International Journal of Advances in Engineering Research http://www.ijaer.com

(IJAER) 2014, Vol. No. 8, Issue No. V, November **ISSN: 2231-5152**

DATA LEAKAGE DETECTION SYSTEM FOR CLOUD-BASED STORAGE SYSTEMS

K. Manoj Kumar+ G. Shubhang+ G. Rajesh Chandra *

**Assistant professor, Department of Electronics and Computer Engineering, K L University, India.*

+ B Tech, Department of Electronics and Computer Engineering, K L University, India

ABSTRACT

In this age of cloud computing the incidences of theft/leakage of valuable, sensitive data is on the rise. Data security has gained greater significance than ever before. This paper attempts to deal with the problem of data leakage and its detection, a rising phenomenon especially in big organizations. We are all too familiar with stories of data loss from laptop theft, hacker break-ins, and backup tapes being lost or stolen, and so on. It is possible only by Data Leakers, who usually are authorized persons. The identification of the Data Leaker is very big task. In the past methods such as Watermarking and fake data addition were used to identify leakers. But in such methods reliability, consistency is very low and largely incompetent in the present age. Thus a sophisticated system to confront such activities with reasonable competence is highly essential. This paper proposes a highly effective system employing various techniques to cleverly detect data leakage and the leakers.

Keywords: ***Watermarking, Guilty party, Forged Data.***

INTRODUCTION

This paper examines the phenomenon of data leakage, detection and its impact on organizations. Data leakage may be defined as the illegal transfer of valuable/sensitive data by an entity to unauthorized entities. Data leakage detection is the process of finding the data leaker by using various techniques ranging from interrogation, watermark/fakedata addition to other modern techniques. The effects of data leakage could range from loss of valuable data, privacy, cyber-theft, to threat to economy and national security. In this paper a new approach using various techniques such watermarking, fake data addition, guilty party etc. to build a fail-safe data leakage detection especially suited to cloud-based data storage systems is proposed for enhanced security. The system must be able to perform in various areas and through various means given the wide scope of data leakage today, which was previously limited to email and few/certain individuals.

WATER MARKING

Watermarking is basically an ancient technique in use from hundreds of years. We are using these techniques but here we are providing a new algorithm for water marking concept this algorithm provides high securable water marking. In this paper first our task is to add

watermarking image to the original data. Here we are developing a new watermarking technique that will takes input with an image and gives a unique watermarking image that may be useful for our approach this approach can be developing in the following algorithm

Input:

– f is the original image of sizeM1 ×M2,
– w1 is an set that is belongs to {−1, 1} it is a digital watermark image with the range N1 × N2.
In this algorithm we are giving the input f, W1. The process of algorithm works as,

Algorithm[3]

1. Initiate l is from 1 to L

2. Initiate s from 1 to S generate s1(a,b)

3. Bring into being key Key1 ∈ {0, 1}
if Key1 is zero then don't entrench a spot other wise
– sort the specified coefficients as:
fs1, l(a, b) _ fs2,l(a, b) _ fs3,la, b)

– do quantization by divide fs1,l(a, b) and fs3,l(a, b) into bins using the Following form $\Delta = (fs3,l(a, b) - fs1,l(a, b)) / (2Q - 1)$.

4. The compound convert coefficients in each band are scaled back to the levels of the original image transform coefficients using the min and max coefficient Values.
◦ the fused coefficients fuseed are computed as follows:
Fused = αfs,l(a, b) +W(i, j).

5. An converse make over is now computed to give the watermarked image

Output:

Water marking image

The output of this algorithm gives image of watermarking. Next our task is to add this message with the data. The watermarking concept will be common to data of every trusted agent of organization.

2.1 Adding Forged Entities:

Addition of Forged Entities plays important role in our paper. Without these entities data administrator can't distribute the data to the concerned parties. Here forged entities are basically data that appears to be original but isn't. Once the forged entities are added to the original data in a unique manner, it is distributed to the concerned parties. For creating forged data, there are many algorithms. Here our task is to generate forged data depending on original data we are generating forged data. The basic idea behind the addition of forged data is to ensure each copy of the data or a file has unique data such that it does not raise any suspicion and helps in nabbing the data leaker.

2.2 Generation of forged entities:

The generation of forged data for an agent Ui as a black-box function CREATEFORGEDDATA(Ri, Fi, Condi), that automatically catches input as set of all data Ri, the subset of forged entities Fi that Ui has received so far and Condi and returns a new forged entity. This function wants Condi to gives a valid object that satisfies U's condition. Set Ri is needed as input so that the created forged entity is not only valid but also identical from other original entity.

GUILTY PARTY

The agent who concerned as leaker has an odd value as it is not official to receive data from the agents or owner. Before we present the general formula for calculating the probability P{Gu|Sa} that an agent Ui is guilty, we provide a simple example. Assume that the distributor2 set O, the agent sets A and the target set S are:

O = {o1, o2, o3}; A1= {o1,o2}; A2= {o1, o3}; S = {o1, o2, o3};

In this case, total three of the owner's objects have been leaked and have appeared in S. Let us first consider how the target have obtained object o1, which was given to the both agents. The target either guessed o1 or one of A1 or A2 leaked it. We know that the probability of the former event is p, so assuming that probability that each of the two agents leaked o1 is same, we have the following cases:

The target guessed d1 is leaked with probability p,

- Agent A1 leaked o1 to S with probability (1-p)/2

- Agent A2 leaked o1 to S with probability (1-p)/2.

Process:

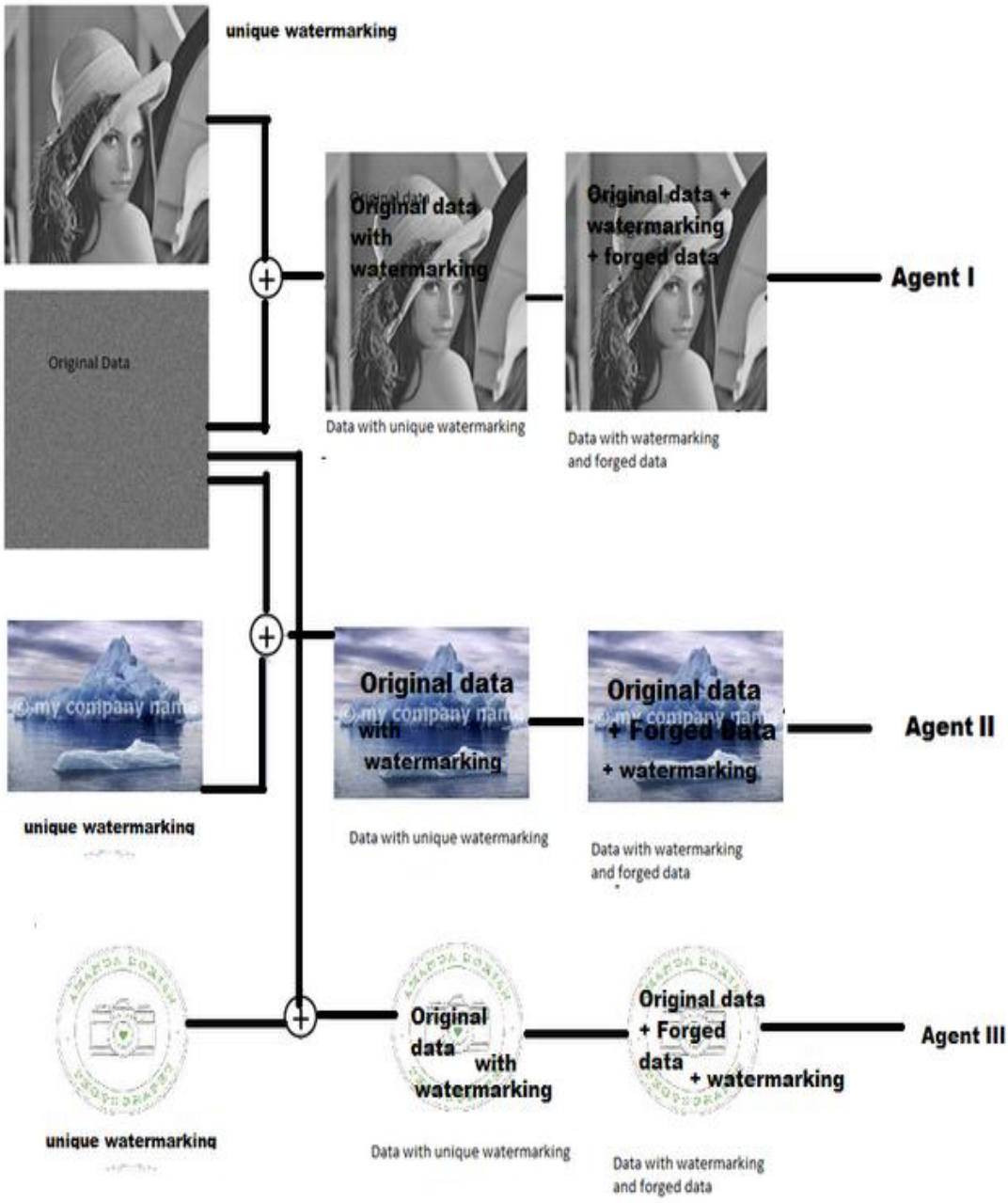

Figure 1: Process of new approach

EXPLANATION OF PROCESS

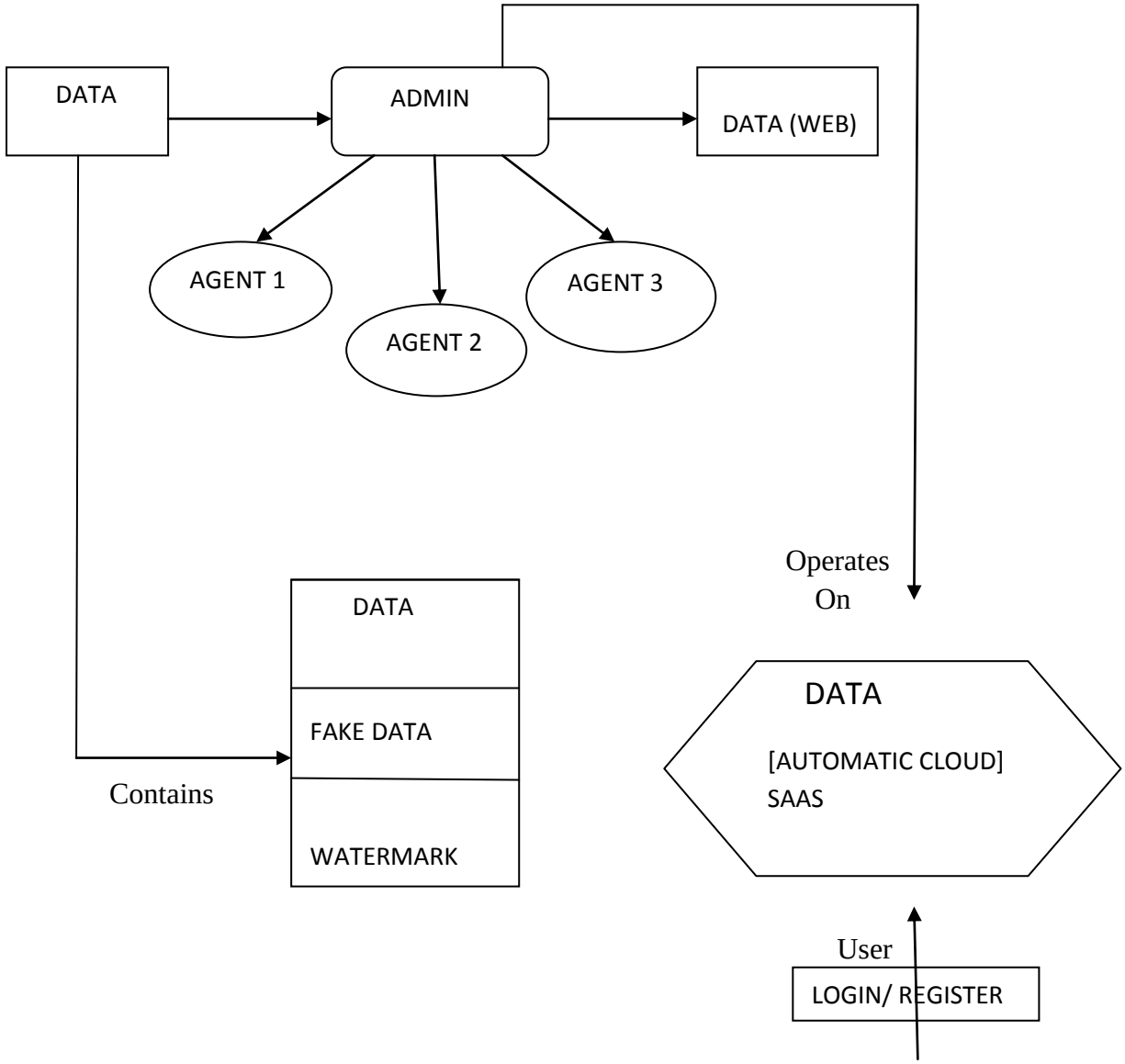

Figure 2: Diagrammatic Representation of the System.

The Data Leakage Detection System using Cloud is basically a cloud based database administration system with an inbuilt system for managing users and detecting data pilferage and the culprits. The administrator distributes the data to trusted agents who may leak the data.
So in our process the System automatically generates a unique watermark and forged data which are embedded in the original data as the user logs in to obtain data with forged data and watermarking. Thus each user gets a virtually unique copy of the same data requested.

Let us take one example for the explanation of this process, in any organization the owner always shares data with trusted workers one of whom may leak the data making it available in other websites or causing misuse of that data by any other organizations thereby causing immense damage to the rightful owners. Our paper gives a new approach for these types of problems, in this approach the System automatically generates a unique watermark and fake data which are added to the original data that is distributed to the users to ease detection of the data leaker.

For putting this concept in practice several algorithms and are used to ensure a high success rate.

CONCLUSION AND FUTURE WORK

In our paper we have proposed a concept on Data Leakage Detection system in cloud computing. In this original data is added with watermarked image and forged data. Now the data will be distributed to the authorized entities, if any data is found in third-party sites the leaker can be identified with the help of forged data. In our future work we will encrypt that original data before distributing it to the agents. In this paper we have given just an approach, in our future work we will develop a project for data distributers with the algorithm of encryption for implementation in cloud computing.

REFERENCES:

1. A Copyright Protection using Watermarking Algorithm INFORMATICA, 2006, Vol. 17, No. 2, 187–198 @ 2006 Institute of Mathematics and Informatics, Vilnius, Abou Ella HASSANIEN

2. International Journal of Computer Applications (0975 –8887) Volume 42–No.6, March 2012 25,Guilt Model Process for Identifying Data Leakage and Guilty Agent in Data transmission,.

3. Digital watermarking: A Tutorial, Dr. Vipula Singh Professor and Head of Electrical and Computer Engineering Department Geethanjali College of Engineering and Technology, Hyderabad India

4. Cloud Computing: Concepts, Technology & Architecture by Thomas Erl; Architecting the Cloud: Design Decisions for Cloud Computing Service Models (SaaS, PaaS, & IaaS) by Michael J. Kavis; Cloud Computing Protected: Security Assessment Handbook by John Rhoton.