International Journal of Research in Engineering Technology –– Volume 1 Issue 6, Sep - Oct 2016

RESEARCH ARTICLE OPEN ACCESS

A novel Technique for Data Leakage Detection in Cloud Using Fake Objects

V.Vanitrha[1], K.Mythili[2]

[1,2] (Dept of CSA, SCSVMV University Enathur, Kanchipuram)

Abstract:

With virtualization being the buzz word in the business world, organizations rely on distributed computing for various reasons. Cloud computing trend has set many standards for resource allocation due to which organizations have gradually started to migrate to cloud environment. Though there are many pros of cloud computing, security of data is a serious disadvantage. Cloud service providers can guarantee security through their administration to certain extent. But again malicious leakage of data happens on regular basis. Though many strategies have been devised previously for detection and prevention of data leakage, each have their own set of drawbacks. Traditionally watermarking is used for data leakage detection. Watermarking is a process of embedding unique code in the data. Ill-willed and tech savvy agents easily decode such watermarked data and this leads to huge problems. In this research an innovative technique for data leakage detection using fake object and OTP generation for secured distribution is proposed. A data distributor has given sensitive data to a set of allegedly trusted agents (third parties). Some of the data is leaked and found in an illegal place (e.g., on the internet or some unauthorized person's laptop). The distributor must assess the likelihood that the leaked data was gathered from a lone person or more agents, as opposed to having been independently gathered by other means. In the proposed system fake objects are injected in to the data file to improve the chances of identifying the leaker. This technique doesn't rely on alterations of the released data (e.g., *watermarks*). The proposed system is advantageous in many aspects like immediate leakage detection and prevention of further damage being caused to the organization.

Keywords- **Data leakage, detection, fake objects, prevention.OTP.**

I. INTRODUCTION

Data leakage is defined as the uncalculated or premeditated distribution of confidential data to unauthorized third parties. Sensitive data of companies and organizations include employee information, business strategies, management details and other information depending on the size and type of industry. In course of doing business, sensitive data is shared among third parties such as employees working freelance(e.g., on laptops),business partners and customers at various levels. This increases the risk of confidential data being let out unknowingly into other hands. Whether caused by malicious purpose, or an unintentional mistake, by an insider or outsider, exposed sensitive information can seriously hurt an organization either directly or indirectly. Many organizations are moving their business environment into cloud as it provides many benefits like setting up of virtual office, connecting to the business anywhere, anytime. But Security of data is of major concern in cloud. In this

research paper, certain strategies to overcome the data leakage issue existing in today's business domain in cloud computing environment are proposed and implemented.

a) ***Cloud Computing***

Cloud computing is the distributed networking of computers over internet. A number of hosted services are provided in cloud environment which makes it easy for organizations to concentrate on business rather than in-house resource. Cloud computing enables `companies to consume computational resources as a utility.

Cloud computing provides several advantages for businesses and end users. Three of the main pros of cloud computing includes:

- **Self–Service Provisioning**: End users can scale up computing resources according to necessity.
- **Elasticity:** Companies can increase their utilization as computing needs increase and then scale down as the requirement decreases.
- **Pay per use:** The computing resource used by the users is measured at a granular level which leads to paying only for whatever service is availed.

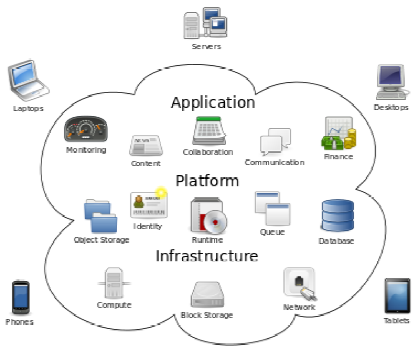


Fig 1: Cloud Computing

II. LITERATURE REVIEW

Throughout the recent past years many systematic research works has been under taken to develop efficient methods to detect data leakage in current virtual world. Some of the prominent works are discussed here. K.Mythili, S.Rajalakshmi and D.Vidya[1], have used identity services to detect data leakage in cloud computing using identity services", Data leakage detection [2] by SandipA.Kale prof S.V.Kulkarni, 2012 introduces subdued technique for detecting data leakage of set of objects. ArchanaU.Bhosale et al,have proposed data allocation strategies that perk up the probability of identifying leakages [3]. Guilt Assessment and Fake Data Allocation Strategies for Traitor Detection by G.N Sowjanya et al [4],have proposed data distribution algorithm and have used Brown Robinson that fixes a tolerance € > 0 and initializes a certain quantity of fake tuples for each agent instead of any random agent.

K.Manojkumar et al [5],in their work have developed a sophisticated system to

confront Data leakage detection employing various techniques like watermarking and forged data.[6]Data Leakage Identification and Blocking Fake Agents Pattern Discovery Algorithm by Karthik R et al[7],proposes data allocation strategies which minimizes the sum objective,leading to increase in the chance of data leaker identification.[8]huDanfeng(Daphne)Yao proposes a novel fuzzy fingerprint framework and algorithm to realize privacy-preserving data-leak detection. Parvathi Maheswari T [9] has probably proposed the idea of encryption of data and fake records using Secure Hash Algorithm(SHA).Further trace log has been used to detect the guilty agent.PriyaWalnuj et al [10],in their paper have proposed a data leakage detection technique which uses the SRF (Shortest request first) algorithm and have implemented a system called data watcher to find out the guilty agent.

III. ARCHITECTURE OF THE PROPOSED SYSTEM

This research paper presents a data leakage detection system using various tracking strategies like OTP generation for file access and fake object insertion, which assess the likelihood that the leaked data came from one or more agents. For safe transactions, allowing only authorized users to access sensitive data through access control policies to prevent data leakage by sharing information only with trusted parties and the data is detected from leaking by means of adding fake object in the data set, which improves probability of identifying leakages in the system.

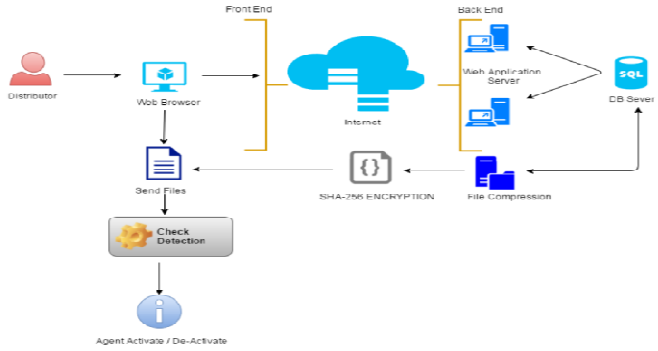


Fig 2: Architecture of the proposed system for Data Leakage Detection using Fake Objects.

The proposed architecture consists of five layers:

- Administrator layer/ Data Distributor layer
- Agent creation layer
- Data Allocation layer (File encryption and Compression is done)
- Leakage detection layer
- Activate/Deactivate agent layer

a) Implementation Strategy

File allocation process:

- Since the data is going to exist in the cloud environment, in the proposed system Google Snappy compression algorithm is used for file allocation and storage.
- SHA256 Encryption algorithm is used for encrypting the data

before transmission on the web for security purpose.

Leak Detection:

- File level – a Fake object is hid in the original data file, which will help in identifying the leaker in case of Leakage.
- Link level: Combination of user based Web Id and OTP are used to find the Data Leak.

4. METHODOLOGY OF THE PROPOSED SYSTEM

a) *Problem Definition*

A distributor owns a set of confidential data objects T ={t1,tm*}*. The distributor, due to business need wants to share some of the data with a group of agents *U*1*, U2…,Un* but does not want the objects be leaked to other unauthorized parties. An agent *Ui* receives a set of data *Ri* which belongs to *T* on request. The data objects in *T* could be of any type and size. After giving data to agents, the distributor finds that a set *S* of *T* has leaked. The third party who is in the possession of the leaked data is called the target. The third party may manipulate the data possessed by him and cause damage to the organization in many ways. Since the agents *U*1*, U2,…, Un*, have some of the records, it is reasonable to suspect them for leaking the data. However, the agents can deny the charge and claim that they are innocent, and that the target through other means obtained the S data.

b) *Implementation of the Proposed System*

- Login Module

The login module is the gateway for the administrator to gain access in to the implemented system. The administrator is asked to enter the user name and password. The system starts the authentication process. On successful authentication, the user can gain access into the system else a message is displayed saying "Authentication Failed".

- Create Agent Module

This module is used to create a new agent. The distributor can add a new agent as per need. The agent details are stored in the database in SQL Server, which is the backend source in the proposed system.

- Add File Module

In this module the distributor selects the agent and the file to be sent to the particular agent. The file selected is compressed using Snappy algorithm and encrypted before sending to the agent.

- Check Leaker Module

This module traces the agent name hidden in the file chosen for tracing the hidden fake object. The result is the agent name, responsible for leaking the data.

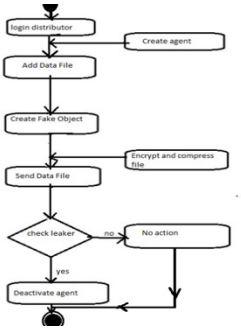
- Agent Activate/Deactivate Module

This module handles the action to be taken on the agent found

guilty. Once a agent is found to be guilty, he can be deactivated and any future business transaction with him is stopped. The activation of the guilty agent can be done only after the distributor is convinced by the agent's explanation.

5. PERFORMANCE EVALUATION OF THE IMPLEMENTED SYSTEM

a) Activity Diagram For Process Flow



b) *Test Case for the Implemented System*

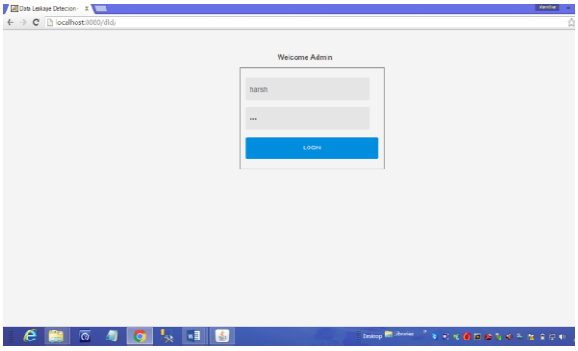


Fig 3: Administrator Log In Page

In case the login username and password entered is incorrect, authentication fails and the following window appears.

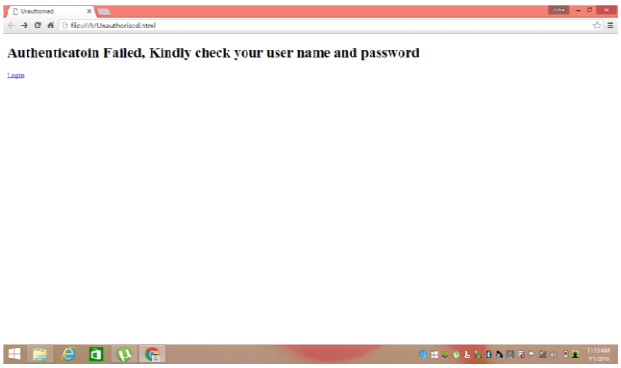


Fig 4: Window panel for "Authentication Failed" message display

On successful authentication the admin enters into the home page as shown below.

Fig 5: Home page

The admin can register new agents, assign data files to the agents, check the leaker agent detail in case of leakage detection and finally activate or deactivate leaker in course of action taken against the guilty agent.

6.CONCLUSION AND FUTURE WORK

a) Conclusion

In this research study it is concluded that the data leakage detection model proposed is very robust as compared to the existing watermarking model. The security of data during its distribution or transmission is taken care off and leakage is traced. Thus, using this model security as well as tracking system is developed and implemented. Watermarking can just provide data security using various algorithms through encryption, whereas this model provides security plus detection technique and also file encryption for compression purpose which will be handy in cloud storage. This model is comparatively simple, but it captures the essential tradeoffs. The algorithms devised implements a variety of data allocation strategies that can improve the distributor's likelihood of identifying a Leaker . It is shown that distributing objects prudently can make a significant difference in identifying guilty agents, especially in cases where there is huge overlap in the data that agents must receive. The detection of data leakage in the proposed model is limited to data file alone. In future the same strategy can be applied to other data sources.

b) Future Work

- Future work includes the study of agent guilt models that capture leakage scenarios like Email leakage, Database leakage etc.
- Finding the guilt agent with proper MAC address via networking concepts.
- Sending notification to distributor via android application in case of leakage.

REFERENCES

[1] K.Mythili, S.Rajalakshmi and D.Vidya,"*Data leakage detection in cloud computing using identity services*",International Journal of Computer Science and Engineering,Volume-04,Issue-04,Pages (59-63),Apr-2016,E-ISSN:2347-2693.

[2] Sandip A. Kale, Prof. S.V.Kulkarni ,"*Data Leakage Detection*",IJARCCE,Vol 1,Issue 9,Nov 2012.

[3] ArchanaU.Bhosale,Prof .Vharkate M.N, Aparna U.Bhosale,"*A Stydy of Data Allocation Problem for Guilt Model Assessment in Data Leakage Detection Using Cloud computing*",IJSR,Vol 3,Issue 4,Apr 2013.

[4] G.N Sowjanya,Mrs N.Nagasubrahmanyeswari,"*Guilt Assessment and Fake Data Allocation Strategies for Traitor Detection*",IRACST,Vol 2,Issue 3,June 2012.

[5]K.ManojKumar,G.Shubang,G.Rajesh Chandra."*Data Leakage Detection for Cloud-Based Storage Sytsems*",IJAER,vol 8,Issue V,Nov 2014.

[6]Karthik .R,Ramkumar .s,Sundaram .K ,"*Data Leakage Identification And Blocking Fake Agents Using Pattern Discovery Algorithm*",IJIRCCE,Vol 2,Issue 9,Sep 2014.

[7] BhagwanD.Thorat,P.RDevale,"*A Model to Find The Agent Who Is Responsible For Data Leakage*",IJRET ,Vol 3,Issue 11,Nov 2014.

[8] SonamChugh,Sateesh Kumar Peddoju."*Access Control Based Data Security in Cloud Computing*",IJERA,vol 2,Issue 3,May – Jun 2012.

[9] XiaokiShu ."*Data Leak Detection As a Service :Challenges and Solutions.*

[10] ParvathiMaheswariT."*Prevention and Data Leakage Detection using Trace Log*",IJESRT,July 2014.

A
s
s
i
s
t
a
n