IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Data Leakage Detection Strategy in Cloud Computing Environment

Dr. S Adinarayana[1], M Ajay Kumar[2], K Prasanna[3], M Sai Krishna Pawan[4], M Sri Nikhil[5]

Professor, Department of Computer Science and Engineering[1]

U.G Scholars, Department of Computer Science and Engineering[2345]

Raghu Institute of Technology, Visakhapatnam, AP, India.

Abstract: *Anxiety is a useful technique that modifies data and makes it "less sensitive" before sending it to the agent. For example, some products can be added to random popularity or the amount of money can be changed by many factors. But sometimes it is important not to change the original data of the distributor. For example, if a professional pays us, he must have a salary and a bank account. Medical researchers will need accurate data about patients (as opposed to statistical calculations) if they are to treat patients. Watermark detection is traditionally done by watermark, that is, by placing a watermark on each printed document. If this document is later found in the hands of an unauthorized person, the leaker can be identified. The watermark can be useful in some cases, but again there are some changes to the original file.*

Additionally, watermarks can sometimes be affected if the recipient receives the information of poor quality. In this article, we explore non-invasive methods for detecting leaks in products or materials. Specifically, we examine cases where, after the broker has been given a set of goods, the seller finds some of the original goods in an illegal location.

Index Terms *- Data leakage, audit server, integrity, unauthorized, optimization*

I. INTRODUCTION

In the course of business, it is sometimes necessary to hand over sensitive information to trusted third parties. For example, a hospital can provide patient information to researchers who can develop new treatments. Likewise, a company can join a partnership with another company that wants to educate its customers. Another business will outsource data processing and therefore must make the data available to many other companies. By distributor we mean the owner of the document and as the agent we mean the assumed third party. Our goal is to determine when a business owner's information was accessed by an agent and, if possible, to identify the agent who leaked the information. Senders can verify that the information comes from one or more agents and is not collected by in other ways. Using the example of stealing cookies from a cookie jar, if we catch Freddie with the cookie, he can claim that his friend gave him the cookie. However, if we caught Freddie with 5 cookies, he would claim that his hand was not in cookie dough. If the supplier finds "sufficient evidence" that the agent has leaked information, it must stop doing business with it or legal action will be taken. In this article, we develop a model to measure the "guilt" of the agent. We also provide an algorithm for distributing items to employees, improving the way we identify leaks. Finally, we are also considering the option of adding "fake" products to the product distribution. These objects do not correspond to real places, but are real to the agent. In a sense, the fake item acts as a kind of watermark for the entire collection without replacing a single member. If it turns out that the agent has received one or more counterfeit goods, the sender can be more confident that the agent is guilty.

Exporters can add counterfeit goods to their export information to be more efficient at detecting violators. However, counterfeit products may affect the authenticity of the recording, so they will not always be allowed. For example, the idea of corrupting data to detect leaks is not new. However, individual elements are often affected, such as adding noise to sensitive salaries or adding watermarks to images. In our case, we interfered with the export by adding counterfeit products. In some applications, counterfeit products may cause less problems than real products. For example, suppose the data distribution is a medical record and the representative is a hospital. In such cases, even minor changes to actual patient information may not be approved. However, it may also confirm some inaccurate medical information, as there are no patients objecting to this information and therefore no one will be considered false information. Our use of counterfeit products results from the use of "working" data on mailing lists.

In this case, Company A sells a disposable postcard to Company B (for example, to send an advertisement). Company A adds tracking information that includes addresses of company A. For this reason, a letter sent to company A is sent for every mailing list purchased by the bus company. This information is a type of artificial information and helps identify inappropriate use of information. The d istributor creates the fake product and adds it to the information it distributes to the broker. We allow Fi_Ri t o be a subset of artifacts received by the agent UI.

As described below, counterfeit products must be manufactured with care so that a representative cannot dis tinguish them from real products. In general, the supplier can limit the number of counterfeit products it can produce.For example, a product may have an email address and a real inbox must be created for each fake e mail address (but the agent will see that the product is fake).Because creating and monitoring email account s uses resources, senders can limit spoofing. If there is a limit, we see it as B's artifact. Similarly, distributor s may want to limit the number of fakes each agent receives to avoid surprises and affect the agent's busines s. Therefore, we recommend that the distributor can send most counterfeit products to the Ui Creation repres entative.

The fake but really good thing is a trivial topic whose investigation is better than the scope of this article. Here, we model the creation of a fake object for agent Ui as a black-box function CREATE FAKE OBJECT(Ri; Fi; Condi) that takes as input the set of all objects Ri, the subset of fake objects. Fi is what Ui has taken so far and Condi is returning the new artwork. This work requires Condi to create us able products that satisfy the Ui gang. Ri should be set as a strategy so that the counterfeit product is not onl y valid, but also indistinguishable from other real products. For example, generating fake salary data that inc ludes employee levels and salary characteristics may reflect employee classification, salary distribution, and the relationship between these two characters. If agents are going to use statistics in their business, it is imp ortant to ensure that important statistics remain unchanged by showing fake things.

2 LITERATURE REVIEW

2.1 Agent Guilt Model

A Ui agent will commit a crime if it throws one or more objects at the target. The event in which Agent U_i is bound to the given infiltration set S is denoted by $G_i|$

S. The next step is to estimate Pr { $G_i|$ S }, i.

That is, to calculate S.

Pr {$G_i|S$}, the probability that agentGi will commit a crime given evidence, estimate the probabi lity that the value in S can be "guessed" by the target. For example, suppose some items in t are people's emails.

Do an experiment where one person sees the email of 100 people and only 20 people will see it, resultin g in an estimated 0.2. Call the prediction pt, the probability that the target predicted the object is t.

Two ideas about the relationship between various events.

Assumption 1: For all t, t ∈ S such that t ≠ T, the proof of t is independent of the proof of T. . The sour ce can be an agent with t in its configuration, or the target itself.

Assumption 2: An object t ∈ S can be recovered from the target in only two ways.

An agent Ui leaks t from its set Ri, or the target guesses (or receives) t without the help of an agent n .

To find the probability that Agent U_i is guilty of emitting the S light, consider the target t1with probability p and the agent going from t1 to S with probability 1-p.

First calculate the probability of leaking a product t to S. To calculate this, define the set of representations o f t to V_t= {U_i | t<-R_t} in their data. Then using conjecture 2 and knowing the probability p, we have,Pr {some reps leaked t into S} = 1-p

(1.1)that all representatives of V_t leak t into S with equilibrium And let's pretend you're used to it. thought 2 obtain,

$$\Pr\{U_i \text{ leaked t to S}\} = \begin{cases} \frac{1-p}{|Vt|}, & \text{if } Ui \in Vt \\ 0 & \end{cases} \quad (1.2)$$

Equation 1.2, agentU_i is guilty,

$$\Pr\{Gi|S\} = \prod t \in s \cap Ri\left(1 - \frac{1-p}{|vt|}\right) \quad (1.3)$$

2.2 Data Allocation Problem

Distributors provide "intelligent" information to agents to increase their chances of finding the culprit. T here are four aspects to this question, depending on the type of information the header requests and whet her "falsification" is allowed [4]. Name makes two types of requests, called standard requests and specia l requests. When prompted, add the artifact to the database.

A counterfeit product is a product manufactured by the importer and not included in the T collection.The se products are designed to look like real products and are distributed to employees along with T product s to prevent employee leaks.

2.3 Optimization Problem

Distributors have a limit and purpose for distributing information to vendors. Sellers have done everythi ng to satisfy the reps' requests, providing the products they wanted or all the available products that mad e them happy. Its purpose is to detect agents that have leaked some of their information.

We consider the restrictions to be strict. Sellers should not refuse the requested name and should not offe r different versions of the same product to the agent.The distribution of counterfeit products according to the limit can only be relaxed. The goal is to quickly identify the perpetrator who leaked all product infor mation.

The Pr {G_i|S =R_i } isthe probability that agent U_i is guilty if the distributor discovers a leaked table *S* that contains all R_i objects.

The difference functions Δ (i, j) is defined as:

$$\Delta(i, j) = \Pr\{\frac{G}{Ri}\} - \Pr\{\frac{G}{Ri}\} \quad (1.4)$$

1) Problem description

Allow distributor to receive information requested by n representative. Distributors want to supply R1, .Rn paper. For a representative, U1. .. , Un

- a distribution that satisfies the rights of the intermediary; and
- Maximize the crime rate difference Δ(i, j) for all i, j = 1, respectively. .n and i = j.

Assuming that the configuration meets the requirements of the agent, we can create the problem a ccording to various rules.

2) OptimizationProblem

The approximation of objective of the above equation does not depend on agent'sprobabilities and therefore minimize the relative overlap among the agents as Minimize (. . . ,($|R_i \cap R_j|$) / R_i , . . .) i != j (1.6)(over R1 , . . . ,Rn)

This approximation is valid if minimizing the relative overlap, ($|R_i \cap R_j|$) / R_i maximizes Δ (i, j).

3 Allocation Strategies Algorithm

There are two types of strategy algorithms

Explicit data Request

Distributors do not allow forgery to be added to distribution without clear proof that counterfeiting is prohib ited. The distribution of data is therefore carried out entirely at the request of the data representative. Sellers must not delete or modify the R claims of authorized representatives who claim false information. But suppl iers can add fakes.

In the data allocation algorithm for strict requests, the input is a series of requestsR1, R2, ..., Rn, and various requests from n agents. The eoptimal algorithm finds suitable agents to receive counterfeit products. Then c reate the object in this iteration and assign it to the selected agent. The e-optimal algorithm generates optimal solutions by adding more number of b_i of fake objects to each R_i cluster so that each term of the objective sum is reduced.

4.EXISTING SYSTEM

Stealth Oracle also uses a similar technique the authors refer to as black box blurring to monitor network connections for interference from different devices. However, the system needs to complete the application and rollback to repeat with different strategies, it cannot provide real-time protection, and the process of identifying different components in the network is very sensitive to retransmission of packets. Another new and traditional way to reduce information leakage is information flow. However, static information requires access to the source code, so it cannot support older software. To solve this limitation, dynamic blot analysis is used at a very high cost to monitor the propagation of private information throughout the system. However, such strategies do not allow the sharing of personal data on limited connections and may lead to delays of 20 times or more.

Disadvantage

- It is not efficient enough for practical data leak inspection in this setting.
- The customer or data owner does not need to fully trust the cloud provider.
- Keywords often do not cover information that is sensitive enough to cause information leaks.
- It does not aim to provide a remote service.
- In this case, using it to find the leaked information is not useful enough.

Advantge

- Computation cost is low.
- The computation burden is not huge for the users. Data Process also work in very efficient manner.
- Cloud Audit Server (CAS) need not to be trusted.

Prove resilient to reboot attacks in a better security model while maintaining public service and managing data in motion

5.Proposed System

In proposed system, the system proposes a new cloud storage scheme in proof of retrievable for cloud storage, in which a trustworthy audit server is introduced to preprocess and upload the data on behalf of the clients. On the other side we improve the semi-honest trust worthy and ensure dynamic data process in cloud. And This system develops a strengthened security model for considering data security against Data Leakage and the storage server in the upload phase of an integrity verification scheme. And presents an efficient verification scheme for ensuring remote data integrity in cloud storage.

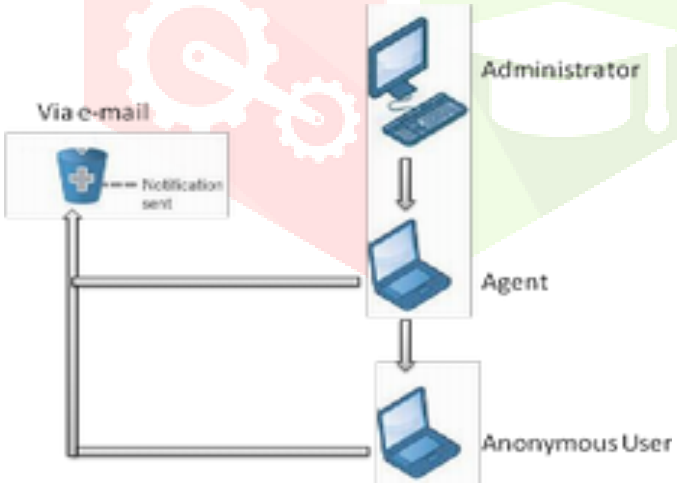

Figure 1.System Implementation

Study phase includes careful planning, analysis of existing methods and limitations of use, design of proced ures to effect project change, and measuring change.

5.1.2 Modules

- Client Phase
- User service privilege in cloud
- Integrity proof for user Service Access
- Cloud Audit server Phase

5.1.2 Module Description

1. **Client Phase**

- In this Module we Verify the cloud user authentication
- When enter the system, The person has validate using their id and cloud password. If new user enter this phase, they go to registration phase and enter their personal detail, as well as secret code , In Security Purpose We encrypt The code Using AES algorithm .
- If the user doesn't trust, their process has abort.
- It proved secure against reset attacks in the strengthened security model while supporting efficient public verifiability and dynamic data operations simultaneously.

2. **User service privilege in cloud**

- In this module describe about user privilege in cloud. Here cloud has provide three type of service for authorized users.
- Data upload
- Data download
- Dynamic Data deletion in cloud
- Data upload : User data has upload in cloud while satisfy the user integrity proof.
- Data download : Data has download from cloud after check the security code.
- Data Deletion : Data has download from cloud after check the security code.

3. **Integrity proof for user Service Access**

- In this Module Integrity Proof check the users current request, if the user has a Come from data upload request , then they allow to select their data , calculate that data size and source location and simultaneously cloud audit server check the availability space on ur account, if available space is exist in cloud , then only data allow to store in cloud. If the cloud free space not sufficient in ur account, file doesn't allow to store in cloud server.
- If User request has come for download data from cloud, the integrity check user's secret code. This security checking process has follow to delete un wanted data from cloud in dynamically under the security verification.

4. **Cloud Audit server Phase**

- The cloud server reduces the clients burden by maintaining their files and sequentially monitor user's cloud account while end of user's action it dynamic update user's account .
- The process takes over by using Integrity Verification algorithm.
- Its work carrying out multiple challenges-responses. Under the Shacham and Waters proposed a security model for PoR system which achieve number of checking.

The client periodically challenges the storage server to ensure the correctness of the cloud data and the original files can be recovered by interacting with the server and it dynamically as well as for download and delete also.

Future work

The idea of an environment of trust is a little shaky. The departure of a trusted employee with access to s ensitive information can create a data breach if the trusted employee continues to access the information after the trust has expired. In distribution, this can also happen when the trust of the website is broken. M ost of these cases reported by the media, for example;Loss of business or government information such a s social security numbers (such as trade secrets, important company information, contract information, et c.); Publishing such situations may cause more harm than losing the file itself.

While such a situation poses a risk of theft or other serious consequences, there is usually no permanent d amage; even if the security breach is fixed before unknown persons can access the data, or the thief is onl y interested in the device, not the data inside. However, when such cases are widely publicized, criminals often try to minimize losses, for example by giving the victim a name on the organization's credit report.

CONCLUSION

Protecting sensitive information from leakage is an important and practical research problem. Algorithms in this project achieve results achieved by iterative or limited discounting by agreement. .This uses and evaluat es a new privacy storage data leak detection system that allows data owners to securely transmit on-premises or send traffic audits to cloud providers that do not disclose sensitive information. Addressing the challenge of reducing the computational burden on users.
In cloud computing, users store their data files on cloud servers. For this reason, it is important to prevent un authorized access to resources and to ensure safe sharing of resources. In the traditional management system , we usually assume that the data manager and storage are in the same security center and the server is comp letely trustworthy. These attacks can leak users' personal data for business gain, as data owners often store d ecrypted data on cloud servers. How to manage access to encrypted data in an insecure environment and ho w to ensure the confidentiality of user data is a problem that cloud computing technology is used and applic ations need to be solved.

REFERENCES

[1]Muhammad Azizi Mohd Ariffin, "Data Leakage Detection in Cloud Computing Platforms", Lub Yim Hli 2019,International Journal of Advanced Trends in Computer Science and Engineering 8(1.3):400-408, DOI:10.30534/ijatcse/72019 .

[2]K. Ciam teb thiab A. Prakash, "Digitization of Information Leakage in Outbound Network Traffic", Proc. 30. IEEE Symptoms.Safety. Confidential, May 2009, p. 129-140.

[3]Grobauer, B., Walloschek, T. and Stocker, E. (2011). Learn about cloud computing vulnerabilities. IEEE Security and Privacy, Vol.9, no. 2 one, p. 50-57, Three.

[4] Data Leakage Detection, an IEEE article, Panagiotis Papadimitriou, IEEE Fellow and Hector Garcia-Molina, IEEE Fellow, October 2010.

[5] Chandu Vaidya and Prashant Khobragade, 2015, "Data Security in Cloud Computing", ISSN: 2321-8169.

[6] Michael Miller, "Cloud computing" Webbased applications that change the way you work and collaborat e online, Pearson Education, 2012.

[7] X. Shu and D. Yao, "Data Leak Detection as a Service", Transaction 8. International Meeting. Safety. Privacy Community Networks, 2012, p.222–240 : ca.

[8]H. Yin, D.Nkauj, M. Egele, C. Kruegel, thiab E. Kirda, "Panorama: Capturing System-Wide Information Flow for Malware Detection and Analysis", Proc. 14th Annual ACM Conferencecomp uter. argument. Safety. , 2007, p. 116-127.

[9] K. Borders, E. V. Weele, B. Lau ve A.Prakash, "Securing confidential information on personal computer s using Storage Capsules", Proc. 18. USENIX Security. symptoms. , 2009, p. 367-382.

[10]Payne and W. Lee, "Gyrus: User Interface Monitoring Framework for Web Applications", Proc. 23. US ENIX SAFE. Symptoms, 2014, p.79-93: I.

[11]P. Buneman, S. Khanna, W.C.Tan. Why and where: characteristics of data sources. Bussche and V.Vian u, ed., Database Theory ICDT 2001, 8th International Conference, London, UK, 4-6 Ocak 2001, Proceedings, Lecture Notes in Computer Science Cilt. 1973, p. 101–1 316–3 Springer, 2001.

[12] Ambrust Michael, Amando Fox, Rean Griffith, Anthony D, Joseph, Randy Kats, Andy Konwinski, Gu nho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica, thiab Mati Zaharia (2009): Saum toj in Huab Hua b : Huab Cave information Berkeley Cave Pom Hivanat Bahçesi, s.

[13] Shaw Jack: Dynasis Blue Paper: Cloudcomputing Public, Private and Hybrid(www.Dynasis.com, Değe rlendirme tarihi Kasım, 2013)

[14]M.Sai Charan Reddy, T.Dr. Venkata Satya Yaswanth, T. Gopal, L. Raji, K.Vijaya, "Data Leak Detectio n Using Cloud Computing", International Journal of Research in Engineering and Technology (IRJET) e-ISSN: 2395-0056, Vol: 06 Issue : 2019 -03-03.