International Journal of Engineering &Technology, 7 (3.12) (2018) 366-368

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Data Leakage Prevention and Detection System

M. Rajasekaran[1], Amisha Gupta[2], Padmini Sharma[3]

[1]Department of Computer Science and Engineering, SRM University, Kattankulathur, Tamilnadu, India.
**Corresponding Author Email: [1]smrajamani1981@gmail.com ,[2]rashigupta96@gmail.com, [3]bornville13@gmail.com*

Abstract

Our system includes a data distributor who sends some important information to some third parties, called agents in our system. Some data gets leaked and found in an unauthorized place (e.g., on the internet or somebody's PC). The distributor must be able to detect that the data got leaked and came from one or more agents. Our system detects the guilty agent, who has leaked the data to untrusted agents. For identification of leaked data in existing system uses watermarking technique. However, it has a deficiency that watermarking data can be modified or changed. So, In proposed system another technique is used for improving chances of detecting guilty agent is MAC (Media Access Control) address with some more prevention strategies to increase the security of the system.

Keywords: *Data leakage, Data prevention, MAC address, Guilty agent, AES algorithm*

1. Introduction

There are many websites that publish information and provide access of information on the internet. All that websites has many different web application programs that contains potential information or important data that has to be protected. In many organizations, business situations and company outsource its data to the other company or organization or agents, all these agents are recognised as trusted third party agents. The data distributor gives the confidential data to the trusted third party agents and there are chances that any of that authorised agent can leak the potential data. Data leakage is defined as unintentional or accidental distribution of private or sensitive data to an unauthorized party. Detection of leaked data and guilty agent is major challenge in many industries.

If any agent will leak the confidential data, it leads a greatest financial loss. So, there is a need to provide protection and privacy to the data and identify the faulty agents, which are from the trusted agents, who leak the data to untrusted person. For improving the possibilities of identifying guilty agents, another new method has been proposed. This technique will use the MAC address and AES algorithm. The MAC address improve the possibility of identifying guilty agents.

2. Literature Survey

There are many works that have been proposed by various researchers in the past [10], [11], [12], [13]. Among them, Panagiotis Papadimtriou and Hector Garcia-Molina designed "DATA LEAKAGE DETECTION" In there model it is explained that there is a chance to assess the likelihood that a third party agent is responsible for a leak ,based on the overlapping of his data with the leaked data and the data from the other agents, and based on the probability that objects can be "guessed" by other techniques.

The algorithms presented implements a variety of data distribution techniques that can improve the distributer's possibilities of

detecting the leaker. It is also observed that distributing objects thoughtfully can make a important difference in detecting guilty agents, especially in scenarios where there is large similarity in the information that agents must get.

Simon Liu and Rick Kuhn designed "Data Loss Prevention" in which types of losses discussed are Leakage in which important data is no further under the organization's control and disappearance or loss in which a accurate data copy is no further available to the organization.

Data loss prevention technologies were introduced for govt. and industry requirements and intellectual property protection which aimed on preventing various sensitive information from going out of the organisation's private confines. Loss Modes are Data at rest, Data at the end point and Data in motion Best practices like, Prioritize loss modes, Protect without disruption, flexible and modular architecture were introduced.

Maulik Bhagat and Prof. G.B. Jethava proposed "Detection of guilty agent using encryption algorithm and mac address" in which desired objective of distributor is to recognise the guilty agent who leaked the data.

This system identifies the guilty agent using mac address and AES Encryption algorithm without adding fake object and without calculating guilt probability.

AES algorithm encrypts the potential content so unauthorized agent is unable to use the confidential information.

MAC address is used to identify the guilty agent, so this system identifies the guilty agent by providing the confidentiality of data.

3. Proposed Work

In this paper, a new ensemble approach is proposed for effective prevention of the data in the system. Our goal is to detect if the data has been leaked by any of the trusted agent and also to get the name of the guilty agent who leaked the data. Using the technique

of incorporating MAC address improves the possibility of identifying guilty agents.
Also, to prevent the sensitive content from getting leaked we tend to use prevention strategies.

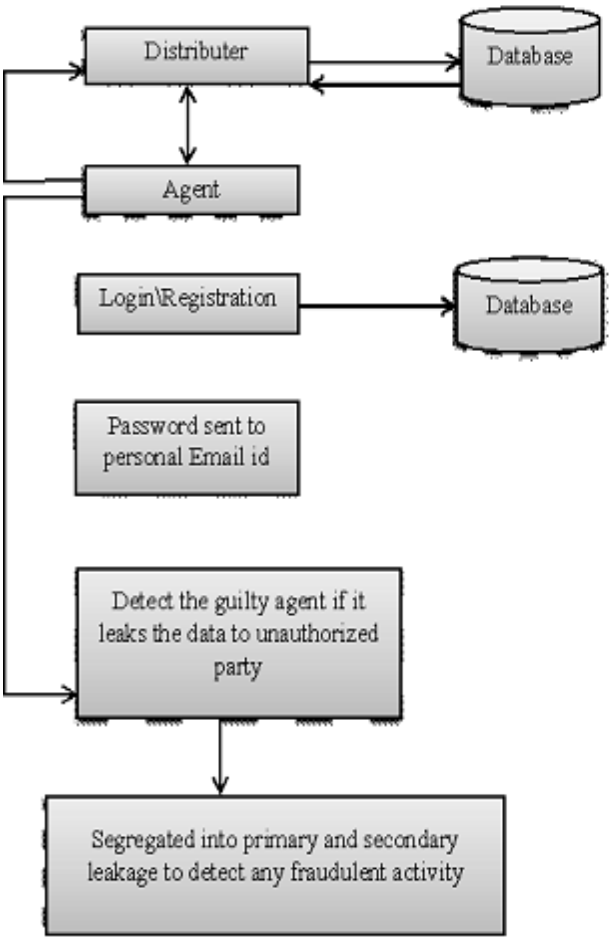
Techniques Used:

Prevention Strategy 1:

Using Password Encryption
Step 1: The authorized agent should fill the registration form and provide a password.
Step 2: After clicking on 'register', the computer will take all the details and it will get stored in SQL Database.
Step 3: The encrypted form of the password entered will be displayed in the database, using AES Algorithm. So, only an authenticated agent can send a data request with the required credentials.
Step 4: As the agent provides its email id and a password, a copy of password will be sent to the provided email id, to prevent any fraudulent activity.

Prevention Strategy 2:

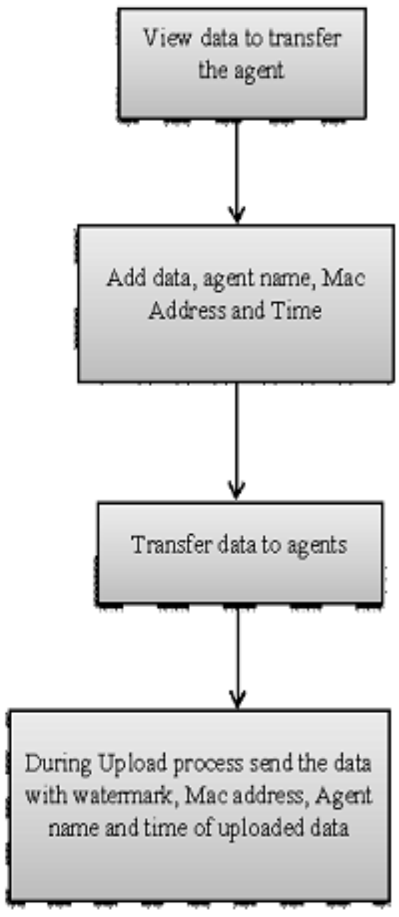
Primary and Secondary Leakage



Step 1: The authorised distributor, can check the Data Leakage Details, if there is any, after signing up.
Step 2: It is segregated into Primary and Secondary Leakage.
Step 3: Primary leakage will display the guilty agent's name, file name, MAC address, date and data receiver.
Step 4: Secondary leakage has two columns, namely 'Distributor send file to agent' and 'Agent Received file from distributor'. It shows the no of files that the distributor has sent and the no of times the agent has downloaded it.
Step 5: If the no. of times the file has been downloaded exceeds the no of files that were actually sent by the distributor, then that agent can be suspected of data leakage.

4. Result and Discussion

The proposed system has been implemented using JavaScript. We have created our own dataset by registering the agent in the proposed system. The architecture diagram is shown below: agent. So, our project aims at identifying the guilty agent by providing the confidentiality of data.



5. Conclusion

In this paper, it is explained that in an organization, most of their potential information is shared to the trusted agents. But, they cannot assure that the agent is loyal. So the desired objective of the sender of the data is to detect the guilty agent who leaks the data. So, this data leakage prevention and detection system can identify the guilty agent using MAC address and Primary and secondary Leakage. The AES algorithm encrypts the potential information. Hence, unauthorized agent is unable to obtain the confidential data. And MAC address is used to identify the guilty

References

[1] Panagiotis Papadimitriou, Hector Garcia-Molina, A Model for Data Leakage Detection Stanford University.

[2] Panagiotis Papadimitriou, Hector Garcia-Molina, Data Leakage Detection IEEE Transaction on Knowledge and Data Engineering, Vol-23,No 1, January 2011 pp. 51-63

[3] Ajay Kumar, Ankit Goyal, Ashwini Kumar, Navneet Kumar Chaudhary, Sowmya Kamath ,Comparative Evaluation of Algorithms for Effective Data Leakage Detection, IEEE Conference on Information and Communication Technologies (ICT 2013), pp. 177-182.

[4] Anush Koneru, G.Siva Nageswara Rao, J. Venkata Rao, Data Leakage Detection Using Encrypted Fake Objects, International Journal of P2P Network Trends and Technology Vol-3 Issue-2 2013 pp. 104-110

[5] B. Sruthi Patil, Mrs. M. L. Prasanthi, Modern Approaches for Detecting Data Leakage Problem, , International Journal Of Engineering And Computer Science, Vol-2, Issue-2, Feb 2013 pp. 395-399.

[6] Jaymala Chavan, Priyanka Desai, Relational Data Leakage Detection using Fake Object and Allocation Strategies, International Journal of Computer Applications, Vol-80, No.16 , October 2013 pp. 15-21.

[7] Rudragouda G Patil, Development of Data leakage Detection Using Data Allocation Strategies, International Journal of Computer Application in Engineering Sciences Vol-1, Issue-2, June 2011 pp. 197200.

[8] Sandip A. Kale, Prof. S. V. Kulkarni, Data Leakage Detection, International Journal of Advanced Research in Computer and Communication Engineering Vol-1, Issue-9, November 2012 pp. 668-678

[9] Jagna Ajay Kumar, K. Rajani Devi, An Efficient And Robust Model For Data Leakage Detection System, Journal of Global Research in computer Science, Vol-3, No-6, June 2012 pp. 91-95

[10] Chapter 2 Data Leakage http://www.springer.com/978-1-4614-2052-1.