# Data Leakage Detection

**Ghagare Mahesh[1], Yadav Sujit[2], Kamble Snehal[3], Nangare Jairaj[4], Shewale Ramchandra[5]**

*Dept. of Computer Science & Engineering, DACOE Karad, Maharashtra, India*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** A *data distributor of the any company has given sensitive data about their work or business to one or more authorized person. If that data are shared by these agents who does not authority to share this data to any employee and that data are leaked and found in an unauthorized place. In this project we are implementing the system for detection of leaked data and possibly the agent who is responsible for leakage of data. The distributor must access the leaked data came from one or more agents. We use any duplicate data which does not known to the agent's for identifying leakages. In some cases, we can also use "realistic but fake" data records to further improve our chances of detecting leakage and identifying the unauthorized person who leaked data.*

*We provide an alert service to the distributor. When the information has been leaked by the agent then it will send a message to the distributor or authorized person the data is going to leak as well as it can identify that leaker by the showing its IP address.*

*We use the different or one or more fake objects to the different sharing data files. We use one fake object for one data file records.*

***Key Words*: Distributor, Guilty Agent, Third Party, Sensitive Data, Alert.**

## 1. INTRODUCTION

A data distributor or head of the any company, etc. has given sensitive data that means the important data or information about their work or business to one or more agents or the authorized person or employee (third parties)[3][4]. It should not be handling without authority and it should not interfere by any unauthorized person.

The idea of our project is to find guilty agent who leaked the sensitive or confidential data of the company. And give alert message to the guilty agent if he/ she are break rule again and again then take a legal action.

The distributor can register their name and information then he/ she has authority of distributor. The client or agent can register their information and send request to the distributor. After request is receive the distributor provide unique username and password to the agent. Agent has login by entering that username and password. The agent has send request to the distributor for data the request is explicit or sample. Then distributor can check the request send by the agent is our agent? Then distributor checks the type of request i.e. explicit or sample request. It can collect data from database and add fake object. After that the distributor check whether the data is already sends that agent or not. If data is already sends then it will send message sorry data is already sent. If data is already not sends then it can send that data to the particular agent.

If the third party or any client send request to the agent and the agent send them sensitive data that has does not authority to send or share data. The alert message is send to the distributor that the agent is guilty. Then distributors sends warning message to that agent don't do this again if it happens again by that agent then distributor take action against that agent.

Goals and Objectives:

To detect the agent who leaked the confidential data and send alert message to the distributor.

The objectives of the "Data Leakage Detection" are as follows:

- Detection of guilty agent
- Send message or email to the distributor with identification of guilty agent
- Send alert message to the guilty agent
- Take legal action on agent when he/she break rule after the alert message

### 1.1 Need:

The need of our project data leakage detection:

- Protect of complex and important data[2]
- Find guilty agent
- Secure confidential data [2]

### 1.2 Literature Survey

An enterprise data leak is a scary proposition. Security practitioners have always had to deal with data leakage issues that arise from email and other Internet channels. But now with the use of mobile technology, it's easier for data loss to occur, whether accidentally or maliciously. The guilty detection approach we present is related to the data provenance problem tracing the lineage of S objects implies essentially the detection of the guilty agents [2]. And assume some prior knowledge on the way a data view is created out of data sources. Our problem formulation with objects and sets is more general As far as the data allocation.

Strategies are concerned; our work is mostly relevant to watermarking (Stenography) [1] that is used as a means of establishing original ownership of distributed objects. Watermarking is a unique code is embedded in distribute copy [5]. Data leakages can be identified using these original data [5]. Thus watermarking is a useful methodology. But sometimes the watermarks can be destroyed if the data recipient is malicious [5]. Hence this technique proves to be inefficient. Finally, there are also lots of other works on mechanisms that allow only authorized users to access sensitive data through access control policies [2]. Such approaches prevent in some sense data leakage by sharing information only with trusted parties. However, these policies are restrictive and may make it impossible to satisfy agent's requests.

## 2. Implementation Details

### Distributor Module:

Distributor is the main authorized user of the company. It can accept the registration request of the agent and send the user ID and password for authorized person of the company. It accept data request from the agent and check the request is sample or explicit. It maintains sensitive data in database and distribute as per agent request by adding the fake object in the original data.

If the distributor has received the alert message the agent has leaked data or the agent is guilty then it can send warning message to the guilty agent and take action against that agent. Distributor also changes the password. It can delete any agent by click on delete agent option.

### Agent Module:

Agent can register their information and send registration request to the distributor. When distributor can accept their registration request and accept user ID and password for login. Then agent login by enter user ID and password. Agent can send request for data by using two requests Explicit and Sample request. The explicit request is for particular data and sample request for number of data. Sometimes the agent can accept the request of the third party and it share the data and the agent is guilty.

### Data Allocation or Distribution Module:

As per request, Fake objects are objects generated by the distributor that are not in set T. The objects are designed to look like real objects, and are distributed to agents together with the T objects, in order to increase the chances of detecting agents that leak data.

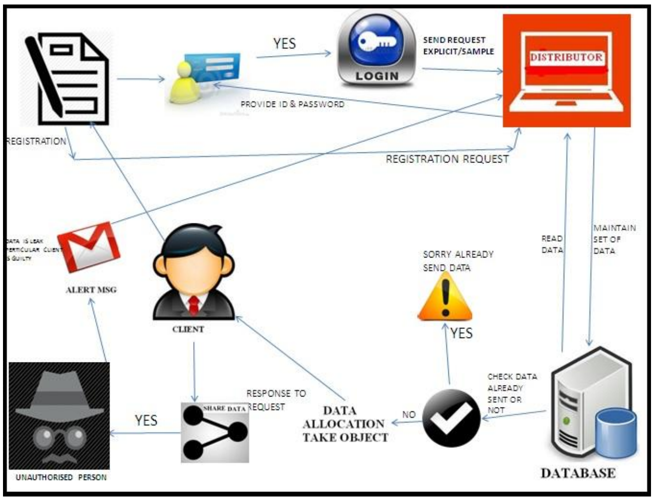

Figure1. Architecture Diagram

## 2.1 Allocation Strategies:

### Explicit Data Requests:

In this case distributor is not allowed to add fake objects to the distributed data. So, the data allocation is fully defined by the agents' data requests. Therefore, there is nothing to optimize. It uses two algorithms based on request type i.e. either e-Random algorithm to randomly add fake objects. We used e-Random algorithm to randomly select agent for fake object allocation.

### Sample Data Requests:

An object allocation that satisfies requests and ignores the distributor's objective is to give each agentUi a randomly selected subset of T of size mi. We denote this algorithm as s-random.

### Data Leakage Detection Module:

In this module distributor detects guilt agents based on assigned fake objects to corresponding agents. It audits the probability of getting guilty agents**.** Send the alert message to the distributor the agent is guilty.

### 2.1.1. For Explicit Data Request

**Algorithm 1.** Allocation for Explicit Data Requests (EF)

```
Input: R1; . . .;Rn, cond1; . . . ; condn, b1; . . . ; bn, B
Output: R1; . . .;Rn, F1; . . . ; Fn
R←Ø Agents that can receive fake objects
for i= 1; . . . ; n do
    if $b_i$ > 0 then
        R⟵ R ∪ {i}
    Fi ⟵Ø
while B > 0 do
    i⟵SELECTAGENT(R;R1; . . .;Rn)
    f ⟵CREATEFAKEOBJECT($R_i$,$F_i$,$cond_i$)
    Ri ⟵Ri ∪ {f}
    $F_i$ ⟵$F_i$ ∪{f}
    bi ⟵bi - 1
    if $b_i$ = 0then
        R⟵ R\{$R_i$}
    B⟵ B - 1
```

**Algorithm 2.** Agent Selection for e-random

```
**function** SELECTAGENT (R,R1, . . .,Rn)
    i ⟵select at random an agent from R
    return i
```

**Algorithm 3.** Agent Selection for e-optimal

```
function SELECTAGENT (R,R1, . . .,Rn)
    i ⟵$\underset{i' : R_{i'} € R}{\text{argmax}}(1/|R_{i'}| - 1/|R_{i'}|+1)\Sigma|R_{i'} \cap R_j|$
    return i
```

### 2.1.2 for Sample Data Request

**Algorithm 4.** Allocation for Sample Data Requests (SF)

```
**Input:** m1; . . .;$m_n$, |T|    Assuming mi ≤|T|
**Output:** R1, . . .,Rn
a ⟵$0_{|T|}$     a[k]:number of agents who have received object $t_k$
R1 ;⟵Ø, . . .,Rn⟵Ø ;
remaining⟵$\Sigma^n_{i=1}$ $m_i$
while remaining > 0 do
    for all i=1,. . . ,n : [$R_i$] <$m_i$ do
        k ⟵SELECTOBJECT(I,$R_i$)  May also use additional parameters
        Ri ⟵Ri ∪ {$t_k$}
        a[k]⟵a[k]+1
        remaining⟵ remaining - 1
```

**Algorithm 5.** Object Selection for s-random

```
**function** SELECTOBJECT(i,$R_i$)
    k⟵ select at random an element from set {k'|$t_{k'}$€$R_i$)
    **return** k
```

**Algorithm 6.** Object Selection for s-overlap

```
**function** SELECTOBJECT (i,$R_i$, a)
    K ⟵{k | k =$\underset{K'}{\text{argmin}}$a[k']}
    k ⟵select at random an element from set
      {k'|k'€K∧$t_{k'}$€$R_i$}
    return k
```

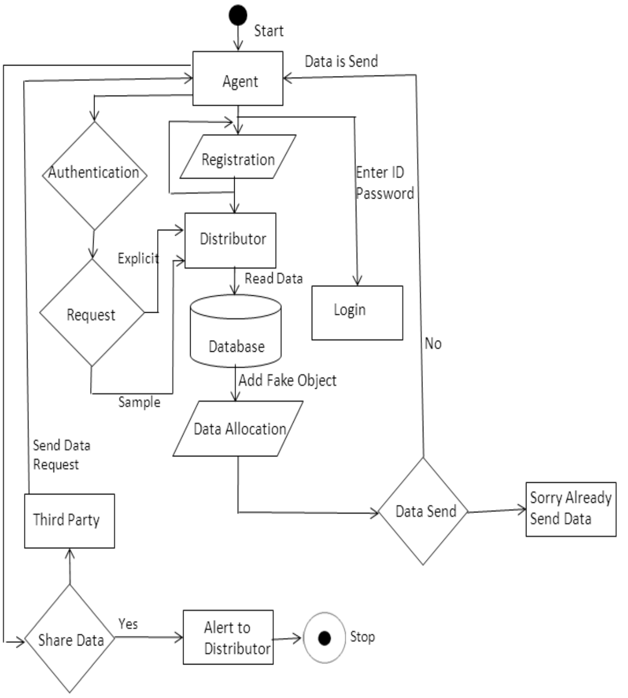

Figure 2. Data Flow Diagram of Data Leakage Detection

## 3. CONCLUSIONS

Thus, these modules successfully work according to IEEE paper. It can successfully login distributor to the system and register the new agent request and show confirmation message for registration. In our work the distributor can check the list of registration request for new agent and the agent and distributor also updates its information successfully.

In this project in next modules we can implement following idea: User ID and password send to the agent for login to system. The agent should send data request to the distributor and distributor check the request and send data to the agent by adding fake object in data allocation module. In agent guilt module we can check send alert message to the distributor when the agent has share any confidential data. This is main goal of our project.

## ACKNOWLEDGEMENT

It is our privilege to acknowledge with deep sense of gratitude to our project guide Prof. Jadhav J.H. and our H.O.D Prof. Birnale M. A. whose supervision, inspiration and valuable discussion has helped us tremendously to complete our project. Their guidance proved to be the most valuable to overcome all the hurdles in the fulfillment of this mega project on "Data Leakage Detection".

We are grateful to Principal Dr. Jalindar R. Patil for direct or indirect help in the completion of this project. Last but not least, this acknowledgement would be incomplete without rendering my sincere gratitude to all those who have helped us in the completion of this project.

## REFERENCES

[1] R. Agrawal and J. Kiernan, "Watermarking Relational Databases,"Proc. 28th Int'l Conf. Very Large Data Bases (VLDB '02), VLDBEndowment, pp. 155-166, 2002. IEEE Transaction and knowledge and data engineering, Vol.23, No.1, January 2011

[2] P. Bonatti, S.D.C. di Vimercati, and P. Samarati, "An Algebra for Composing Access Control Policies," ACM Trans. Information and System Security, vol. 5, no. 1, pp. 1-35, 2002.

[3] P. Buneman, S. Khanna, and W.C. Tan, "Why and Where: Characterization of Data Provenance," Proc. Eighth Int'l Conf. Database Theory (ICDT '01), J.V. den Bussche and V. Vianu, eds., pp. 316-330, Jan. 2001.

[4] P. Buneman and W.-C.Tan, "Provenance in Databases," Proc.ACM SIGMOD, pp. 1171-1173, 2007.

[5] Ms. Aishwarya Potdar1, Ms. Rutuja Phalke2, Ms. Monica Adsul3, Ms.Prachi Gholap4
B.E, Department of Computer Engineering, KJCOEMR, Pune University, Pune, India, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 4, April 2013

## BIOGRAPHIES

Mr.Mahesh Mohan Ghagare
mahesh.9800@gmail.com
Student of BE.CSE
Dr.Daulatrao Aher College Of Engineering, Karad.

Mr. Sujit Suresh Yadav
sujit.yadav315@gmail.com
Student of BE.CSE
Dr.Daulatrao Aher College Of Engineering, Karad.

Miss. Snehal Balasaheb Kamble
sksnehalkamble1@gmail.com
Student of BE.CSE
Dr.Daulatrao Aher College Of Engineering, Karad.

Mr.Jairaj Vilasrao Nangare
jairaj.nangare@gmail.com
Student of BE.CSE
Dr.Daulatrao Aher College Of Engineering, Karad.

Mr.Ramchandra Sadanand Shewale
shewale2012@gmail.com
Student of BE.CSE
Dr.Daulatrao Aher College Of Engineering, Karad.