A Study on Determining Data Leakage Detection

Senha Sewale¹, Jyoti Navse², Punam Chougale³, Prof.D.M.Bhosale⁴

¹²³⁴Keystone College of Engineering, Hadapsar, Savitribai Phule Pune University.

Abstract: Now a days every company wants data security by the cybercrime and data is an important part of every company so that data loss and destruction is necessary. Huge data is being exchanged among multiple people and third parties at every moment in IT sectors. There are huge chances of data vulnerability, leakage or alteration during exchanging of data. So, to securely overcome these problems, a thorough study on web data leakage detection and prevention system is needed. This survey talks about the idea, causes and techniques to find out the data leakage. Businesses analyze the facts and figures to mine the raw data for fetching the useful information. Businesses use this information in order to generate and implement revenue at every mile stone. Thus, along with data accessibility and availability along with data security is also vital for consideration.

Keywords: Water Marking, Fraud or Guilty Agent, Data Allocation, Data Leakage Detection.

I. INTRODUCTION

The importance of the data is undefined, so it should not be leaked or copied. Though there are number of techniques developed for the data security by using different encryption algorithms, there is a major issue of the integrity of the users of those systems. It is very difficult for any system administrator to find the data leaker among the system users. It creates a lot many ethical issues in the working area of the office Therefore Data leakage is the big challenge of the enterprise. The paper is modified as follows. Data classification, Data identification, Organization of Prevention Techniques, Improved security (measuring outcome). Literature review introduces several techniques used in same systems along with their shortcomings. Now coming section shows the completeness.

II. LITERATURE REVIEW

Data security is difficult for most businesses and even home or office computer users. Client information data, payment details or transaction, personal data, bank account information - all of this data can be critical to replace and more dangerous if it falls into the wrong hands. The possibility of data lost due to disasters such as a flood, earthquake or fire is destroying, but losing it to hackers or a malware infection can have much greater consequences. However, some of this information isn't intended to give up the system. This data can be access by the unauthorized system and could lead to various problems for the larger corporation or even the personal home user. Having your bank account information stolen is just as damaging as the system administrator who was just robbed for the client data in their database.

INFORMATION IS CLASSIFIED AS FOLLOWS:

- Most critical
  - Used oftentimes
  - Instantly availability
  - Significant and immediate financial impact

- Business
  - On a regularly basis used
  - Reasonably available
  - Significant long-term financial impact

- Essential
  - Used periodically
  - Available within defined timeframe
  - Potential long-term financial impact

- Consequential
  - From time to time used
  - Available within extensive timeframe
  - Likely compliance problem.

- Non-critical
  - Rarely used
  - Limited accessibility
  - Improbable financial affect
  - Potential conformity impact

Classification of above data are as follows:

Classification of Most critical

- Used oftentimes
- Instantly availability
- Significant and immediate financial impact

Classification of Business data

- On a regularly basis used
- Reasonably available
- Significant long-term financial impact

Classification of Essential data

- Used periodically
- Available within defined timeframe
- Potential long-term financial impact

Classification of Consequential data

- From time to time used
- Available within extensive timeframe
- Likely compliance problem.

Classification of Non critical data

- Rarely used
- Limited accessibility
- Improbable financial affect
- Potential conformity impact
other thing which comes into our minds is digital used. Since we are discussing about crime and the allocation, cryptography, and steganography can be altered in any way and this is the reason data information. These information should not be the main reason why they need to safe the have strategies to face off these criminals and this is fire like leakage, besides the police section also mentioned. These records are more vulnerable to criminal and the data of the crime they have forbid data from leakage.

Business information:  
As a business firm or any other firm it handle large amount of valuable data which contain of business strategies, list of customers and their personal information, personal data of the employees, fiscal records and budget, salary accounting of the employees etc. The above acknowledgement every information keep a value to the firm and that is the causes the company would assure the security of such sensible data. Customers and employee inside information is a type which can afford modification so watermarking and fraud object idea can be appropriate to forbid them from leakage but on the other way financial data can be not adapted so in this case we can use steganography and data allocation techniques.

Chemical formulas: Most of the products have like medicines, soap, cosmetics etc. have their personal chemical formula which is used for production. Every specific brand wants their products to be unique, look at a medicine brand make a medicine or more, the producers will never want their challenger get to better them in market for which the chemical formula mandate to be unbroken secure from leakage while dispersion. The importance of these formulas are required to be exact therefore it is not applicable to use such techniques which alters the data and there techniques such as steganography, data allocation can be used to forbid data from leakage.

Criminal information: The police department contain of criminal record or data wherein they have list threatening or baleful criminals as well as top criminal and the data of the crime they have mentioned. These records are more vulnerable to fire like leakage, besides the police section also have strategies to face off these criminals and this is the main reason why they need to safe the information. These information should not be altered in any way and this is the reason data allocation, cryptography, and steganography can be used. Since we are discussing about crime and the other thing which comes into our minds is digital forensics. Mostly the people have judgment that the meaning of digital forensics is investigating of crime. All important rule in digital forensics are three basics principles: that the evidence is acquired without altering it; that this is demonstrably so; and that analysis is conducted in an accountable and repeatable way. The rule of digital forensics says ‘The use of scientifically derived and proven methods toward the collection, preservation, identification, validation, analysis, interpretation, documentation and presentation of digital evidence derived by the digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions.

Military information: Defensive strategies, weaponry, offensive strategies and information of any particular soldier are mainly in the military information. These information cannot be affordable to be found by any non-military individual or any enemy since it can guide to defeat of that particular military force who data has been leaked and this is the scenario where cryptography comes into picture to safe the information.

PREVENTION TECHNIQUES:  
Now we discuss about the prevention and after that come to the discussing of techniques.  
A. Encryption  
Encryption has now a difficult security feature for booming networks and active home users alike. This security idea uses mathematical schemes and algorithms to scuffle data into unclear text. It can only by decrypted or decoded by the party that possesses the associated key. (FDE) Full-disk encryption offers some of the best security available. This technology give authority to you to encrypt every part of data on a hard disk drive or a disk. Full disk encryption is also powerful when hardware solutions are used in conjunction with software elements. This combination is frequently referred to as end-point or end-based full disk encryption.

B. Strong User Authentication  
Certification or authentication is some other part of data security that we face off with everyday computer usage. Let, think about when you log into your blog account or email account. That one sign-on process is a form authentication that allows you to log into files, folder, applications and even an full computer system. After logged in, you have various given privileges until sign out. Few systems will cancel a session if your system has been constant for a few amount of time, requiring that you prove certification once again to re-enter. Every sign-on scheme is also improve into strong user
authentication systems. However, it requires particular to sign in using different factors of authentication. This may include a one-time password, a password, a fingerprint or even a smart card.

Personal data: This difficult information is powerful ammunition to an catch the thief as it could give them easy access to your credit cards, bank account, medical records and other critical property. An organization kept highly sensitive or confidential private data (such as data about individuals’ health or finances) which could cause destroy or distress to those individuals if it fell into others hands. The organization’s data security measures should concentrates on any potential threat to the organization’s information systems or to the information. Being the part of security measures, an organization assure that information on computer laptops issued to faculty is secured by encryption, and that desk-top pc screens in its offices are positioned so that they cannot be seen by casual passers-by. Paper waste is collected in secure bins and is shredded on site at the end of each week. This paper focuses on stop of sensitive data as well as detection of any data been passed. There are two basic techniques first where we need to alter the data and second without alteration.

1. Alteration Techniques
   - Fake objects
   - Watermarking
   - Cryptography

2. Techniques without alteration
   - Data allocation
   - Steganography

![Diagram](image)

Fig1. Fake objects that are not identified by agents

Now study about these techniques in details.

1. Water marking technique

Traditionally, leakage detection is accessed by watermarking, e.g., a ideal or unique code is built in every distributed copy. If that copy is later founded in the hands of third or an unauthorized party, the data leaker can be recognized. Watermarks can be very useful in few cases. If we define Real time example then watermarking can also be used for edited or compressed video data. To provide copyright protection and copy protection for digital video and audio data, two complementary techniques are being involved: watermarking and encryption. Encryption idea can be used to safe digital information during the transmission from the source to the destination. However, after the receiver has received the information and decrypted the data, the data is in the clear and no longer protected. Watermarking techniques can complement encryption by embedding a secret imperceptible signal, a watermark, directly into the clear data. This watermark signal is built in such a way that it cannot be removed or modify without changing the quality of the video or audio data. The watermark signal can for instance be used for copyright protection as it can hide information about the author in the data. The watermark can now be used to prove ownership in court. Another important application for which the watermark code can be used is to find the source of unauthorized copies by means of finger print techniques.

Drawbacks-

Watermarks contain few modification of the original data. Furthermore, watermarks can sometimes be demolished if the data recipient is malicious.

1. Fake object technique

Fake object are another technique in original data which plainly increases the chance to find guilty agents. The distributors have authority to add fake objects to the distributed data in order to increase his effectiveness in detecting guilty broker. The techniques of disturbing or modifying data to detect leakage is not new. However, in most cases, particular objects are modified, e.g., adding a watermark to an image or by adding random noise to sensitive salaries. Talking about real time use of false objects our survey says we can take trace records one of them. Trace records are basically owned addresses by entities. Suppose we consider two companies X & Y, suppose company X sells a mailing list to company Y for say advertising. At that time company X add trace records to the mailing list because of which every time when company Y uses that mailing list company X receives a copy of it. This therefore can find unauthorized use of data.

Disadvantages-

It moderate to alteration of original data. Altering original data may not be right every time for e.g. let we have financial information like employee’s
salary or budget, such sensitive information cannot be altered as any alteration can lead to financial crises of the company.

3. Steganography:
The science and art of hiding information by embedding messages within other, apparently innocuous messages. Sometimes steganography is used when encryption is not permitted. Commonly steganography is used to supplement encryption. Using steganography an encrypted file may still hide the information, so even if the encrypted file is deciphered, the hidden message is not seen. Steganography can be reasoned as technique which embeds sensitive data with other data using rules and techniques because of which we can find authorized person or entity. Apparently watermarking widely used in steganography.

Steganography can be reasoned as technique which embeds sensitive data with other data using rules and techniques because of which we can find authorized person or entity. Apparently watermarking widely used in steganography.

Watermarking are two types invisible and visible watermarking. As the name suggest is visible it is a logo or text which show the ownerships. For example the logo of the popular company like Audi, Mercedes, BMW very apparently notify their particular cars. These logo can even be for advertisement and promotion of these company. After that we consider invisible watermarks, they can be built in video, images, audio as well as text. The remarkable view about invisible watermarking is that they seem to be the original object and besides this idea can be used for copyright prevention which authorizes authors, writers, creators etc.

Disadvantage:
It takes long time to make the code. If you were to send a code to another person the decryption and the encryption technique usually takes a very long time. Overall cryptography is time taking and a very long process.

5. Data allocation
The main goal of the data allocation problem as how can the distributor smartly give data to agents in order to improve the probability of detecting a guilty agent. Authenticated user receive the file send by the admin and users can edit their account details or any other information etc. Agent see the secret key information through the mail. In order to gain the chances of finding agents that leak data. Coming to the details description of our proposed system, suppose the admin sends file or data to user at that particular time system create 4 digit random number also known as secret key and mails it to intended user, the user needs to enter the key while downloading the file or data this apparently ensure the security of the sensitive data. Now if someone other than the intended user tries to acquire the data and enters wrong secret key the system acknowledges the intended user via mail. The intended user can lock the sensitive file or data by setting desired password.
III. CONCLUSION

From this survey we contain that the data leakage detection system model is very important and useful as compare to the existing watermarking model. We can give the security to our data during its transmission or distribution and even we can find if that gets leaked. Thus, using this model security as well as tracking system is developed. Watermarking can just give the security using different algorithms through encryption, whereas this model gives security plus detection technique. This model is very helpful in different companies, where data is distribute through any private or public channel and shred with third party. Now, industry & various offices can rely on this security & detection model.

Acknowledgements:

It is with a great sense of gratitude that we acknowledge the support, time to time suggestions and highly indebted to my guide Prof. ABC.

REFERENCES

Journal Papers: