LIME:PRIVACY PRESERVING TECHNIQUE FOR DATAMINING SYSTEMS

[1]SHAIK.MOHASEENA, DEPT OF COMPUTER SCIENCE AND ENGINEERING, JNTUA, ANANTHPUR, INDIA
[2]DR.B.LALITHA, DEPT OF COMPUTER SCIENCE AND ENGINEERING, JNTUA, ANANTHPUR, INDIA

Abstract : In the existing system we have seen that data security is not sufficient in organizations it's not because of the system security errors but also from the persons that are presente in the firm. This paper proposes a method called insider collision attack which operates with the kernal values and which is very strong to defend the security attacks but in here there is a chance to occur a known-plain text attack and it is limited to non-homomorphic encryption method thus we use technologies which are very effective for finding the data leakage and data security problems when we are sending data through a network there may be a chance to some unauthorized users accesing the data for that here we use a data transfer protocol and data lineage framework for finding the data leakage situations while transmitting the data in the outsourcing and in the network transmission.

Keywords:Data security,known plain text attack,data leakage,unauthroized users,data transfer protocol,Lime

INTRODUCTION

Data **s**ecurity is the important concern for the organizations for example if we consider a bank confidential the data of one customer leaks such as PIN number then it leads to so many issues and because of that customer may loss the money and that bank losses the customer's trust and may loss the goodwill and market share in the industry,to prevent from all these kind of things we have to control the data flow within the organization. Restricting data leakage is the solution for controlling unexpected data problems and providing security to the sensitive information of the organization from unauthorized or illegal persons.here sensitive data may refers to company internal processing documents and financial statements etc. Not only the companies are effected by the data leakage,the growth of social networks and smart phone usage makes the situation very worse.In social networks the users provide their personal information to third party appciations,if there are no customized protocols between users and the third party applications then uses personal information will be shared throughout the internet and some hacking companies.

Even with control access mechanisms if there exists a malicious authorized user he can dispose the sensitive data when he recieves that data. Primitives like encrypting the data offer encryption of the data only. But once malicious user receives and decrypts it the situation becomes tough to control user to miss use of the sensitive information.

What is the need of data lineage

1. In organizations majority of the users doesn't know about data security so we have to take care of data there.
2. when we are using new devices in our organization and the downloading the softwares there may be a chance to some malicious users to access the system data.
3. some employees in the company may become malicious and leakout the companies confidential data to the third parties.

LITERATURE SURVEY:

A computational model for watermark robustness[1]

As LIME may be a general model and may be applicable to all or any cases, we have a tendency to abstract knowledge the information kind and decision each data item document. There area unit three completely different roles which will be appointed to the concerned parties in LIME: information owner, information client and auditor. The data owner is accountable for the management of documents and also the client receives documents and may do some task victimization them. The auditor isn't concerned within the transfer of documents, he's solely invoked once a run happens then performs all steps that area unit necessary to spot the source. All of the mentioned roles will have multiple instantiations once our model is applied to a concrete setting. A key position in LIME is taken by the auditor. He's not concerned within the transfer. The need for formal security definitions of watermarking schemes provide suitable abstractions to analyse and prove the security of applications imprtent in watermarking schemes Moreover, the existing formal definitions for watermark security still suffer from conceptual deficts. In this paper we make the first essential steps towards an appropriate formal definition of watermark robustness.

Data Lineage in Malicious Environments[2]

Intentional or unintentional leakage of confidential data is undoubtedly one of the most severe security problem that organizations faceing. In this work, we present a generic data lineage framework Lime for data move across multiple entities that take two characteristic principal roles (i.e., owner and consumer). We using data lineage mechanism to identify duplicate entity, and identify the simplifying non-repudiation and honesty assumptions. We develop data transfer protocol between two entities within a malicious environment by building upon oblivious transfer, robust watermarking, and signature primitives.

A survey paper on data lineage in malicious environments[3]

A data distributor has given importent data to a set of trusted agents. Some of the data are leaked and found in an unjustified place. The distributor must assess the likelihood that the crevice data came from one or more agents, as opposed to having been individually gathered by other means. We propose data allocation strategies that improve the probability of identifying crevices. These methods do not build on alterations of the released data. In some cases, we can also implant "realistic but fake" data records to further improve our chances of detecting crevice and identifying the duplicate entity. While sending data over the network there is lots of illegitimate users trying to get useful information. There should be proper security provided to data which is send to network.

Chronology of data breaches[4]

The data breaches have been reported because the personal information avilable to unauthorized pers. Such as Social Security numbers, account numbers, and driver's license numbers. Some

breaches that do NOT expose such sensitive information have been included in order to underscore data breaches problems.

Related work

There are a lot of mechanisms that provide data access to authorized users and restricting unauthorized users to access the sensititve information through some control policies such mechanisms prevent the data leakage by providing data access to trustworthy employees.

LIME(Lineage in the malicious environment) is the one type of mechanism which can be used with any type of data for which watermarking schemes exist.Most water marking schemes are designed for multimedia files such as audio,image and video files.water marking techniques have been also developed for other data types such as relational data bases,text files and even android apps[1].the first two are especially used to apply LIME to user databases or medical records. for water marking of texts there are mainly two methods the first one embeds the information by changing the text appearance(like spaces and distance between lines)and the second one is reffered as language water marking and works on the sysmantic level of text instead of its appearance.In the robust watermarking technique the authors say rather than removing the existing information make changes to old information by adding some new information.

Trusted sender:

In case of a trusted sender or authorized user there is no need to take another security

mechanisms.the sender who is the owner of the document D,creates a water marking key k,embeds a triple $\sigma=(c_s,c_r,T)$ consisting of two parties identities and time stamp t,into D to create $Dw=w(D,\sigma,k)$, then sends Dw to the recipient who are the authorized persons requested for that document.

Untrusted sender:

In the case of untrusted sender we have to take additional security action to prevent sender from cheating.Here the sender divides the original document into n parts for each part he creates two different watermarking versions.he then transfers one of each of these versions to the recipient.The recipients combines the document with the parts that he received.

DataLineage Generation:

The auditor is the person who find the victim if any data leakge occurs in the system.follwing are the responsibilities of the auditor:

1. Auditor initially suspends the data sender.
2. Adds that suspect to the list of data lineage
3. Sends the leaked document to the suspect and verifies the detection keys k1 for the watermarks in the docuemt and the water mark σ.
4. If with key k1,σ cannot be detected then,the auditor outputs the lineage that last entry is resoponsible for the data leakage.
5. If the suspect is trusted the auditor checks that σ is of form (c_s,c_r,T) where is cs is the identitiy of the current suspect

Necessary data required for finding the data lineage:

Bothe the data senders and receivers has to provide the document related data like watermarking key σ and the keys of the document by using this information auditor finds the lineage from where that document has moved to in of the firm.

If we consider a organization as a sender or data owner who gives tasks to other companies which work as the consumer.there is a chance to send the companies sensitive information other companies,there is no chance to owner to trust that outsourced company.fingerprinting is the technique that is used for the transmission of the data. If we consider this transformation between the companies as a tree and if at anypoint of the outsourced company leaks the data the organization

invokes the auditor for finding the data leakage.Then the auditor follows the finger prints of the data transmission and adds all finger prints of users who participated in to the data lineage.finally the auditor find outs the responsible person for the data leakage By using LIME.

Online social Network:

If we consider social networking as the owner&users upload their personal data into online social network.The online social network uses all this data as a consumer scenario.third party applications has to acces the application in terms of the other users like fingerprinting scenario[8].the users give this information to the online social network which can relay that information to third party applications like fingerprinting.If any data leakage occurs auditor will find outs the party responsible for data leakage .

EXPERIMENTAL RESULTS :

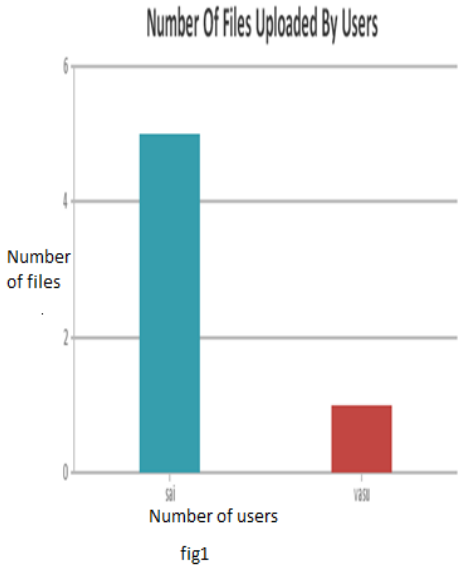

fig1

The experimental results shows that initially users upload the file after registering into the system,these files contain data like image this data will be encrypted and stored in the database, when the user searches for the data(search based on content) and if the searched data matches with the stored data then it will be displayed in encrypted form to the user.

After successful registration of the user user searches for the data then if any matching content is there that file will be displayed and it will be encrypted so that user has to send a request to the data owner for key when the user receives that key he can view and download the data. The above graphs fig1 show that the users data who uploaded data.

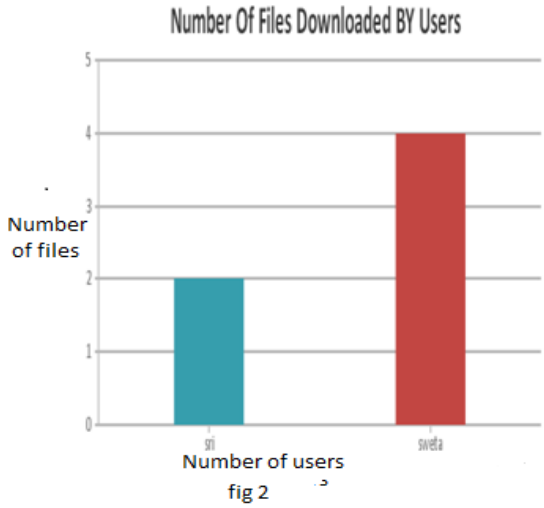

fig 2

fig2 shows downloaded data based on this auditor can find the data transmission and if any data leakage occurs there will be a chance of finding user or person the responsible for data leakage.

CONCLUSION:

Here we presented LIME model,to identify the data lineage that is happened in the organization application or system.Here we are using data transfer protocol for tranfering the data between the users the is called as fingerprinting and there are different users presented in this like data owner who gives the data and the consumer who access that data auditor who observes the data flow that is happening in the system if any data leakge occurs auditos has to find the responsibility party.

LIME is flexible as we are differentiating the senders as trusted sender and Untrusted sender when we are dealing with a trusted sender then we use simple protocols and untrusted snder requires more complicated protocol for the transmission of data.

REFERENCES

[1] A. Adelsbach, S. Katzenbeisser, and A.-R. Sadeghi, "A computational model for watermark robustness," in Information Hiding. Springer,2007, pp. 145–160.

[2] Michael Backes, Niklas Grimm, and Aniket Kate, "Data Lineage in Malicious Environments" 2015.

[3] Bhamare Ghanashyam,Desai Kiran,Khatal Supriya,Mane Vinod,Prof. Hirave K.S.,"a survey paper on data lineage in malicious environments, M28-2-4-10-2015

[4] Chronology of data breaches, http://www.privacyrights.org/data-breach.

[5] Data breach cost, http://www.symantec.com/about/news/release/article.jsp?prid=20110308 01.

[6] Privacy rights clearinghouse, http://www.privacyrights.org.

[7] Electronic Privacy Information Center (EPIC), http://epic.org, 1994.

[8] Facebook in Privacy Breach, http://online.wsj.com/article/SB10001424052702304772804575558484075236968.html.

[9] B.Lalitha Evaluating the Privacy of User Profiles in Personalized Information Systems ISSN: 2277 128X.