment. The goal of this research will be to boost the effectiveness of information security mechanisms, as well as to identify proposed security facilities, third-party compliance, and efficient authentication methodologies[9].

4. FUTURE SCOPE OF RESEARCH

The most essential element is to provide a security tool or techniques capable of handling information security ideas including secrecy, integrity, and accessibility. When it comes to the protection of information in the cloud, cloud providers and consumers must guarantee that such a cloud is safe from all threats and vulnerabilities. Big data is a fast-increasing trend; big data management generally requires a comprehensive understanding of big data items as they pertain to security. It has been found that the best cyber security strategies are those that would focus on architecture, data confidentiality, information management, and authenticity. The expansion and functionality of the latest advancement in information and communication techniques, like social media and the Internet of Things, and also cloud computing, are extremely crucial to big data protection and privacy. However, given the current constraints in information security, suggestions for future research require particularly addressing the incredibly huge commitment of research scholars and practitioners. There is a scarcity of sophisticated functionality to security and privacy concerns. The integration of various remedies from multiple categories of conventional and emerging current technologies overcomes safety issues led by something like a solitary security mechanism. Cloud resources are widely used by big data companies, and their security can be a top priority. This paper describes the strength of the research technique and identifies the areas which have to be improvised. There is still a requirement to handle these voluminous data security challenges concerning cloud computing and are always available with the most effective and suitable approach to secure infrastructure. Big data security necessitates extensive research to fully detect risks and adequately solve security concerns.

5. CONCLUSION

We will discuss various methods for protecting huge data in a cloud computing context in this paper. This comprises the methods used in numerous research publications, as well as their benefits and drawbacks. All of those studies lacked data security methods for a variety of reasons, including a lack of support for dynamic data operations, a lack of data integrity, and a lack of high resource and compute costs. This paper examined some critical big data security issues, classified various existing solutions, and compared their potential. We identified major security concerns and proposed some specific techniques for preventing threats. When outsourcing cloud computing, big data security issues must be addressed at all levels of the cloud environment using essential techniques and tools. These techniques exhibit better results in terms of access control and multi-factor authentication for the scattered environment as compared to other mechanisms. More research is required to tackle the protection of huge data rather than current security technologies and methods. In the future, prominent security challenges and issues must be addressed by academicians and researchers to make sure the long-term success of data security in an exceeding cloud computing environment and to collectively explore new security techniques to handle the challenges of huge data. Hence this paper gives an overall description of all existing techniques for big data security and methods proposed for ensuring security in the cloud.

6. REFERENCES

[1] Hashem, I.A.T., et al., 2014. The rise of "big data" on cloud computing: Review and open research issues.Information Systems, 47, pp.98–115.

[2] JM. Mart´ınez Sesmero, "Big Data"; application and utility for the healthcare system,” *FarmHosp*, vol. 39, no. 2, pp.69-70, 2015.

[3] R. Saranya, V.P.MuthuKumar, “Security issues associated with Big Data in cloud computing”, International Journal of Multidisciplinary Research and Development, Volume 2, Issue 4, 580-585, April 2015.

[4] Chaturvedi, A., & Lone, F. A. (2017). Analysis of Big Data Security Schemes for Detection and Prevention from Intruder Attacks in Cloud Computing. International Journal of Computer Applications, 158(5)

[5] A.F. Barsoum, A. Hasan, "Enabling dynamic data and indirect mutual trust for cloud computing storage systems", IEEE Trans. Parallel Distrib. Syst., 24 (2013), pp. 2375-2385

[6] J. Lai, R H Deng, C. Guan, and J. Weng. "Attribute-Based Encryption with Verifiable Outsourced Decryption." IEEE Trans. Inf. Forms. Security, vol 8, pp 1343-1354, 2013.

[7] Sudhansu Ranjan Lenka and Biswaranjan Nayak, "Enhancing Data Security in Cloud Computing Using RSA Encryption and MD5 Algorithm” International Journal of Computer Science Trends and Technology (IJCST) – Volume 2 Issue 3, June 2014.

[8] Wu, Z.; Xu, G.; Lu, C.; Chen, E.; Jiang, F.; Li, G. “An effective approach for the protection of privacy text data in the Cloud DB”. World Wide Web 2018, 21, 915–938. [CrossRef].

[9] Chatterjee, K. Biometric re-authentication: “An approach towards achieving transparency in user authentication”. Multimed. Tools Appl. 2019, 78, 6679–6700.