International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 03 | Mar 2022 www.irjet.net p-ISSN: 2395-0072

Data Outflow Detection with Guilty Agent Tracking

Kumara Huliyurdurga[1], Sakshi Jadhav[2], Rutuja Patil[3], Deepti Lawand[4]

[1]*Student, Dept. of IT Engineering, Pillai College of Engineering New Panvel, Maharashtra, India.*
[2]*Student, Dept. of IT Engineering, Pillai College of Engineering New Panvel, Maharashtra, India.*
[3]*Student, Dept. of IT Engineering, Pillai College of Engineering New Panvel, Maharashtra, India.*
[4]*Professor, Dept. of Computer Engineering, Pillai College of Engineering New Panvel, Maharashtra, India.*

Abstract -*Data leakage can be defined as accidental leakage of data by a security illiterate or a deliberate attempt by a guilty employee within the firm. When a sender sends confidential data to a third party, for example, a company may have partnerships with other companies that require sharing customer data, there are chances that sensitive data gets leaked accidentally or purposely by an attacker. The research paper focuses on detecting whether the owner's sensitive data has been leaked by trusted third parties, and if possible identifying the agent that leaked the data. If the unauthorized user is identified, then the owner of the data will get an alert and can stop distributing their data with the third parties and send warning alert to that third party that if the same thing happens again by them then the owner of the data can take action against them and even can legally penalize them for data Cases of leakage. This will further improve the chances of finding the guilty person who leaked the data.*

***Key Words*: Data leakage, Distributor, Agent, IP, Warn, Alert.**

1.INTRODUCTION

Data leakage, also known as data outflow, is the unauthorized communication of data from within an organization to an external destination or recipient, such as a third party. It's essentially delayed data theft, and it's a big concern for data security, with serious consequences for any firm, regardless of size or industry. This is a threat that any firm would wish to avoid, from decreased revenue to a tarnished reputation. Data can be leaked in a variety of ways, and it's crucial to remember that the problem might originate from either an external or internal source.

The Accidental/Unintentional Breach:

"Unauthorized" facts leakage does now no longer always imply supposed or malicious. For example, an employee may by mistake choose the wrong recipient when sending an email, which contains confidential data. Unfortunately, these unintentional data leakage also result in the same penalties and reputational damage, as they do not lighten the legal responsibilities.

The Ill-Intentioned Employee:

This type of leakage is usually conducted manually by an authorized employee with access to company systems, who are paid in lump sum by the attackers to do this theft. Whether the threat is present inside your organization or externally, it is necessary that one must be aware of the risk.

Electronic Communications with Malicious Intent:

Malware is frequently used to target various mediums, and it has a high success rate. A cybercriminal, for example, may easily fake a growing business email account and request important data be provided to them. The information, which could contain financial data or critical pricing information, would be sent without the user's knowledge.[8]

Today's world is reliant on information transfer, or the flow of data from one person to another. This system is required for the following reasons:

- Protection of complex and important data,
- For Finding the guilty agent,
- And to keep the confidential data secure.

2. LITERATURE SURVEY

2.1 Data Leakage Detection And Security In Cloud Computing Environment, November 2019: The paper's proposed model addresses the issue of providing compelling proof against the data leaker and the data he or she has released. This is implemented with the four modules namely the data allocation module, fake object module, optimization module & data distribution module.

Technique used in this paper:

Author Sanidhya Gupta utilizes the SHA technique, which stands for Secure Hashing Algorithm, to compress and encrypt data supplied as input and produce hash or hash values, which appear to be random values. These seemingly random values are actually the input data in an encrypted or coded form. Hash values of data are easier

for computers to implement than the original data because hashes make it easy for the computer to do various operations or calculations over files and data strings. This hashing algorithm is deterministic, which means it always returns the same result for the same input value.

2.2 A Survey: Data Leakage Detection Techniques, August 2018:The proposed model in this paper deals with problems such as when important information goes to unapproved hands it will prompt the direct and indirect loss of particular industry then information leakage is an outcome in vulnerability so to overcome it there must be an efficient and effective system to avoid and protect authorized information.

Technique used in this paper:

The sampling algorithm used by author K. S. Wagh was based on context-aware selection. It produces predictable and subsequence-preserving outputs. In Dynamic Programming, the Recurrence Relation technique is used to work on a compact sample list. This alignment-based technique uses order-aware comparison to detect data leaks and has a high tolerance for pattern deviations.

It explains the domain system's flow as well as the exact implementation techniques for data input, data preprocessing, and other relevant tasks. The results reveal that this method has performed admirably in this domain.

2.3 Data Leakage Detection with K-Anonymity Algorithm, 2016: This paper focuses on identifying the agent that released the distributor's critical information, and it is possible to identify the agent who leaked the data. Sensitive data is created using the K-anonymity technique, which hides the data set. The original datasets will not be accessible to third parties.

Technique used in this paper:

Author B. M. Patil used the K-anonymity algorithm to eliminate the watermarking technique and create fake objects, allocating data, using the guilt probability model to find the guilty agent, and creating sensitive data. The K-anonymity algorithm provides simple and effective approaches to protect individuals' private information by only releasing k anonymous views of a data set. The k-anonymity approach has grown in popularity as a result of its positive outcomes.

2.4 Data Leakage Detection, March 2016: The author is creating a system for detecting leaked data and the agent who is accountable for data leakage in this research paper. In some circumstances, "realistic but fake" data records can be utilized to increase the chances of discovering data leakage and identifying the unauthorized individual who leaked the data.

Technique used in this paper:

For finding the guilty agent, the author used a variety of allocation strategies, including six algorithms such as s-overlap, s-random, and others on explicit and sample data requests. There are three modules in this paper: an agent module, a distributor module, and an admin module. The results reveal that this method has performed admirably in this domain.

2.5 Fast Detection of Transformed Data Leaks, March 2016: For searching complex data-leakage patterns, the author employs a sequence alignment method. This technique is used to identify both long and essential data patterns. This recognition is combined with a sampling technique that enables for comparison of two independently tested successions. This structure achieves a high level of precision in detecting altered leakage.

Technique used in this paper:
For detecting complex data-leak patterns, sequence alignment techniques were used. Alignment of sampling algorithms as well as sampling-oblivious algorithms are comparable.

3. Proposed Work

In order to achieve better domain results, researchers came up with new techniques to build Data leakage detection systems, which seek to inherit advantages and eliminate the disadvantages of the existing systems.

3.1 System Architecture

This system allows agents/users to request data from the distributor, access the allowed data request and get messages from the distributor. The distributor can check the data request, allow/deny requests, check the leaked data, track the responsible leaker and send messages to an agent/user.

Distributor Module: The company's main authorized user is the distributor. It accepts the agent's data requests and determines whether they are sample or explicit. It stores sensitive information in a database and distributes it to agents based on their requests by inserting a bogus object into the original data. If the distributor receives notification that an agent has leaked data or is guilty, it can send a warning message to the offending agent and take action against him. The password is also changed by the distributor. It has the ability to delete any agent by selecting the delete agent option.The Distributor can also track the guilty agents and take necessary action. From the guilty agents section, the distributor can view where the leakage has occurred with the details, and has the authority to report and remove the agent and also send alerts to the agent to warn them about the identified

leakage.

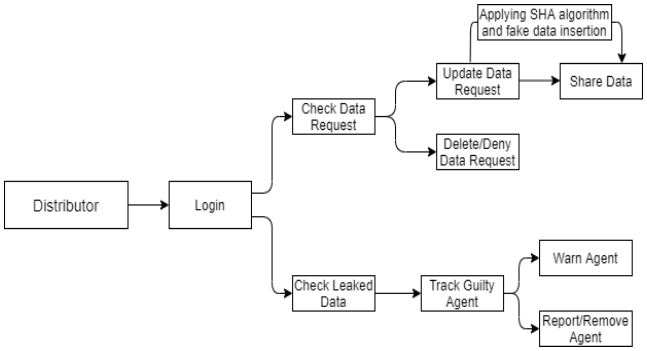

Fig. 3.1 Overview of the system: Distributor's side

Agent Module: By entering his or her user ID and password, the agent can log in. Agents can make data requests to the distributor. Once the distributor accepts the request of the agent, the agent can access the data. If the agent shares the data by sharing his key, he/she can be tracked as a guilty agent. The agent can also receive messages from the distributor.

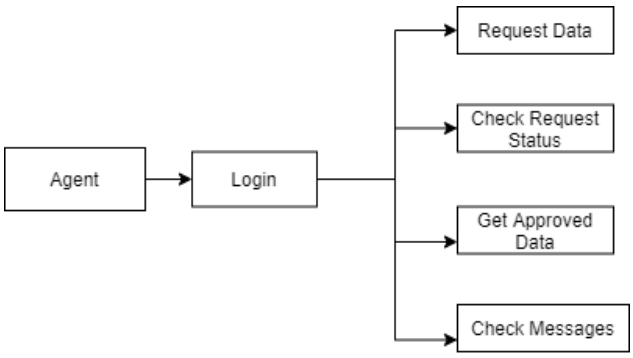

Fig. 3.2 Overview of the system: Agent's side

3.2 Technique used

Ip addressing:

An IP address works as an identifier for a device on the internet or within a local network. IP addresses are the identifiers that allow data to be exchanged between devices on a network they are connected on: they contain location information and make devices communicable. Different computers, routers, and webpages all need to be distinguished on the internet. IP addresses are a necessary aspect of how the internet functions. We are using this technique to check whether the files shared within an organization are accessed by someone outside the organization. Or accessed at two different locations by the same profile.

Random Key Generation method:

In the system each file which is present or the file distributor uploads needs to be assigned with a key. This key will be the essential element which the agent will need to access the files. The key will act as a major feature in securing the files/reducing the outflow of data. To assign the files with the key, a python function is used which randomly produces a key with a certain combination of digits and alphabets. When the distributor adds a file, the function instantly generates a key and assigns the file with it. This key is stored in the database which will be used to authorize the agent to access the file. If the agent enters a wrong key while accessing the file, the agent will be denied the access. This helps in securing the files in a simple yet reliable manner.

4. CONCLUSIONS

Thus, on the basis of the literature survey and research on the topic the idea is implemented and works efficiently for secure sharing of data and tracking the guilty agent in an organization.

The constraints of checking the data leakage is the key, the agent id and the location. The data leakage occurring is covered through sharing of keys to access the data without authorization, accessing the data from a device which is not on the same network as of the organization. IP addressing is used to check if the agent is accessing the data from some other location.

If the agent is found as guilty the distributor can warn him or remove the agent from the system. The warning is sent in the form of email as well as an alert to the agent. If the distributor removes the agent, the agent's account will be deleted from the system and he/she will no longer be able to log in or access his/her account on the system.

ACKNOWLEDGEMENT

We express our deepest gratitude towards our project guide, Prof. Deepti Lawand for her valuable and timely advice during the various phases in this research paper and for motivating us to do better. We would also like to thank our H.O.D of Information Technology department Dr. Satish Kumar Varma for this opportunity and for providing us with enough resources to complete our tasks.

We would also like to thank Our Principal "Dr. Sandeep Joshi" for providing us the best environment possible. Finally we would like to thank everyone who has helped us directly or indirectly in the research paper.

REFERENCES

[1] S. Visnu Dharsini, Mrinal Pramanik, Sanidhya Gupta, Saurav Pahadiya, Data Leakage Detection And Security In Cloud Computing Environment International Journal of Scientific & Technology Research Vol 8, Issue 11, November 2019.

[2] K. S. Wagh, A Survey: Data Leakage Detection Techniques, International Journal of Electrical and Computer Engineering (IJECE), Vol. 8, No. 4, August 2018.

[3] Wakhare Yashwant R, B.M.Patil, Data leakage detection with K-Anonymity Algorithm, International journal of computer Science and Information Technologies, Vol.7(4),2016.

[4] Xiaokui Shu, Jing Zhang, Danfeng (Daphne) Yao, Wu-Chun Feng, IEEE, Vol 11, NO. 3, MARCH 2016.

[5] Ghagare Mahesh, Yadav Sujit, Kamble Snehal, Nangare Jairaj, Shewale Ramchandra, Data Leakage Detection International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 03, Mar-2016.

[6] Panagiotis Papadimitriou, Hector Garcia-Molina, IEEE, VOL. 22, NO. 3, MARCH 2010.

[7]Xiaokui Shu, Danfeng Yao, Elisa Bertino, IEEE Transactions On Information Forensics and Security, Vol. 10, No. 5, May 2015.

[8]www.forcepoint.com/cyber-edu/data-leakage

[9]https://www.kaspersky.com/resource-center/definitions/what-is-an-ip-address

BIOGRAPHIES

Mr. Kumara Huliyurdurga
Student of BE-IT
Pillai College of Engineering

Miss. Sakshi Jadhav
Student of BE-IT
Pillai College of Engineering

Miss. Rutuja Patil
Student of BE-IT
Pillai College of Engineering

Prof. Deepti Lawand
Professor
Pillai College of Engineering