Journal of IoT in Social, Mobile, Analytics, and Cloud (ISSN: 2582-1369)
www.irojournals.com/iroismac/

Review on Advanced Cost Effective Approach for Privacy with Dataset in Cloud Storage

Vijesh Joe

Data Scientist, WoTo Technologies, Chennai

E-mail: vijesh.joe@gmail.com

Abstract

Cloud computing allows customers to run compute and data-intensive applications without the need for a large investment in infrastructure. Additionally, a significant amount of intermediate datasets are created and often saved, in order to reduce the expense of re-computing these applications. It becomes difficult to protect the privacy of intermediate datasets because attackers may be able to retrieve information that is sensitive to privacy via the analysis of several intermediate datasets. Existing techniques to deal with this problem generally endorse the use of encryption for all cloud datasets. For data-intensive applications, the time and expense of repeatedly decrypting and encrypting intermediate datasets are prohibitive; hence, encrypting all intermediate datasets does not make sense. Big heterogeneous data storage concerns and challenges, countermeasures (security and administration) and cloud storage prospects, are discussed in this article. New questions arise for cloud storage researchers, when they examine these issues in depth.

Keywords: Privacy in cloud, cloud storage, authentication, data security, data protection, homomorphic encryption

1. Introduction

From the collection of hardware, networks, storage, services, and interfaces that may be gathered together to supply components of computing as a service, cloud computing is born out of the concept. Software, infrastructure, and hold over the Internet are the major components of this service, which may be offered separately or as a comprehensive platform. You don't need to buy any extra hardware to store a huge amount of data and run programs on the cloud. Data storage and computing are also more flexible with this technology. It is

DOI: https://doi.org/10.36548/jismac.2022.2.001
Received: 16.05.2022, received in revised form: 18.06.2022, accepted: 03.07.2022, published: 13.07.2022

possible to process and create enormous amounts of processed data from these types of applications. In addition, certain useful intermediate datasets may be stored in order to prevent the high expense of putting together the final product. To use cloud computing, you pay a monthly fee for accessing the network storage and for computational power. In order to use the Cloud, you must have an Internet connection, which might be cumbersome [1-5].

When it comes to data storage, processing, and handling, a "cloud" is a group of servers scattered throughout the Internet. The Internet is used to provide hardware and software as a service in the cloud. In this way, big data may manage and disseminate its stored data effectively [6]. Big data in the cloud is protected by Hadoop, file and network environments, encryption, logging, and nods authentication, as well as a multi-tiered architecture for cloud assurance. The mining of large data employs a variety of models. There are a number of advantages to cloud computing, including the ability to handle large volumes of data. Connectivity between devices and data is facilitated by sharing data and connecting to other devices, which in turn enhances the interconnectivity. In essence, big data are utilized to turn raw data into usable information. It is the major goal of big data to store, organize, display, and analyze enormous volumes of data every day [7-10]. Figure 1 shows the data privacy recent advancement consideration parameter graph.

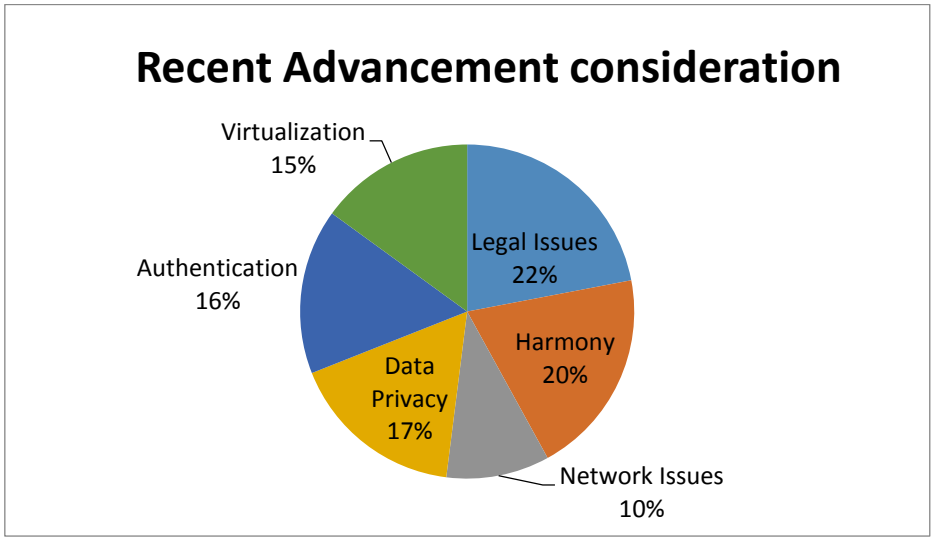

Figure 1. Data privacy recent advancement

Many embedded devices now link to the Internet, allowing for a greater amount of data to be created. These gadgets monitor and connect everything (such as roads, buildings, the environment, and lakes). Traditional data management approaches can't keep up with the ever-increasing volume of data on the Internet [11].

Organizations' computer needs have expanded tremendously since the introduction of technology, necessitating additional processing and storage capacity. As a result of the time and money required to build up large-scale systems, business clients are increasingly likely to outsource their computing and storage resources. Researching storage in the cloud is becoming more common since new applications need a lot of storage space and the quantity of data consumed is growing every year.

1.1 Cloud computing

Recent years have seen the rise of cloud computing. Users may expect a wide range of services at a fair and steadily dropping cost from it, such as infrastructure, platforms, or software. Cloud computing, on the other hand, is still in its infancy [12]. In addition, cloud computing's expansion is hampered by a lack of standards, security concerns, and interoperability challenges [2, 3]. Because of this, businesses choose cloud providers depending on how well their services perform. Aside from this issue, it is difficult to gauge how cloud service providers manage security quality since many don't provide clients access into their infrastructures.

As a result, cloud service providers should take security into consideration. Cloud computing data centers are designed to fail, according to the "design-for-failure" paradigm. Scalable, lower-cost, and purpose-built cloud storage solutions are required for worldwide provisioning of storage services in the public cloud. Various gears including servers, networking equipment, and storage systems, may be used in these solutions. In addition, it should be able to take advantage of significant economies of scale by using conventional delivery patterns. Many "off-the-shelf" IT solutions developed for use in conventional IT markets are too costly and do not fit the unique needs of a cloud data center environment. This research examines the architecture of cloud storage, as well as the problems it faces and the potential solutions it offers, considering the future of cloud storage, as well as its potential benefits and drawbacks. There are several issues with cloud storage, including not limited to security and privacy, data dynamics and integrity, data access, data separation and authentication, data breaches, backup issues, and virtualization vulnerabilities.

1.2 Confidentiality Issues

Many clients' data are kept on a single server in the cloud, making privacy an important issue. The cloud's essential criterion for secrecy is the promise of preserving sensitive or secret information that is stored or processed in the cloud. Data that are externally

stored, the identities of users who have access to the data, or the activities that users conduct on the data might all be considered, depending on the circumstances. Confidentiality is ensured in these systems by the use of encryption methods. The internet and servers are used to maintain and administer data and applications in the cloud. Thus, cloud computing has made it possible to increase processing power without requiring large expenditures.

2. Literature Survey

2.1 Data Privacy Issues

Tenants have the right to expect certain protections for their personal information. Cloud storage providers that offer high levels of security are more likely to attract customers. Companies that provide cloud storage services are looking for ways to better secure and regulate access to their customers' data. The number of data assaults and interceptions is rising in tandem with the amount of the data being sent. The data is stored in a vitalized environment where the user has no control over the data. Cloud computing security, privacy, and authentication have been the subject of several studies.

Although the cloud includes fragmented data, the pieces are always disseminated and replicated on physical storage devices. In addition, a restitution tool is built into the cloud to help in getting the data back. To get this result, the specified fragmentation's granularity must be high [13].

Petit et al., have made a significant contribution to the protection of Web users' privacy. This project's goal is to make it possible for a user to do a search on a search engine while still maintaining their privacy. An applicant's identity cannot be deduced by a search engine or any other adversary listening in on the network (the user). Fake applications (fake inquiries) are the goal of the writers, who seek to keep opponents from finding their applications (or engine research).

The addition of irrelevant responses to obfuscate the search request in order to protect the apps is a big drawback of this technique (interference). The findings are less accurate and the network is excessively busy with this solution [14].

Using a unique upper limit privacy leakage constraint-based technique, Zhang et al., proposed to determine the encrypted process with dataset, in order to save money while still

meeting data owners' privacy demands. For managing intermediate datasets, data provenance was used [15].

Threshold Filtering was used in Praveena et al.'s, study to categorise the dataset that is being encrypted. There will be a predetermined value for each intermediate dataset depending on the data's privacy information. The intermediate dataset will be encrypted and anonymized if its value exceeds a certain threshold [16]. To make finding and accessing the encrypted dataset simple, Two-Round Searchable Encryption (TRSE) was used. It claims to first anonymize and then encrypt any datasets before storing them in the cloud or sharing them with others.

For current data-intensive applications, it was estimated that huge computing power and storage capacity can be delivered. It is possible to access or analyze original datasets more regularly in this context. As a result, the original datasets were transformed into intermediate ones. Encryption and anonymization were combined in this method to save money [17].

Sedic was developed by X. Zhou et al. Map-reduce processes were made more secure by using this technique. A particular MapReduce feature allows Sedic to split a computing operation based on the data's sensitivity level. As a response to concerns about data privacy, it divides a project into two parts: one for private cloud computing and the other for public commercial cloud computing [18].

2.2 Motivation

Organizations with a new business model may take use of cloud computing, which is a rapidly expanding technology that gives them free access to enormous quantities of data. However, for the time being, most businesses are hesitant to expand their operations into the cloud due to concerns about security. When data is kept on a cloud server, there are ramifications in terms of data ownership and management isolation. This means that the Cloud Service Provider (CSP) has unrestricted access to the user's data on the cloud server. In the meanwhile, an attacker attempting to get access to the cloud server's data may conduct an attack. The cloud environment also has a slew of security vulnerabilities, including man-in-the-middle attacks and data breaches. All of the aforementioned issues pose a significant risk to the privacy of the user. Loss of data by users is impeded, and data leakage issues may arise as a result.

In addition, if the user uploads data directly to the cloud, there may be concerns such as high bandwidth needs, excessive latency, and a big amount of data. These are the main roadblocks to cloud adoption. As a result, identifying a solution and working on it is essential.

3. Virtualization Issues and Vulnerability

Cloud computing relies heavily on virtualization to isolate different instances operating on a single computer from one another. Unchecked, it poses a serious threat to the security of a cloud-based infrastructure. Administrative control over operating systems, both as guest and host, and their inability to provide isolation or scalability difficulties are the second problem. VMMs currently on the market include flaws that enable VMs to be escaped from. Since virtualized guest systems might interfere with the host operating system, "root security" is the need in these situations [19-22]. There have been reports of vulnerabilities in virtualization software that might enable unauthorized access by a local user or an attacker. Examples of such vulnerabilities include the Microsoft Virtual PC and Server flaws, which enable guest OS users to run, code on other guest even the host OS itself. In doing so, it might lead to individuals having greater access to sensitive information, which could result in their access being misused by others.

3.1 Handling of dataset

It is assumed that an intermediary data collection has been anonymized in order to meet specified privacy standards. Combining various datasets, on the other hand, may provide a significant challenge in terms of disclosing information that should be kept private and may lead to privacy violations.

3.2 Privacy upper limit constraint

The privacy upper limit constraint specifies how to quantify privacy for a particular data collection. The issue is identifying a consistent upper-bound for privacy leaks from diverse data sources. As a result, the cost of preserving privacy may be reduced by limiting the number of intermediate datasets that must be encrypted [23-25]. Costs of privacy protection must be included into the equation. There must be less limit to what can be done than the threshold. Figure 2 contains the block diagram of advanced cost effective approach for data privacy.

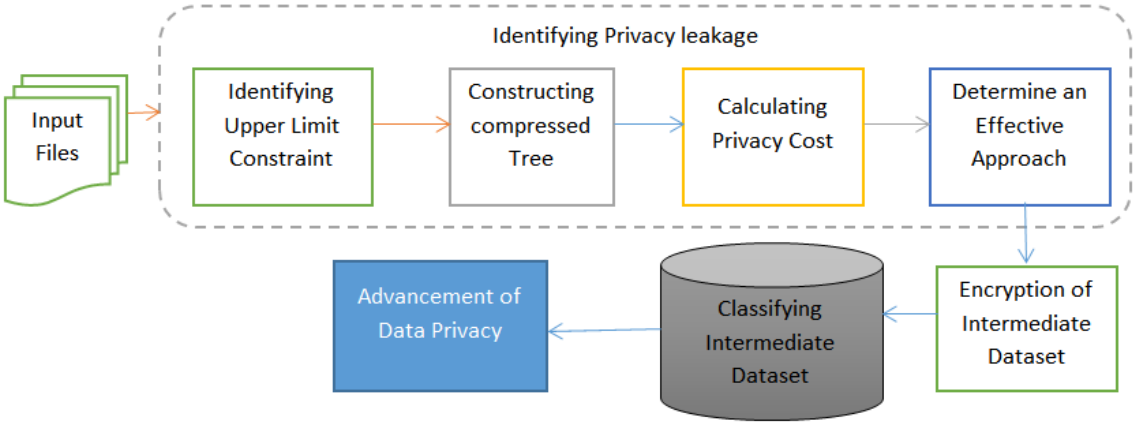

Figure 2. Advanced Cost Effective Approach for Data Privacy

3.3 Homomorphic Encryption

Ronald Rivest and his colleagues initially proposed the idea of homomorphic encryption. In the 30 years that have elapsed since, nothing has changed. In 1982, Shafi Goldwasser and Silvio Micali created a proven security encryption technology that attained a remarkable degree of safety. However, it could only encrypt a single bit using additive homomorphic encryption. Pascal Paillier also presented an additive homomorphic encryption scheme as part of the same theory. Later, Dan Boneh and his colleagues came up with a means to prove that encrypted data is safe. Additions are unlimited, whereas multiplications are limited to one per system.

Homomorphic encryption is a kind of encryption in which the operations are performed on the encrypted data rather than the original data, resulting in the same outcome as if the operations had been performed on the original data. The cipher text may be subjected to complex mathematical operations without affecting the integrity of the encryption. With homomorphic encryption, services may be chained together without revealing any data [26].

3.4 Authorization and Authentication

Authentication is essential in any system that requires perfect security, like a cloud's front door, which only lets in those who are known and trusted. Authentication is necessary to gain access to crucial information. As a result, the authentication mechanism must be rigorous in order to verify that only legitimate users may access the service. Data confidentiality and integrity may be ensured via encryption by limiting access to only those

who have been verified. A comprehensive authentication system may alleviate the majority of the security problems.

4. Future Directions

Cloud storage technology, despite its simplicity of use and cost advantages, nevertheless has a number of drawbacks. In the cloud storage architecture, security concerns (such as confidentiality, integrity, access, authentication and data breaches) and data management difficulties are mostly obscured by the cloud (e.g., dynamics, data segregation, backup, and virtualization). A variety of strategies have been offered in the literature to tackle these dangers. When it comes to data security, reliable timestamps and digital certificates are often used together. Cryptographic solutions are employed to secure data secrecy, whereas attribute-based encryption is utilized to protect data access. Authentication and permission are used to control access. Using a global transaction manager, many databases may be reliably managed without compromising data integrity. In the section devoted to shared data, researchers in cloud computing models plan to address the issue of group data sharing in the future.

5. Conclusion

This research article presents in the perceptive to determine encrypted dataset through privacy policy with cost effectiveness to match the privacy needs of the data owner. To ensure the validity of the approaches, it gives comprehensive access to a secure cloud and efficient use of resources. The data leakage and data disposal may be prevented by using security measures. Crumbling privacy leakage limitations then break down this issue into smaller sub problems. A workable heuristic technique to select datasets that should be encrypted is what has been recommended in the end.

References

[1] Maximilian Wöhrer, Uwe Zdun, " Smart contracts: Security patterns in the ethereum ecosystem and solidity", International Workshop on Blockchain Oriented Software Engineering (IWBOSE) ,IEEE, 2018.

[2] Qiwu Zou, Yuzhe Tang, Ju Chen, Kai Li, Charles A. Kamhoua, Kevin Kwiat, Laurent Njilla, "ChainFS: Blockchain-Secured Cloud Storage", IEEE 11th International Conference on Cloud Computing (CLOUD), 2018.

[3] Deepika Saxena and Ashutosh Kumar Singh. workload forecasting and resource management models based on machine learning for cloud computing environments. arXiv preprint arXiv:2106.15112, 2021.

[4] Jitendra Kumar, Ashutosh Kumar Singh, and Rajkumar Buyya. Ensemble learning based predictive framework for virtual machine resource request prediction. Neurocomputing, 397:20–30, 2020.

[5] Sakshi Chhabra and Ashutosh Kumar Singh. Dynamic data leakage detection model based approach for map reduces computational security in cloud. In 2016 Fifth International Conference on Eco-friendly Computing and Communication Systems (ICECCS), pages 13–19. IEEE, 2016.

[6] AK Singh and Jitendra Kumar. Secure and energy aware load balancing framework for cloud data centre networks. Electronics Letters, 55(9):540–541, 2019.

[7] Aman Singh Chauhan, Dikshika Rani, Akash Kumar, Rishabh Gupta, and Ashutosh Kumar Singh. A survey on privacy-preserving outsourced data on cloud with multiple data providers. In Proceedings of the International Conference on Innovative Computing & Communications (ICICC), 2020.

[8] Ehsan Hesamifard, Hassan Takabi, Mehdi Ghasemi, and Rebecca N Wright. Privacy-preserving machine learning as a service. Proc. Priv. Enhancing Technol., 2018(3):123–142, 2018.

[9] Ping Li, Jin Li, Zhengan Huang, Chong-Zhi Gao, Wen-Bin Chen, and Kai Chen. Privacy-preserving outsourced classification in cloud computing. Cluster Computing, 21(1):277–286, 2018.

[10] Xindi Ma, Jianfeng Ma, Hui Li, Qi Jiang, and Sheng Gao. Pdlm: Privacy-preserving deep learning model on cloud with multiple keys. IEEE Transactions on Services Computing, 2018.

[11] B. Mao, S. Wu, and H. Jiang, “Exploiting workload characteristics and service diversity to improve the availability of cloud storage systems,” IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 7, pp. 2010–2021, July 2016.

[12] M. H. Chen, Y. C. Tung, S. H. Hung, K. C. J. Lin, and C. F. Chou, “Availability is not enough: Minimizing joint response time in peer-assisted cloud storage systems,” IEEE Systems Journal, vol. PP, no. 99, pp. 1–11, 2016.

[13] Cigref, “cloud computing and protection of data”, in publication of network companies enterprises 2015.

[14] A. Petit , Ben Mokhtar, L. Brunie and H. Kosch, “Towards Efficient and Accurate Privacy Preserving Web Search,” MW4NG ’14 December 8-12, 2014, Bordeaux, France. 2014.

[15] X. Zhang, C. Liu, S. pandey and J. Chen, “A Privacy Leakage Upper Bound Constraint Based Approach for Cost-Effective Privacy preserving of Intermediate Data Sets in Cloud,” IEEE, Vol 24, No. 6, June 2013.

[16] T. Praveena, G Raja, “Threshold Based Filtering approach For Cost Effective over Encrypted Cloud Data ", IJISET, Vol. 1 Issue 3, May 2014.

[17] Ms. C. Celcia, Mrs. T. Kavitha, “Privacy Preserving Heuristic Approach fpr Intermediate Data Sets in Cloud,” IJETT, Vol 9, March 2014 ISSN 2231-5381.

[18] K. Zhang, X. Zhou, Y. Chen, X. Wang, Y. Ruan, “Sedic: Privacy Aware Data intensive Computing on Hybrid Clouds,” Proc. 18th ACM Conf. Computer and Comm. Security (CCS”11), pp 515-526, 2011.

[19] Y. Hua, B. Xiao, X. Liu, and D. Feng, “The design and implementations of locality-aware approximate queries in hybrid storage systems,” IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 11, pp. 3194–3207, Nov 2015.

[20] C. Lalit, I. Majitar, and B. Rabindranath, “A comprehensive survey on internet of things (iot) toward 5g wireless systems,” IEEE Internet of Things Journal, vol. 7, no. 1, pp. 16 – 32, 2020.

[21] P. Samuela, C. Claudia, B. Marina, T. Marco, P. Valeria, G. Andrea, and F. Manuel, “Iot enabling technologies for extreme connectivity smart grid applications,” in CTTE–FITCE: Smart Cities & Information and Communication Technology (CTTE–FITCE). IEEE, 2020, pp. 16–32.

[22] X. Shuming, N. Qiang,W. Liangmin, andW. Qian, “Sem-acsit: Secure and efficient multi-authority access control for iot cloud storage,” IEEE Internet of Things Journal, pp. 1 – 1, 2020.

[23] G. Tudor, C.-C. Andrei, A.-C. Madalina, and Z. Alexandru, “Cloud storage a comparison between centralized solutions versus decentralized cloud storage solutions using blockchain technology,” in 54th International Universities Power Engineering Conference (UPEC). IEEE, 2019, pp. 16 – 32.

[24] B. Mao, S. Wu, and H. Jiang, “Exploiting workload characteristics and service diversity to improve the availability of cloud storage systems,” IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 7, pp. 2010–2021, July 2016.

[25] M. H. Chen, Y. C. Tung, S. H. Hung, K. C. J. Lin, and C. F. Chou, "Availability is not enough: Minimizing joint response time in peer-assisted cloud storage systems," IEEE Systems Journal, vol. PP, no. 99, pp. 1–11, 2016.

[26] D. Pritam and M. Chatterjee, "Enforcing role-based access control for secure data storage in cloud using authentication and encryption techniques," Journal of Network Communications and Emerging Technologies (JNCET), vol. 6, no. 4, 2016.

Author's biography

C. Vijesh Joe is presently working as Data Scientist at WoTo Technologies in Chennai. His major area of research includes cloud computing, speech processing, wireless networks security, data science analytics, and computer graphics in multimedia.