Data Leakage Detection in Financial System with Optimized Data Allocation by Identifying Critical Data From Database

Vaishali B. Langote*, Prof. Vidya Deshpande
MIT Pune, Department of Information Technology

Abstract

In any financial system it is required to hand over their data to trusted third parties when the growth of company is very fast and also data cannot be handled by themselves. If an outsourcer is doing the database maintenance, he must have the exact no of customers, customer identification numbers and all confidential information about the customers. So the chances of leakage are very high, traditionally, leakage detection is handled by watermarking, i.e., a unique code is embedded in each distributed copy. If that copy is later discovered in the hands of an unauthorized party, the leaker can be identified but the problem with watermarks is it can be removed. We call the owner of the data the distributor and the supposedly trusted third parties the agents.

Our goal is to identify the sensitive data; allocation of data in optimized manner and detect when the distributor's sensitive data has been leaked by agents, also identify the agent that leaked the data and if possible, preventing the leakage of data.

1. Introduction

There arise lots of challenges while distributing data to trusted third parties and also monotonous task of detection agents who leaked data. Here, we propose two techniques one for detection of sensitive data and another steganography technique which allows data allocation in more secured manner with optimized allocation strategy by minimizing overlap.

Optimization approach for detecting the critical data on a database using which we can easily identified the sensitive data from database. We demonstrate that through purposeful introduction of malicious transactions. The problem of critical data detection is reposed as an optimization problem over graphs. We further pose the generalized critical data detection problem and develop methods for introducing deterministic and linguistic constraints for the critical data detection. After the detection of sensitive we propose data allocation strategies (across the agents) that improve the probability of identifying leakages. These methods do not rely on alterations of the released data. In some cases, we can also inject "realistic but fake" data records to further improve our chances of detecting leakage and identifying the guilty agent.

2. Related Work

P. Papadimitriou and H. Garcia-Molina, defined handling of data request by two methods simple and explicit method which allows data allocation in random or optimal strategy with minimum data overlap as well as detection of leakage with watermarking technique [1].

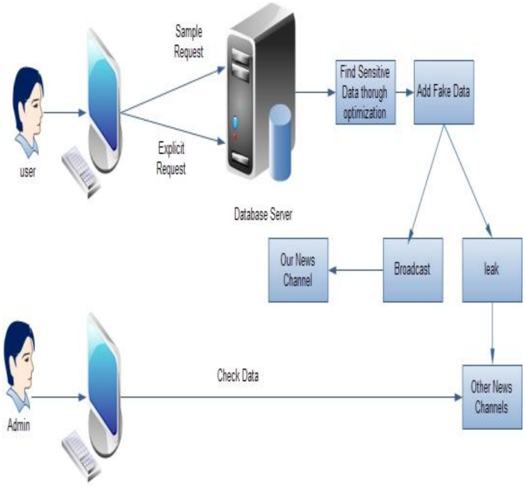
R. Agrawal and J. Kiernan, proposed technique of Watermarking on relational databases [2].

Umamaheswari, S., Geetha, H.A. defined the way to detect the guilty agents with the impact of probability guessing [3].

Sandip A. Kale, Prof. S.V.Kulkarni proposed various techniques of watermarking for detection of data leakage.

3. Proposed system

In the proposed system we have introduced Optimization approach for detecting the critical data on a database. This critical data we are allocating to more reliable agents which decreases chances of leakage. This system is more reliable it provides Steganography with LSB technique for detecting guilty agent with highest impact of probability. This system is specifically designed for financial system. In financial system data is more confidential and needs to be maintained more securely hence cryptography is used. The main problem of detection of guilty agents [3] is guessing the probability that who leaked the data is being eliminated and this system ensures the detection of leakage of the data and agent accurately. Here the additional feature of avoiding the guilt leakage is done.



4. Text Steganography with LSB Technique

In the field of Data Communication, security-issues have got the top priority. So, of late the degree of security provided by a security tool has become the main evolutionary criteria of it. Classical cryptography is one of the ways to secure plain text messages. Along with that at the time of data transmission, security is also implemented by introducing the concept of steganography, watermarking, etc. In this types of combined approach, there exits some drawbacks. In remote networking, at the time of transmission of hidden encrypted text message, if the eavesdroppers get the track of the hidden text, then they could easily get the encrypted text.

Now breaking of encrypted text message can be achieved by applying some brute force technique. So, there remains some probability of snooping of information. So, this type of techniques incurs another level of security which can route the Cryptanalyzer or Steganalyzer in a different direction. The work proposed here represents a heuristic approach to introduce the concept of Multi Layer Data Security algorithm in the field of combined Cryptography and Steganography. The algorithm, that we have proposed here, will secure the text message in multiple protection layers. Here, we have used the concept of Cryptography and Steganography (as two Layers of Security).

LSB Technique:

Least significant bit (LSB) insertion is simple approach to embed information in a cover object. In this technique, the data in the binary form is to be hidden into the LSBs of the carrier bytes.

5. Algorithm for allocating data to agents in optimal way.

Embedding Algorithm

The proposed algorithm includes the hiding of encrypted data within the binary text file using LSB technique. The data is hidden inside the binary text file using LSB.
Steps to embed data are as follows:

Inputs: Text file and text to embed.
Output: Secrete code embedded text file.

1. Agent Request for data {R1,R2,R3…..Rn} and Condition {C1,C2,C3….Cn}
2. For Each R → C1,C2,C3…Cn upto Rn
3. Responses to client with data by adding fake objects
4. The original recordset stored as text file is taken as an input. Calculate the size of the file convert the bytes to bits.
5. Extract all the characters from the text file for encryption and store them in a character-array.
6. Convert the encrypted text to the number system using ASCII code to generate the binary matrix.
7. For embedding data inside the Text File, the LSB of each record is stored into the character array.
8. Choose the first character from the binary matrix and pick the data bits of the corresponding character. Replace the LSB of with data bits from the binary matrix.

Extraction Algorithm

In extraction algorithm the original cover object is not required. It requires only the output object and the steps are:

Step 1: Consider Character Array.
Step 2: Extract all the values of the output object.
Step3: Start scanning from first value of the character-array and extract LSB. Store the extracted LSB in the character array.
Step 4: Generate the message by combining the bits of the character array.

6. Optimized approach for Detection of sensitive data

Problem of critical data detection is reposed as an optimization problem over graphs. We further pose the generalized critical data detection problem and develop methods for introducing deterministic and linguistic constraints for the critical data detection

1. For deciding sensitive data we will take subset V' of V of n nodes is said to be critical if it is a global maximum of the following cost function

$$F := S^2 + \beta(1 - \frac{n}{N})^2$$

Where S is the soiled measure due to the n nodes into which purposeful malice is introduced, N is the total nodes on the graph.

2. Using Following formula we can calculate soiled measure of some nodes.

The cost of the above Equation may be enhanced to include that less nodes is more preferred.

$$S^2 + (1 - \frac{n}{N})^2$$

3. Through the language of clean and soiled measures, the notion of critical subset of the graph is posed as an optimization problem over the graph.

Finding Guilty Agent and Guilty Agent Avoidance

Here we proposed algorithm which, ensures the data leakage detection probability is very high and also the detection of the guilty agent is more accurate. It also ensures the detection of the guilty agent and respective agent is blocked i.e. access to the system is forbidden hence further leakage is avoided.

Conclusion

In the real scenario it's very difficult to handle large data. Hence there is always a need to hand over sensitive data to the trusted third parties. Even if we had to hand over sensitive data, to the agents, we could use encryption along with steganography each object so that we could trace its origins with absolute certainty. The data distribution strategy improves the distributor's chances of identifying a leaker.

References

[1] Panagiotis Papadimitriou, and Hector Garcia-Molina, “Data Leakage Detection”, IEEE transactions on knowledge and data engineering, vol. 23, no. 1, january 2011
[2] R. Agrawal and J. Kiernan, “Watermarking Relational Databases,” Proc. 28th Int'l Conf. Very Large Data Bases (VLDB '02), VLDB Endowment, pp. 155-166, 2002.
[3] Umamaheswari, S., Geetha, H.A., “Detection of guilty agents”, 2011 National Conference on Innovations in Emerging Technology (NCOIET), Feb 2011
[4] Sandip A.Kale, Prof. Kulkarni S.V. “Data Leakage Detection”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 9, November 2012
[5] S.K.Bandyopadhyay, Debnath Bhattacharyya, Poulumi Das, “Text Steganography: A Novel Approach”, International Journal of Advanced Science and Technology Vol. 3, February, 2009
[6] A.A.A Gutub, M.M Fattani, “A Novel Arabic Text Steganography Method Using Points and Extensions”, Proceedings of WASET, 28-31, May 2007.
[7] N. N. Emam, “Hiding a large amount of data with high security using steganography algorithm”, Journal of Computer Science,223 – 232, April 2007.
[8] Prashanth Alluvada, “Optimization Approach for Detecting the Critical Data on a Database”, 20 Apr 2008