# Data Leakage Detection and Security in Cloud Computing

**Chandu Vaidya**
*Assistant Professor*
*Department of Computer Science & Engineering*
*Rajiv Gandhi College of Engineering and Research, Nagpur*

**Prashant Khobragade**
*Assistant Professor*
*Department of Computer Science & Engineering*
*Rajiv Gandhi College of Engineering and Research, Nagpur*

**Ashish Golghate**
*Assistant Professor*
*Department of Computer Science & Engineering*
*Rajiv Gandhi College of Engineering and Research, Nagpur*

## Abstract

Data leakage ad security of data is an essential component of cloud computing, most of the data has been proceed through third party application (TPA) and user are unaware about the security essentials. The internet based computing model provide shared resource to provide data on demand, such model provide service over the internet with computing resources. The TPA who expertise and provide crossing point between the user and the cloud service provider who facilitate service between them, the data store over cloud is passes through TPA in plain text and then it encrypted using the third party application. It might be possible that data can easily leaked over the internet, in this paper the proposed approach is to detect when the distributor's sensitive data has been leaked by agents, and if possible to identify the agent that leaked the data. Perturbation is a very useful technique where the data is modified and made less sensitive before being handed to agents and encryption algorithm SHA provide security of sensitive data before uploading over the internet cloud. The proposed model also provide security of data before uploading through TPA and after successfully upload data the secret key is shared with original user without intercepting TPA.

**Keywords- Data Leakage, Cloud Computing, data Security, SHA, Third Party Agent, Fake Object**

## I. INTRODUCTION

Today the present world regularly depends on transfer of information i.e. transfer of data from one person to another person. The data send by the distributor must be secured, confidential and must not be repeat as the data shared with the trusted third parties are confidential and highly important [1]. The data either access or read by any of the user or data tamper while uploading of data over the cloud may lead to data leakage. The enterprises have confidential data and if such data is shared or access without permission with any other person then it may damage to the organization. In cloud computing, software as a Service (SaaS) is define as software is act midway between the user and internet which deploy over the internet [4]. With SaaS, a source licenses an application to customers either as a service on demand, through a subscription, in a "pay-as-you-go" model, or (increasingly) at no charge when there is opportunity to generate revenue from streams other than the user. Software as a service is rapidly growing market as indicate high growth. This rapid growth indicates that SaaS will soon become common place within every organization and hence it is important that buyers and users of technology understand what SaaS is and where it is suitable [2].

Cloud computing security is growing in information security, computer security. It is generally suggested that information security control to be selected and implemented according and in section to the risks, by assessing the threats, vulnerabilities and impacts. Cloud security concerns can be grouped in various ways; Gartner named seven [9] while the Cloud Security Alliance identified fourteen areas of concern.

## II. LITERATURE REVIEW

Chandu Vaidya, Prashant Khobragade [1], suggest the data security problem stored the in cloud data storage, which is mostly a distributed storage system. Existing method rely on erasure-correcting code in the file distribution preparation to support redundancy equality vectors for verification of erasure coded data using the RSA encryption and focus on the scheme which achieves the privacy of client data with security and integration of data error localization and storage correctness insurance.

Prashant Khobragade [8] [10] introduced a way to analyse a browser history data to detect where the attack is happened. Data get collected from browser history and stored in database as evidence. The suggested method and forensic toolkit helps to analyze the data for the law enforcement.

All rights reserved by *www.grdjournals.com*

Sandip A. Kale, S.V. Kulkarni [2] focuses on watermarking, robust watermarking technique can be very useful in some cases, but again, involve some modification of the original data. Furthermore, watermarks can sometimes be destroyed if the data recipient is malicious.The distributor may be able to add fake objects to the distributed data in order to improve his effectiveness in detecting guilty agents.

Sushilkumar N. Holambe, Dr.Ulhas B.Shinde, Archana U. Bhosale [3], gives idea about data leakage is the big challenge in front of the industries & different institutes. They have discussed about distributor creates and adds fake objects to the data that he distributes to agents. The distributor may be able to add fake objects to the distributed data in order to improve his effectiveness in detecting guilty agents. Leakage detection is handled by algorithm, there is a big issue of the integrity of the users of those systems.

Priyanka Barge, Pratibha Dhawale, Namrata Kolashetti [6], give strategy of capturing data and data distribution strategies that improve the distributor's chances of identifying a leakage. The method proposed identifies guilty agents in case of overlap in the data agent. It is also possible to assess the likelihood that an agent is responsible for a leak, based on the overlap of his data with the leaked data and the data of other agents, and based on the probability that objects can be "guessed" by other means.

Data leakage [7][12] happens every day when confidential business information such as customer or patient data, source code or design specifications, price lists, intellectual property and trade secrets, and forecasts and budgets in spreadsheets are leaked out. When these are leaked out it leaves the company unprotected and goes outside the jurisdiction of the corporation. This uncontrolled data leakage puts business in a vulnerable position. Once this data is no longer within the domain, then the company is at serious risk.

## III. EXISTING METHODOLOGY

### *A. Fake Objects Method*

In some applications, fake objects may cause fewer problems that perturbing real objects. For example, say the distributed data objects are medical records and the agents are hospitals. In this case, even small modifications to the records of actual patients may be undesirable. However, the addition of some fake medical records may be acceptable, since no patient matches these records, and hence no one will ever be treated based on fake records. In this case, company A sells to company B a mailing list to be used once (e.g., to send advertisements). Company A adds trace records that contain addresses owned by company A. Thus, each time company B uses the purchased mailing list, A receives copies of the mailing. These records are a type of fake objects that help identify improper use of data.

### *B. Adversary Model*

This model [5] capture all kinds of data integrity threats and this cloud data not denoted at cloud client side but at the cloud service provide domain address. This can come from two attacks:

1) Internal Attack: Cloud service provider can be untrusted.
2) External Attack: This attack comes from outsiders and who are beyond control domain of cloud service providers.

## IV. METHODOLOGY

### *A. Proposed Method*

The proposed system, shown in figure 1, for data leakage is to detect, when the distributor's sensitive data has been leaked by agents, and if possible to make out the agent that leaked the data.

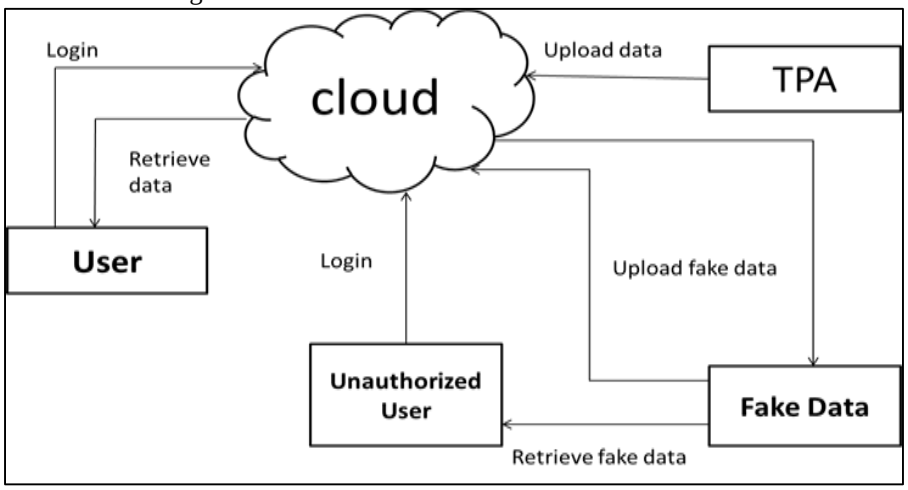


Fig. 1: Proposed Architecture of data leakage Detection

All rights reserved by *www.grdjournals.com*

Perturbation is a very useful technique where the data is modified and made "less sensitive" before being handed to agents. To enhance the security of users data the model have encrypted algorithm that provide the secure transmission of data between the user and TPA to cloud, the present algorithms for distributing objects to agents, in a way that improves our chances of identifying a leaker. With the adding of falsie data objects to the distributed set provide wrong information in case of data has been access by the any other third person [12]. Such objects do not correspond to real entities but appear realistic to the agents. In a sense, the fake objects acts as a type of watermark for the entire set, without modifying any individual members. If it turns out an agent was given one or more fake objects that were leaked, then the distributor can be more confident that agent was guilty. The TPA will upload the original data on cloud and data is in encrypted form and the fake data (predefined) is also uploaded on the cloud shown in figure 2.

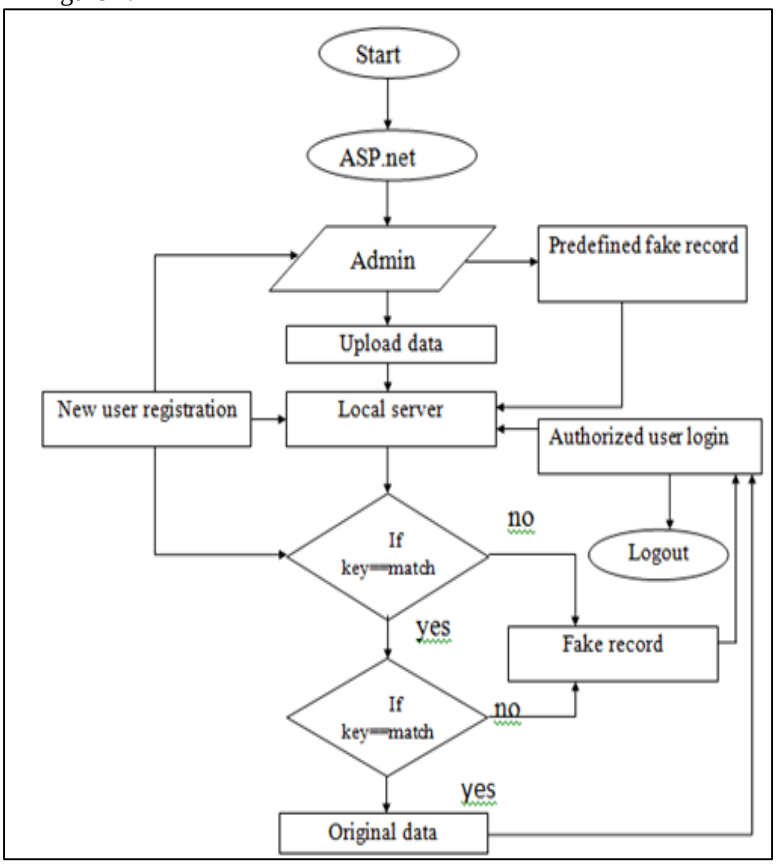


Fig. 2: Workflow of Proposed method

Admin here act as TPA is authorized data owner of the company and has the right to distribute the company's sensitive data to its employees. Admin is authorized to register the new employees, upload the data to server or send the data to the respective employees of the company.

### *B. SHA Algorithm*

The SHA secure algorithm provide the hash function at the time of data upload with he reference to the third party agent and the to the actual cloud, this security feature is added to enhance the security of sensitive data transfer over the internet. SHA algorithm is used because it used in data integrity and confidentiality of data with the use of message digest. In this approach SHA-2 (Secure Hash Algorithm 2), a subset of the cryptographic primitive family Keccak, is used to provide security to the data. The design od SHA-2 used internal 1024 block size and will operate And, Xor, Rot, Add (mod 264), Or, Shr and has 256-128-security bits.

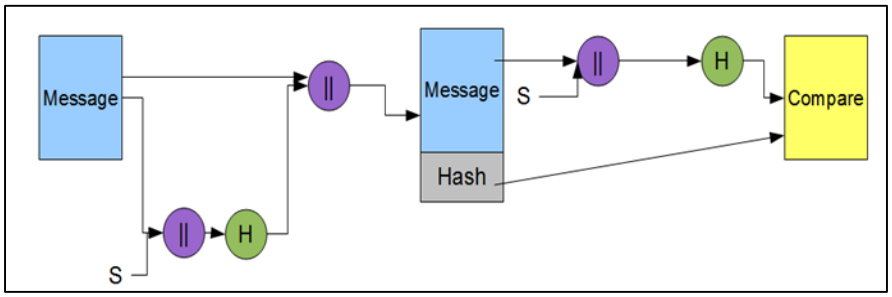


Fig. 3: Basic hash function working

All rights reserved by *www.grdjournals.com*

### *C. Basic Operations*

- Boolean operations AND, XOR and OR
- Bitwise complement, denoted by $\bar{}$.
- Integer addition modulo $2^{32}$, denoted by A + B.
- Each of them operates on 32-bit words. For the last operation, binary words are interpreted as integers written in base 2.
- RotR(A, n) denotes the circular right shift of n bits of the binary word A.
- ShR(A, n) denotes the right shift of n bits of the binary word A.
- A||B denotes the concatenation of the binary words A and B.

## V. CONCLUSION

The proposed method proposes data allocation strategies and adding "realistic but fake records" that improve the probability of identifying leakages and cryptographic [12] algorithm improve the security of data. It also provides security to our data during its transmission and even we can detect if that gets leaked. To deal with the problem of Data leakage, it presented implementation variety of data distribution strategies that can improve the distributor's chances of identifying a leaker. Scope of this system can be extended by making provisions for the generation of fake records dynamically according to the agent's request.

## REFERENCES

[1] Chandu Vaidya and Prashant Khobragade,2015, "Data Security in Cloud Computing", ISSN: 2321-8169 .
[2] Sandip A. Kale, Prof. S.V.Kulkarni," Data Leakage Detection", International Journal of Advanced Research in Computer and Communication Engineering,Vol. 1, Issue 9, November 2012
[3] Prof. Sushilkumar N. Holambe, Dr.Ulhas B.Shinde, Archana U. Bhosale,"data Leakage Detection Using Cloud Computing ", International Journal Of Scientific & Engineering Research, Volume 6, Issue 4,( April-2015)
[4] Ms. N. Bangar Anjali, Ms. P. Rokade Geetanjali, Ms. Patil Shivlila, Ms. R. Shetkar Swati,Prof.NB Kadu," DATA LEAKAGE DETECTION", IJCSMC, ISSN 2320–088XVol. 2, Issue. 5, May 2013
[5] Anusha Koneru, G. Siva Nageswara Rao, J.Venkata Rao," Data Leakage Detection Using Encrypted Fake Objects", IJCSNS International Journal of Computer Science and Network Security, VOL.14 No.3, March 2014
[6] Priyanka Barge, Pratibha Dhawale, Namrata Kolashetti," A Novel Data Leakage Detection", International Journal of Modern Engineering Research (IJMER) ISSN: 2249-6645 ,Vol.3, Issue.1, Jan-Feb. 2013
[7] Archana Vaidya, Prakash Lahange, Kiran More, Shefali Kachroo and Nivedita Pandey," Data Leakage Detection", International Journal of Advances in Engineering & Technology, ISSN: 2231-1963, March 2012.
[8] Prashant Khobragade, Latesh G. Malik,"A Review on Data Generation for Digital Forensic Investigation using Datamining", IJCAT International Journal of Computing and Technology, Volume 1, Issue 3, April 2014.
[9] Jon Brodkin, "Gartner: Seven cloud-computing security risks". InfoWorld. 2008-07-02. Retrieved 2010-01-25.
[10] Khobragade, P. K., & Malik, L. G., "Data Generation and Analysis for Digital Forensic Application Using Data Mining". In Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on (pp. 458-462). IEEE. April, 2014.
[11] Panagiotis Papadimitriou, Hector Garcia-Molina," Data Leakage Detection", IEEE transactions on knowledge and data engineering, vol. 23, no. 1, January 2011
[12] Chandu Vaidya etl. & BE scholars "Data leakage Detection and Dependable Storage Service in cloud Computing" IJSTE volume 2 issues 10 April 2016 ISSN online 2349-784X

All rights reserved by *www.grdjournals.com*