IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12364

A Study on Reducing Information Leakage in Multi-Cloud Storage Services

Saana Venkata Krishna Reddy[1], Dr. Gobi Natesan[2]

Student, MCA, Jain (Deemed-to-be-University), Bengaluru, India[1]

Professor, MCA, Jain (Deemed-to-be-University), Bengaluru, India[2]

Abstract: In recent times, there have been significant advancements in techniques for storing data across multiple clouds. This has resulted in users having a certain level of control over information leakage, as data is spread out across several cloud storage providers. However, the chaotic distribution of data pieces can still result in severe data loss, even with the use of multiple clouds. This article highlights the issue of a data leakage caused by the chaotic distribution of data in multi-cloud storage systems and introduces a solution called Store Simulation. This system aims to minimize customer data leakage across multiple clouds by keeping syntactically related data on the same cloud.

The efficient creation of similarity-preserving functions is a key aspect of Store Sim. These functions use signatures, such as Min Hash and Bloom filters, to compute information leakage. Additionally, the system utilizes a clustering-based storage plan creation method to efficiently distribute data chunks across multiple clouds.

To test the efficacy of this strategy, two real datasets from GitHub and Wikipedia were used. The results show that the Store Sim blueprint can reduce information leakage by up to 60% compared to no-plan placement. Furthermore, our investigation into system vulnerabilities reveals that this approach makes information attacks more challenging.

Keywords Multi-cloud storage systems, Information leakage, Data distribution, Syntactic similarity Min Hash, Bloom filters, Clustering-based storage plan creation, Real datasets, Information attacks Data loss prevention.

I. INTRODUCTION

As the digital world expands, people are increasingly relying on devices such as laptops, phones, and tablets to manage their digital lives. This has resulted in a demand for network resources, which has led to the popularity of cloud-based file-sharing and storage services such as Dropbox, Google Drive, and Amazon S3. These services provide user-friendly interfaces and affordable storage pricing. However, the use of such cloud storage solutions has also led to concerns over data privacy and security. Storage companies may use user data for analytics and advertising, while malicious insiders or external threats may exploit backdoors or engage in bribery or coercion to gain access to sensitive information.

To avoid these risks, users are often advised not to place all their data in a single cloud storage solution. However, the reality is more complex. Cloud storage providers use rsync-like protocols to synchronize local files with remote data stored in their centralized clouds. Local files are divided into smaller segments, and fingerprints are taken and hashed for efficient management.

Despite these measures, the risk of data breaches remains a concern. In recent privacy violations, third-party attackers have attempted to access data stored across multiple cloud storage providers. Therefore, users need to remain vigilant and choose their cloud storage providers carefully to ensure the privacy and security of their data.

II. LITERATURE REVIEW

The need for storage solutions expands along with the daily volume of data being created. Despite the nearly limitless storage capacity offered by cloud storage providers, data owners prefer variety in the location of their data to avoid vendor lock-in, improve availability, and boost durability. Furthermore, a certain cloud provider can be more economical than another depending on the manner in which customers access their data. In order to overcome these difficulties, Bonvin and Aberer suggest Scalia, an online cloud brokerage solution that continually modifies the placement of data according on its use pattern and is subject to efficiency goals like storage costs. Just those items that have been demonstrated to be more cost-effective than static locations and are close to the appropriate location for data are those that Scalia considers moving.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 3, March 2023
DOI: 10.17148/IJARCCE.2023.12364

Rehaman and Aberer suggest a distributed environment model that entails collaboration between a number of small data centres (SDCs) to boost performance. Due to growing worries about security and information control, SDCs are posing a threat to the centralized system used by the majority of cloud providers. Nevertheless, resource inelasticities can affect SDCs, which can lead to a decline in productivity and income. The authors create a general strategy function for SDCs to measure cooperation success across several dimensions of resource sharing.

Rehaman and Zhang suggest Store Sim, a data leak storage system, to solve the problem of high disclosures in multi-cloud storage systems. By storing data that is syntactically identical on the same cloud, Store Sim seeks to reduce information leakage. The authors provide an approximation of a Min Hash and Bloom filter-based approach to create resemblance signatures for data chunks, along with a function to calculate data leaks based on these signatures. Moreover, they provide a powerful clustering-based storage plan creation technique for efficiently spreading data chunks among many clouds with the least amount of data loss.

III. RELATED WORKS

In the field of cloud storage, there has been a growing interest in developing solutions that can provide efficient and cost-effective data placement while ensuring data availability and durability. One such solution is Scalia, proposed by Bonvin and Aberer, which is a cloud storage brokerage system that adapts the placement of data based on its access pattern and is subject to optimization objectives such as storage costs. Scalia efficiently considers repositioning selected objects that may significantly lower the storage cost, making it cost-effective against static placements and ideal data placement in various scenarios of data access patterns, available cloud storage solutions, and of failures.

Rehaman and Aberer proposed a decentralized cloud model that addresses the issue of resource in-elasticity in small data centres (SDCs) and aims to improve performance by allowing SDCs to cooperate. The proposed model employs a general strategy function for the SDCs to evaluate the performance of cooperation based on different dimensions of resource sharing. This approach can be beneficial for meeting local demands, improving resource utilization, and ensuring service continuity in the case of resource failure.

In summary, the two works presented provide insights into the development of cloud storage solutions that address issues of cost, efficiency, resource utilization, and service continuity. Scalia provides an adaptive scheme for efficient multi-cloud storage, while Rehaman and Aberer's work proposes a decentralized cloud model that enables small data centres to cooperate to improve performance.

IV. METHODOLOGY

The method outlined in Store-Sim contains a rough algorithm called BFSMinHash that creates resemblance signatures for data chunks using Minhash and Bloom filter. Information leakage is also measured using a bilateral leakage formula based on Jaccard similarity. The team created SPClustering, an effective storage plan generation technique that scatters user data to many clouds while reducing information leakage, using the data leakage evaluated by BFSMinHash.

Two datasets that were crawled from GitHub and Wikipedia were utilised to assess their framework. These datasets included files with several modifications, enabling a complete assessment of the suggested method. The trials carried have demonstrated that StoreSim significantly decreased information leakage across various clouds while while retaining high efficiency.

The team also looked at how easily the system might be attacked and discovered that StoreSim made information assaults more trickier. Having strong safety protocols in place is critical for safeguarding user data given the ubiquity of keyloggers as well as other harmful spyware. By decreasing data leakage and making assaults on information considerably more difficult to carry out, StoreSim offers a possible answer to this issue

V. CONCLUSION

This project deals with the problem of information leaking in multi-cloud storage. Although storing data across several clouds offers some safety, uncontrolled dispersion can still lead to significant amounts of information loss. StoreSim, an unique technique that employs BFSMinHash and SPClustering to store data on the same cloud with little information loss, was offered as a solution to this problem. It was shown that StoreSim is effective and efficient using two actual datasets, and that when compared to unplanned placement, it can cut down on information leakage by up to 60%. Also, the attack ability investigation showed that StoreSim makes assaults on retail information far more difficult in addition to lowering the danger of wholesale information leaking.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 3, March 2023
DOI: 10.17148/IJARCCE.2023.12364

REFERENCES

[1] J. Crowcroft, "On the duality of resilience and privacy," in Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 471, no. 2175. The Royal Society, 2015, p. 20140862.
[2] A. Bessani, M. Correia, B. Quaresma, F. Andr´e, and P. Sousa, "Depsky:dependable and secure storage in a cloud-of-clouds," ACM Transactions on Storage (TOS), vol. 9, no. 4, p. 12, 2013.
[3] H. Chen, Y. Hu, P. Lee, and Y. Tang, "Nccloud: A network-coding-based storage system in a cloud-of-clouds," 2013.
[4] T. G. Papaioannou, N. Bonvin, and K. Aberer, "Scalia: an adaptive scheme for efficient multi-cloud storage," in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis. IEEE Computer Society Press, 2012, p. 20.
[5] Z. Wu, M. Butkiewicz, D. Perkins, E. Katz-Bassett, and H. V. Madhyastha, "Spanstore: Cost-effective geo-replicated storage spanning multiple cloud services," in Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles. ACM, 2013, pp. 292–308.
[6] G. Greenwald and E. MacAskill, "Nsa prism program taps in to user data of apple, google and others," The Guardian, vol. 7, no. 6, pp. 1–43, 2013.
[7] T. Suel and N. Memon, "Algorithms for delta compression and remote file synchronization," 2002.
[8] I. Drago, E. Bocchi, M. Mellia, H. Slatman, and A. Pras, "Benchmarking personal cloud storage," in Proceedings of the 2013 conference on Internet measurement conference. ACM, 2013, pp. 205–212.
[9] I. Drago, M. Mellia, M.MMunafo, A. Sperotto, R. Sadre, and A. Pras, "Inside dropbox: understanding personal cloud storage services," in Proceedings of the 2012 ACM conference on Internet measurement conference. ACM, 2012, pp. 481–494.
[10] U. Manber et al., "Finding similar files in a large file system." in Usenix Winter, vol. 94, 1994, pp. 1–10.

