DATA LEAKAGE DETECTION IN CLOUD COMPUTING ENVIRONMENT

V.Suneetha[1], K.M.C.Meghana[2], K.Sruthi[3], S.M.Sanjna[4], K.Komali[5], M.Haindavi[6]

Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

ABSTRACT

The data saved in the cloud is such a valuable and essential resource, security is a crucial concern. Although it's a common misconception that hackers cause security breaches, insiders account for most data loss. Important data is regularly transferred from the distributor to reliable parties in virtually scattered setups. Given the growing number of user requests, it is essential to ensure the services' stability and safety. When a client leaks critical information, the client, responsible for the breach should be named right away. As a result, monitoring the data flowing from the distributor to the agents is required. The project finds a data leakage detection in cloud computing environment, which looks at data tampering and determines that the information leak was caused by one or more agents. Finally, the process is put into action on a cloud server.

INTRODUCTION

The innovative and quickly developing technology in the field of information is cloud computing. Almost every IT company is attempting to enter the technology market. With cloud computing, shared software and information resources are made available to devices as needed. One of the fundamental services provided by cloud computing is data storage. The staff are completely free from the irksome local data storage and monitoring thanks to the use of the cloud. It also poses a slight threat to the files' privacy, though. Data leakage is a major problem in the contemporary business environment since it must be protected against unwanted access. The unintentional or deliberate release of confidential organisational information to unapproved parties is known as data leakage. It is crucial to guard against unauthorised users misusing crucial data. Information about intellectual property rights, patents, functionality, and other relevant facts is essential. This important organisational information has frequently been distributed to several parties outside the boundaries of the company. As a result, it is challenging to find the person or entity responsible for the data leak. When organisational data is leaked by an agent, the main objective of the proposed job is to find the guilty user.

The internet is completely necessary for cloud technology because it stores data in service providers' data centres. One of the biggest issues with cloud computing technologies is data security. Data leakage, data insecurity, and data attacks by both insiders and outsiders are all possible consequences of having less control over data. Every IT organisation must concentrate on the security challenges of protecting its data from various third parties in order to prevent data leaks. Because current employees tend to be the ones that carry out leaks occasionally, the security must be beyond their awareness, making it impossible for them to know how to break it. Data leaking can occur at any time; there is no set period of time when it will occur. The quality of the private information that was released by the individual is the single factor that determines how much harm is caused. if the institution considers the disclosed material to be of critical importance. That can render the institution defenceless. The leaking can hurt sales and bring about the company's demise. Also, when cloud users access the self-care portal or dashboard, there is a chance that data will leak during the authentication phase of a communication session. Hence, the possibility of data leakage could result in security problems like a data breach on a cloud platform, which would impair data confidentiality and legal compliance. There have been numerous studies on security risks and data leakage in the cloud, but none of them looked into data leakage caused by flaws in the hypervisor and dashboard of cloud management software, or showed how to find data leaking on cloud computing platforms. Users are not completely confident in the cloud servers maintained by the cloud providers, but the data records saved in the cloud may be sensitive and private, such as employee or product details, business policies, etc. As a result Data security

in cloud computing has received considerable attention. Insufficient control over data can lead to serious security problems and threats that could lead to data leaking. The calibre of the sensitive data disclosed determines the degree of defiling caused by the data leak. The organisation may be left in a weak state if the information that was leaked is particularly important to it.

The leaking could harm the company's operations and bring them to ruin. Several methods of identifying data leakage have been developed to address this issue, including the fragmentation method, perturbation method, and others. Each of these methods has been developed to handle detecting data leaking on relational data.

LITERATURE SURVEY

Panagiotis Papadimitriou, "presented Data allotment is a prime focus of approach"

It outlines a technique that distributors can use to carefully distribute documents to customers, increasing the chance that a dishonest customer will be exposed. The distributor creates fake items that are absent from the original data collection. To maximise the likelihood of identifying clients guilty for data breach, the objects are constructed to closely resemble real objects and supplied to clients alongside real data objects. Fake objects, however, would not always be permitted because they might impair the accuracy of the operations carried out by customers. By examining the potential that a client is responsible for the leak based on the overlap of data with the leaked data, this approach demonstrates that it is practical to do so.

AL.Jeeva, "provided comparative analysis of encryption algorithms"

Comparisons were made based on factors such key length, encryption ratio, speed, tunability, and power consumption. In conclusion, the Advanced Encryption Standard (AES) is evaluated as the most preferable alternative among the symmetric encryption algorithms due to its decreased energy consumption, buffer utilisation, and encryption and decryption times.

Simon Liu and Rick Kuhn, "constructed Data Loss Prevention"

It examines different sorts of losses, including leakage—where crucial data is no longer under the organization's control—and disappearance or loss—where the organisation is no longer in possession of an exact data copy. Data loss prevention systems were developed to address governmental, industrial, and intellectual property security needs by keeping diverse sensitive data from leaving an organization's private walls. Best practises include prioritising loss modes, providing protection without interruption, and using flexible and modular design were introduced. Loss Modes are Data at Rest, Data at the End Point, and Data in Motion.

Hector Garcia-Molinam designed "DATA LEAKAGE DETECTION"

According to this model, there is a chance to determine the propensity that a third party agent is behind a leak based on the overlap between his data and the information that was leaked as well as the data from the other agents and the possibility that objects can be "guessed"

The algorithms used incorporate a number of data distribution strategies that can increase the chances of the distributer finding the leaker. Also, it has been found that carefully placing things can make a significant difference in identifying guilty agents, particularly in situations when the information that agents must obtain is highly comparable.

B. Sengupta and S. Ruj, 2016 "Utilizing safe cloud storage that is publicly verifiable for dynamic data"

Storage outsourcing is a service that cloud service providers offer to their customers. The integrity of the client's data is upheld through a secure cloud storage (SCS) protocol. In this work, we build a secure cloud storage system that is publicly verifiable and based on the secure network coding (SNC) protocol, allowing the client to update the outsourced data as necessary. Our protocol is the first SNC-based SCS protocol for dynamic data that is secure in the standard model and offers privacy-preserving audits in a publicly verifiable environment, as far as we are aware. Additionally, we go into great depth about the (im)possibility of constructing a universal SCS protocol for dynamic data (DSCS protocol) from a random SNC protocol. Additionally, we alter a current DSCS scheme (DPDP).

PROPOSED SYSTEM

We investigate the potential for establishing a secure cloud storage for dynamic data using the secure network coding techniques. We design such a protocol (DSCS I) based on a secure network coding protocol in order to demonstrate how some secure network coding schemes can be utilised to build effective secure cloud storage protocols for dynamic data. To the best of our knowledge, the first standard-model safe cloud storage protocol for dynamic data that was developed using secure network coding techniques is called DSCS I. Although append-only data have many uses in the real world, generic dynamic data permit unrestricted insertions, deletions, and alterations. We develop DSCS II, a secure cloud storage system tailored to append-only data that addresses several DSCS I's drawbacks. Lastly, in order to assess the performance of DSCS I and DSCS II, we offer prototype implementations.

We investigate the feasibility of offering a generic DSCS protocol creation from any SNC protocol. We go into great length on the difficulties of a generic construction and name a few SNC protocols that can be used to build effective DSCS protocols.

With an SNC protocol, we create a publicly verifiable DSCS protocol (DSCS I). DSCS I manages dynamic data, allowing a customer to change the outsourced data effectively (via insertion, deletion, and modification). We go through DSCS I's (asymptotic) performance and some of its drawbacks.

We demonstrate the security of DSCS I and offer a formal description of a DSCS protocol's security.

We can utilise DSCS I for append-only data because they are a particular example of generic dynamic data. We do, however, identify specific SNC protocols that can be used to create secure cloud storage protocols that are effective for append-only data and not for constructing secure cloud storage for dynamic data in general. Using an SNC protocol proposed by Boneh et al., we create a publically verifiable secure cloud storage protocol (DSCS II) for append-only data.

We talk about the (asymptotic) performance of DSCS II, which fixes several DSCS I's flaws.

DSCS I and DSCS II are implemented, and their effectiveness is assessed in terms of storage overhead, computational cost, and communication cost.

FLOW DIAGRAM

Fig1: Flow Diagram

The cloud server keeps a lot of data for its customers (data owners). To save space, a malicious cloud server may choose to delete some of the client's data that is viewed seldom. A method to determine whether the server is storing the client's data in an unaltered manner is provided via secure cloud storage protocols (two-party protocols between the client and the server). These protocols are divided into two categories based on the kind of outsourced data: secure cloud storage protocols for static data (SSCS) and for dynamic data (DSCS). After the initial outsourcing, the client cannot modify her static data (such as backup/archival data). The client can change her data as often as necessary with dynamic data, making them more generic.

The customer can audit the outsourced data using secure cloud storage methods without having to examine the entire data file, and they will still be able to spot any unauthorised changes made by a hostile server. The server receives a random challenge from the client during an audit, and the server responds with proofs of storage (calculated using the stored data) that correspond to that challenge. If an audit can be carried out by any third-party auditor (TPA) using open parameters, then secure cloud storage procedures are publicly verifiable; otherwise, they are privately verifiable if the auditor requires some confidential client information. the participants and interactions in a secure cloud storage mechanism.

CONCLUSION

Based on the secure network coding (SNC) protocol, we have suggested the secure cloud storage for dynamic data (DSCS I) protocol. This is the first SNC-based DSCS protocol that, to the best of our knowledge, is secure in the standard model and has public verifiability. We have talked about various difficulties in converting an SNC protocol into an effective DSCS protocol. We have also noted several restrictions on a secure cloud storage technique for dynamic data based on SNC. However, some of these restrictions are a result of the SNC protocol that was employed as a foundation. We can create a more effective DSCS protocol by using a more effective SNC protocol. Also, we have developed an effective DSCS protocol (DSCS II) for append-only data and identified specific SNC protocols that are appropriate for append-only data. We have demonstrated how DSCS II gets around some DSCS I drawbacks. Lastly, in order to demonstrate the viability of DSCS I and DSCS II, we have presented prototype implementations of both standards. We have also contrasted DSCS I's performance with that of DPDP I and SNC-based secure cloud storage for static data.

FUTURE SCOPE

The time and resources needed for downloading, updating, and submitting the material again can be saved for future work and enhanced.

REFERENCES

1. Sengupta and S. Ruj, “Publicly verifiable secure cloud storage for dynamic data using secure network coding,” in ACM Asia Conference on Computer and Communications Security, 2016, pp. 107– 118.
2. International Conference on Computational Intelligence and Communication Networks (CICN) - Bhopal, India (2014.11.14-2014.11.16)] 2014 International Conference on Computational Intelligence and Communication Networks - Detection of Data Leakage in Cloud Computing Environment., (), 803–807. doi:10.1109/cicn.2014.172
3. G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, L. Kissner, Z. N. J. Peterson, and D. X. Song, “Provable data possession at untrusted stores,” in ACM Conference on Computer and 2. Kumar, Neeraj; Katta, Vijay; Mishra, Himanshu; Garg, Hitendra (2014). [IEEE 2014 Communications Security, 2007, pp. 598–609.
4. Papadimitriou, Panagiotis; Garcia-Molina, Hector (2011). Data Leakage Detection. IEEE Transactions on Knowledge and Data Engineering, 23(1), 51–63. doi:10.1109/TKDE.2010.100
5. A. Juels and B. S. Kaliski, “PORs: Proofs of retrievability for large files,” in ACM Conference on Computer and Communications Security, 2007, pp. 584–597.
6. Abhijeet Singh, Abhineet Anand, “Data Leakage Detection Using Cloud Computing” International Journal of Engineering and Computer Science, Volume 6, Issue 4, April 2017.
7. H. Shacham and B. Waters, “Compact proofs of retrievability,” Journal of Cryptology, vol. 26, no. 3, pp. 442–483, 2013.
8. Abdullah Bamatraf, Rosziati Ibrahim and Mohd, Najib Mohd Salleh, “A New Digital Watermarking Algorithm Using Combination of Least Significant Bit (LSB) and Inverse Bit”, International Journal of computing, volume 3, Issue 4, April 2011.

9. C.C. Erway, A. Kupc¸ ¨ u, C. Papamanthou, and R. Tamassia, ¨ "Dynamic provable data possession," ACM Transactions on Information and System Security, vol. 17, no. 4, pp. 15:1–15:29, 2015.
10. Naik, Riya; Gaonkar, Manisha Naik (2019). [IEEE 2019 International Conference on Computer Communication and Informatics (ICCCI) - Coimbatore, Tamil Nadu, India (2019.1.23-2019.1.25)] 2019 International Conference on Computer Communication and Informatics (ICCCI) - Data Leakage Detection in cloud using Watermarking Technique., (), 1–6. doi:10.1109/ICCCI.2019.8821894