International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Data Leakage Detection and Security System over the Cloud Computing

Mohankumar S[1], Jegadeeswaran R[2], Hari prasath S[3], Danusu B[4], Mr.Rajiniganth R[5]

[1234]Students, Department of Computer Science and Engineering, KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India.
[5]Assistant Professor, Department of Computer Science and Engineering, KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India.

A B S T R A C T

Cloud computing is now a day's turns into the most pulled in marvels to use for vast scale association or for person who require different system administrations with slightest expense. Regularly person's information is put away on public Cloud which is accessible to everybody for access. This major raises some issue inverse to adaptable administrations given by cloud suppliers, similar to Confidentiality, Integrity, Availability, Authorization and some more. It seems that in cloud computing environment the major problem that ensure the secure communication and protect responsive data in open networks from unauthorized access. Some of the data are leaked and found in an unauthorized place (e.g., on the web or somebody's laptop). Data leakage happens every day when confidential business information are leaked out. When these are leaked out it leaves the company unprotected and goes outside the jurisdiction of the corporation. The distributor must assess the likelihood that the leaked data came from one or more agents, as opposed to having been independently gathered by other means. We propose data allocation strategies (across the agents) that improve the probability of identifying leakages.

Keywords: Cloud Computing, Data Leakage, Data Security, CSP, SLA, Cloud Cryptography

1. INTRODUCTION

Data loss occurs when data is intentionally or unintentionally removed physically or logically from an organization. Data loss has become the largest problem in organizations today, and organizations are responsible for addressing this problem. Data leakage is defined as the accidental or unintentional distribution of private or sensitive data to an unauthorized entity. Data leakage poses a serious issue for companies as the number of incidents and the cost to those experiencing them continue to increase. Data leakage is enhanced by the fact that transmitted data including emails, instant messaging, website forms, and file transfers among others, are largely unregulated and unmonitored on their way to their destinations.
The main scope of this module is providing complete information about the data/content that is accessed by the users within the website. Forms Authentication technique is used to provide security to the website in order to prevent the leakage of the data. Continuous observation is made automatically and the information is send to the administrator so that he can identify whenever the data is leaked.

1.1. Problem Definition :

Current Data Leakage Detection Project propose data allocation strategies that improve the probability of identifying leakages. In some cases, we can also inject “realistic but fake” data records to further improve our chances of detecting leakage and identifying the guilty party. The distributor cannot blame the agent without any evidence. This project identifies the agent who leaked the data with enough evidence.

1.2. Objectives:

The objective of this project is to improve the probability of identifying leakages using Data allocation strategies across the agents and also to identify the guilty party who leaked the data by injecting "realistic but fake" data records.The goal of Data leakage detection is to detect when the distributor’s sensitive data has been leaked by agents, and if possible to identify the agent that leaked the data. Perturbation is a very useful technique where the data is modified and made “less sensitive” before being handed to agents. we develop unobtrusive techniques for detecting leakage of a set of objects or records. In this section we develop a model for assessing the “guilt” of agents.

1.3. The type of employees that may leak data:

- The security illiterate
    - Majority of employees with little or no knowledge of security
    - Corporate risk because of accidental breaches
- The gadget nerds
    - Introduce a variety of devices to their work PCs

    - Download software
- The unlawful residents
    - Use the company IT resources in ways they shouldn't
    - i.e., by storing music, movies, or playing games
- The malicious/disgruntled employees
    - Typically, minority of employees
    - Gain access to areas of the IT system to which they shouldn't
    - Send corporate data (e.g., customer lists, R&D, etc.) to third parties.

2. DATA LEAKAGE IN CLOUD COMPUTING PLATFORM:

Advances in technology over the past few years have enabled cloud computing to change the way organizations work, moving 4, workloads offsite. Cloud computing enables cost-effective and flexible delivery of IT services and resources, including databases, bandwidth, software, servers, storage, networks, and more over the Internet. Today, this new technology is gaining great popularity with interest from academia and industry . Running a private data centre or having, large secondary storage facilities is over budget for many organizations. Cloud storage is the best option for these organizations due to the flexible service model . As shown in Figure 1, three cloud storage models are available.

Despite the many benefits of cloud computing,, there are some technical hurdles and security issues, such as data integrity, privacy, and privacy. When a user or organization stores data or information in the cloud, their personal information data.Cloud service providers (CSPs) must use a variety of mechanisms to protect their customers' data from modification and corruption . Cloud Services Service Providers (CSPs) are responsible for ensuring information security and are limited by service level agreements (SLAs), but do not provide 100% data integrity.

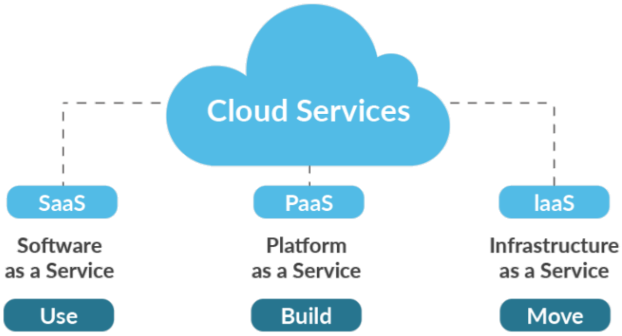

Fig.1.Cloud Service Model

2.1 Data Protection in Cloud Computing

Computing resources such as CPU, memory and storage in the cloud are mainly distributed across multiple hosts or virtual machines. Therefore, it is important to have a data protection scheme to protect your data privacy. Table 1 below shows the three types of data that need to be protected when deploying in the cloud.

Table 1: Type of Data Need To Be Protected in the Cloud

Type	Risk Level	Description
Data in Transit	High	A condition where transferring data via wired or wireless connectivity through a network or Internet make it at risk
Data in Use	Medium	A condition where data is currently used, manipulated or being hold in memory or cache at endpoints of the network such as laptops, USB devices or mobile devices.
Data at Rest	Medium - Low	A condition where the data are stored in databases, storage devices or files systems for persistent storage.

2.2 Data Leakage Threats

Unintentional or accidental distribution of sensitive or confidential data to an unauthorized entity is the definition of data leakage. Unauthorized transfer of classified information from an organization to an outside recipient was also associated with data leakage. By transmitting data processes, such as website forms, emails, instant messaging and file transfer which are unmonitored and unregulated to their destination could make data leakage issue become intensifies. According to the perspective of other researcher, undefined unsolicited revelations of information can also be defined as a data leakage. According to Cloud Security Alliance, the rate of data leakage event being reported in cloud platforms are increasing. Unintentional or accidental distribution of sensitive or confidential data to an unauthorized entity is the definition of data leakage. Unauthorized transfer of classified information from an organization to an outside recipient was also associated with data leakage. By transmitting data processes, such as website forms, emails, instant messaging and file transfer which are unmonitored and unregulated to their destination could make data leakage issue become intensifies. According to the perspective of other researcher, undefined unsolicited revelations of information can also be defined as a data leakage.

According to Cloud Security Alliance, the rate of data leakage event being reported in cloud platforms are increasing. Hardware errors, insecure interfaces and APIs, data loss, and leaks were listed as the top 3 events in Reported Threats. More than half of the reported cloud incidents were caused by these threats. The threat of internal malicious activity, common technical issues, abuse and criminal use, hijacking of accounts or services, and unknown risk profile make up the remaining threats to cloud computing. According to a survey conducted in 2018, the number of data breaches is increasing in the industry, highlighting the importance of remedying data breaches in the cloud.

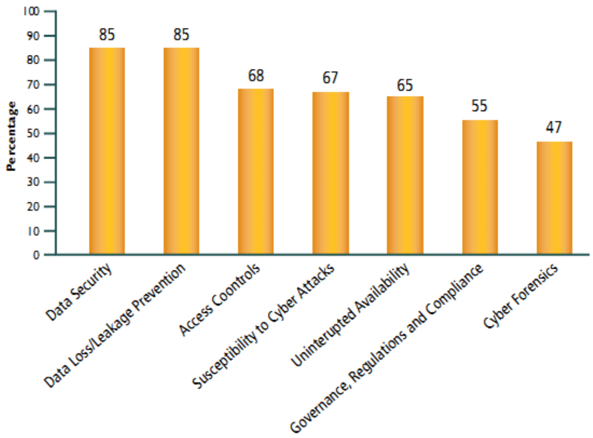

Fig.2.Result of the Survey

1. Attack - An attack is a purposeful attempt to cause damage or loss through technical or social means. Attacks do not necessarily lead to data breaches. For example, denial of service attacks disrupts normal operations but do not expose information.
2. Breach - A successful attack was able to secure sensitive information.
3. Hack - An attack that exploits technical vulnerabilities to secure access that is otherwise unauthorized. Can lead to a breach, but may also be used for ransomware (like WannaCry), establishing botnets, or misusing computing resources.
4. Leak - A leak does not require an external actor, but is caused by some action or inaction of the party who owns the data.

3 LITRATURE REVIEW

Literature survey is the most important step in software development process. Before developing the tool, it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system, the above considerations are taken into account for developing the proposed system.

1) The issue of data security is basically stored in the cloud data store, which is a distributed storage system, and the traditional method uses RSA to validate erasure-encoded data to support duplicate equivalence vectors, erasure correction code for file distribution. It is encryption based and focuses on a scheme where ensures the confidentiality of customer data through the localization of data errors and the security and consolidation of storage.

2) The Uses browser historical data to determine where the attack occurred. Data is collected from browser history and stored in a database as evidence. Proposed methods and forensic tools help analyses data from law enforcement agencies.

3) A strong watermarking method can be very useful in some cases, but again requires minor modifications to the original data. Also, the watermark can be destroyed if the recipient of the data is malicious. Distributors may add fake objects to the data being distributed to increase the effectiveness of detecting guilty agents.

4) Data breaches are a huge problem facing the industry and various organizations. They discussed that distributors create and add fake objects to the data they distribute to agents. Distributors may add fake objects to the data being distributed to improve the effectiveness of detecting guilty agents. Leak detection is handled by algorithms and the integrity of users of these systems is at stake.

4 EXISTING METHODOLOGY

Traditionally, leakage detection is handled by watermarking, e.g., a unique code is embedded in each distributed copy. If that copy is later discovered in the hands of an unauthorized party, the leaker can be identified. Watermarks can be very useful in some cases, but again, involve some modification of the original data. Furthermore, watermarks can sometimes be destroyed if the data recipient is malicious. E.g. A hospital may give patient records to researchers who will devise new treatments. Similarly, a company may have partnerships with other companies that require sharing customer data. Another enterprise may outsource its data processing, so data must be given to various other companies. The Existing System can detect the hackers but the total no of cookies (evidence) will be less and the organization may not be able to proceed legally for further proceedings due to lack of good amount of cookies and the chances to escape of hackers are high.

Drawback

Insider attack: The cloud service provider is not trusted. External Attacks: These attacks originate from third-party and non-domain-controlled cloud service providers. Current approach can detect the hackers but the total number of evidence will be less and the organization may not be able to proceed legally for further proceedings due to lack of good amount of evidence and the chances to escape of hackers are high

4. PROPOSED METHODOLOGY

The aim of this proposed methodology to detect when the distributor's sensitive data have been leaked by agents, and if possible to identify the agent that leaked the data. The proposed method proposes the addition of a "realistic but false record" that increases the probability of data distribution strategy and leak detection and the encryption algorithm improves data security. It can also keep your data safe and detect leaks during transmission. To address data breaches, implementations of various data dissemination strategies have been submitted that may increase the likelihood that a data breach will be detected by a distributor. Data leak detection, the proposed system shown in Figure 3 is to detect when the confidential data of the distributor is lost by the agent, and possibly identify the agent with the data leak. Perturbation is a very useful technique to make data modified and "less sensitive" before being passed to agent. To increase the security of user data, the model has an encrypted algorithm that ensures secure data transfer between the user and the TPA. It includes the cloud as well as algorithms that distribute objects between agents in a way that improves the likelihood of detecting leaks. When adding misinformation objects to a distributed set, misinformation is provided if the data is provided to another third party.

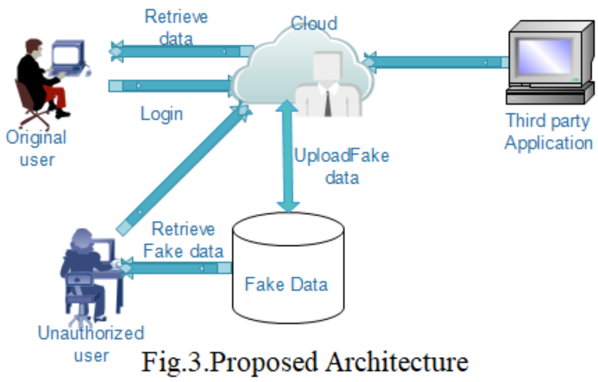

Fig.3.Proposed Architecture

Advantages

1.Algorithm used to distribute the objects to agents that improves the chances of identifying a leaker.
2. Realistic but fake objects is injected to the distributed set. .
3.Leakers cannot argue that they did not leak the confidential data, because this system traces leakers with good amount of evidence.

5. CONCLUSION

The proposed method proposes the addition of a "realistic but false record" that increases the probability of data distribution strategy and leak detection and the encryption algorithm improves data security. It can also keep your data safe and detect leaks during transmission. To address data breaches, implementations of various data dissemination strategies have been submitted that may increase the likelihood that a data breach will be detected by a distributor. The scope of this system can be extended to allow for the dynamic creation of forged records at the request of the agent.This shows that the method can effectively detect data leaks in cloud platforms. When comparing to related operations, the other methods focus more on data leaks occurring at the application level. Going forward, we would like to expand our investigation by discovering potential data breaches in other cloud components such as block storage, API requests, and telemetry.

6. FUTURE ENHANCEMENTS

There is scope for future development of this project. The world of computer fields is not static; it is always subject to be dynamic. The technology which is famous today becomes outdated the very next day. To keep abstract of technical improvements, the system may be further refined. So, it is not concluded. Yet it will improve with further enhancements. It can be improved to include new features. Our application is no different from this. The future enhancements that can be made to Data Leakage Detection are:

- Providing support for other file formats.
- Creation of a web based UI for execution of the application.
- Improving the detection process based on user requirements.
- Provision of quality or accuracy variance parameter for the user to set.

Acknowledgements

We grateful to Mr.R.Rajiniganth,M.E.,AP/CSE for his unprecedented guidance, suggestions for this paper.

REFERENCES

[1] Chaudhary, H., Chaudhary, H., & Sharma, A. K. (2022). Optimized Genetic Algorithm and Extended Diffie Hellman as an Effectual Approach for DOS-Attack Detection in Cloud. International Journal of Software Engineering and Computer Systems, 8(1), 69-78.

[2] Alshammari, S. T., & Alsubhi, K. (2021). Building a reputation attack detector for effective trust evaluation in a cloud services environment. Applied Sciences, 11(18), 8496.

[3] Wang, X., Pan, Z., Zhang, J., & Huang, J. (2021). Detection and elimination of project engineering security risks from the perspective of cloud computing. International Journal of System Assurance Engineering and Management, 1-9.

[4]Sharma, A., Singh, U. K., Upreti, K., & Yadav, D. S. (2021, October). An investigation of security risk & taxonomy of Cloud Computing environment. In 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1056-1063). IEEE.

[5] Al-Shehari, T., & Alsowail, R. A. (2021). An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques. Entropy, 23(10), 1258.

[6]M.Sai Charan Reddy, T. Venkata Satya Yaswanth, T. Gopal, L. Raji, Dr. K.Vijaya, “DATA LEAKAGE DETECTION USING CLOUD COMPUTING”, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Volume: 06 Issue: 03, Mar 2019.

[7]Chandu Vaidya etl. & BE scholars “Data leakage Detection and Dependable Storage Service in cloud Computing” IJSTE volume 2 issues 10 April 2016 ISSN online 2349-784.

[8] Prof. Sushilkumar N. Holambe, Dr.Ulhas B.Shinde, Archana U. Bhosale,”data Leakage Detection Using Cloud Computing ”, International Journal Of Scientific & Engineering Research, Volume 6, Issue 4,(April-2015)

[9] Khobragade, P. K., & Malik, L. G., “Data Generation and Analysis for Digital Forensic Application Using Data Mining”. In Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on (pp. 458-462). IEEE. April, 2014.