© 2019 IJSRST | Volume 6 | Issue 3 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology
DOI : 10.32628/IJSRST19624

Enhancement of Data Leakage Detection Using Encryption Technique

Monali U. Pawar, Shraddha A. Mankar, Snehal S. Mandhare, Siddhi N. More, Rashmi R. Patil
Navsahadri Education Society's Group of Institutions, Pune, Maharashtra, India

ABSTRACT

In the field of information technology data is crucial and sensitive part and it cause serious threat if it leaked. Now a day's huge amount of data is occurred and it requires to be maintained so peoples generally prefer cloud to store their data. Because one can access it from anywhere anytime and it contains huge amount of information and it provides numerous services to huge number of peoples. Statistics shows that lack of information encryption due to human mistakes can lead data loss. The advantages of cloud computing are Reduced Data Leakage, Decrease evidence acquisition time, they eliminate service downtime, they Forensic readiness, they decrease evidence transfer time the main factor to be discussed is security of cloud computing, which is a risk factor involved in major computing fields. So we aim to propose system to achieve highest security of data on cloud. So we proposed a new framework and an Encryption Schemes which encrypt the data and retrieve the data efficiently.

Keywords : Data leakage, Ram usage, Data usage.

I. INTRODUCTION

Cloud computing is advanced technology for IT enterprise. It has different advantages like virtualization, multi-user, scalability and so on. It also gives on demand computational infrastructure which has the power to decrease the cost to build the IT based services. It can provide numerous kind of service over the internet. One of the vital services is provided by the cloud is storage where users can keep their data as per the requirement. Though there are numerous types of services are providing by the cloud but Data storage is one of the upcoming features which is providing by the cloud to the users or any other companies. But due to the lack of proper security control policy and weakness in protection, many client are not ready to implement cloud computing technology.

So, data security is a very crucial task of good quality of services in cloud. Cloud computing faces the challenge of security threats for number of reasons. Firstly adopting the traditional cryptographic approach for the security of data in cloud computing is a threat as the data are stored in remote location and users do not have any control on it. So, it requires a data verification approach and it has no explicit idea about the whole data. So, it is very complex to verify the actual data. It is very tough to verify the correctness of data storage in the cloud as it is located in third party's location.

Stored data in third-party data warehouse may be frequently updated by the user, including updating, removing, insertion, and appending, restoring and other operation. So, we require a more dynamic enhanced technology operation to prevent data leakage from the cloud storage. Lastly, but it is not the last as data centers which are running in parallel in distributed way and all data are stored in different physical locations, so it is very vital to give correctness assurance in the distributed protocols. An encryption schemes based framework is presented and these schemes can hide the plaintext in cipher text which can store to the cloud. It is creating a secret key of small size which is suitable for data centric application. These should protect an unauthorized user to access data from cloud storage. These schemes incorporate storing data and retrieve data efficiently.

Existing system uses Watermarking to identify data leakage. Watermark is inserted in each file distributed over network if it is found to unauthorized user then that user is source to leak data.

Data leak detection variant from the anti-virus (AV) scanning for scanning file systems for malware signatures or the network intrusion detection systems (NIDS). AV and NIDS typically employ automata-based string matching e.g., Aho - Corasick, Boyer-Moore, which match static or regular patterns. However, data leak detection imposes new security requirements and algorithmic challenges:

1. Detection of data leak

Sometimes a data distributor gives sensitive data to one or more third parties. Sometime later, some of the data is found in an unauthorized place (e.g., on the web or on a user's laptop). The distributor must then investigate the source of the leak.

Data leakage is stated as the unauthorized transfer of classified information from a computer or datacenter to the outside world. Data leakage can be accomplished by simply mentally remembering what was seen, by physical removal of tapes, disks and reports or by subtle means such as data hiding.

2. Data Transformation:

The exposed data in the content may be unpredictably transformed or modified by users, and it may no longer be identical to the original sensitive data, e.g., insertions of metadata or formatting tags, substitutions of characters, and data truncation (partial data leak). Thus, the detection algorithm needs to recognize numerous type of sensitive data variations.

3. Scalability: The heavy workload of data leak screening is due to two reasons.

a) Long Sensitive Data Patterns: The sensitive data for instance customer information, documents, source code can be of arbitrary length.

b) Large Amount of Content: The detection needs to rapidly screen content instance gigabytes to terabytes. Traffic scanning is more time sensitive than storage scanning, because the leak needs to be discovered before the message is transmitted. Directly applying automata-based string matching to data leak detection is inappropriate and inefficient, because automata are not designed to support unpredictable and arbitrary pattern variations. In data leak detection scenarios, the transformation of leaked data is not known to the detection method. Creating comprehensive automata models covering all possible variations of a pattern is infeasible, which leads to O(2n) space complexity or O(2n) time complexity where n is the number of automaton states. Therefore, automata approaches cannot be used for detecting long and transformed data leaks. The proposed work is based on two algorithms i.e., RTU and DTU.

How data leakage occurs?

In the course of business, data must be handed over to trusted third parties for some operations. Sometimes these trusted third parties may act as points of data leakage. For instance, a hospital may give patient records to researchers who will invent new treatments. Similarly, a company may have partnerships with other companies that require to share customer data among them.

Network security consists of the policies and practices adopted to prevent and monitor unauthorized access, misuse, modification, or denial of a computer network and network-accessible resources. It consist the authorization of access to data in a network, which is controlled by the network admin. Users select or are assigned an ID and password or other authenticating information that allows them access to information and programs within their authority. Network security covers a variety of computer networks, both public and private, that are used in everyday jobs; conducting transactions and communications among businesses, government agencies and individuals. Networks can be private, such as within a company, and others which might be open to public access. Network security is involved in organizations, enterprises, and other types of institutions. It does as its title explains: It secures the network, as well as protecting and overseeing operations being done. The most common and simple way of protecting a network resource is by assigning it a unique name and a corresponding password.

Traditionally, leakage detection is handled by watermarking, e.g., 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. Watermarks 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. E.g. A hospital may give patient records to researchers who will devise new treatments. Similarly, a company may have partnerships with other companies that require sharing customer data. Another enterprise may outsource its data processing, so data must be given to various other companies. We call the owner of the data the distributor and the supposedly trusted third parties the agents.

The section I explains the Introduction of diabetic retinopathy prediction using methods naïve bays and neural network. Section II presents the literature review of existing systems and Section III present proposed system implementation details Section IV presents experimental analysis, results and discussion of proposed system. Section V concludes our proposed system. While at the end list of references paper are presented.

II. LITERATURE REVIEW

In paper [1] author discussed the issues: A data distributor has given sensitive data to a set of supposedly trusted agents. Some of the data are leaked and found in an unauthorized place. The distributor must assess the likelihood that the leaked data came from one or more agents, as opposed to having been independently gathered by other means. We propose data allocation strategies near the agents that enhance the probability of identifying leakages. These methods do not rely on alterations of the released data e.g., watermarks. In some cases, author also inject “realistic but fake” data records to further improve our chances of detecting leakage and identifying the guilty party.

Priyanka Barge et al. [2] present concept of data leakage, its effects of leakage and numerous techniques to recognize the data leakage. The value of the data is incredible, so it should not be leaked or changed. Huge database is being utilized in IT field.

This database is shared with multiple people at a time. But during this sharing of the data, there are huge chances of data vulnerability, leakage or alteration. So, to prevent these problems, a data leakage detection system has been proposed. This paper includes brief idea about data leakage detection and a methodology to detect the data leakage persons.

Sandip A. Kale et al. [3] present the results of implementation of Data Leakage Detection Model. Currently watermarking technology is being utilized for the data protection. But this technology doesn't provide the complete security against data leakage. This includes the difference between the watermarking & data leakage detection model's technology. This leads for the new technique of research for secured data transmission & detection, if it gets leaked.

In [8], author discussed an additional scenario that shows how the sharing of S objects by agents affects the probabilities that they are guilty. The scenario conclusion matches our intuition: with more agents holding the replicated leaked data, it is harder to lay the blame on any one agent.

R. Sion et al [9] deals with the idea of generating bit patterns on the file at certain location and all the bit patterns merged and make a watermark. The bits inserted are set of numbers which provide right protection to the data that is present in the data base. This also deals with the development of watermark detection application which reads the algorithms of the bit pattern by locating the markings and retrieves the original data at the client side.

This implemented on the technique of watermarking [10] the data utilizing multi-media watermarking technology to prevent the digital content going vital on net by disabling the copy facility. Encryption of the data has its own constraint from protecting the information. If the rights are decrypted then the data cannot be protected from illegally replicating the digital content. But this encryption issues is resolved by sung digital watermark which is embedded on the host data and cannot be eliminated and it includes the copyrights, data protection and monitoring and tracking.

Author [11] deals with generalization and suppression techniques to protect the data from leakage using K-anonymity privacy protection. Where every part of the data is categorized into k numerous subsets and every subset is linked with specific set of details and the final data is obtained at the external source. This technique is a failure as it lacks in clear description on how the data is being secured and what happens to the data if they are not systematically liked to one another.

Ms. Patil Rashmi, Mr Sangve [12] proposed improved Remote Data Possession Checking protocol based on homomorphic hash algorithm. Author syas proposed system supports secure and efficient dynamic operations at block level. Dynamic opreation consist insert, delete, update, and modify. They utilized Merkle Hash Tree to find the location of each data. A third party auditor can also be called as trusted party auditor (TPA) checks the user's data stored in cloud storage for its correctness and accuracy. A third party ensures correctness of user's data. Many times verification is allowed without the requiring the verifier to compare against the original data.

III. SYSTEM ARCHITECTURE

A. Architecture Overview

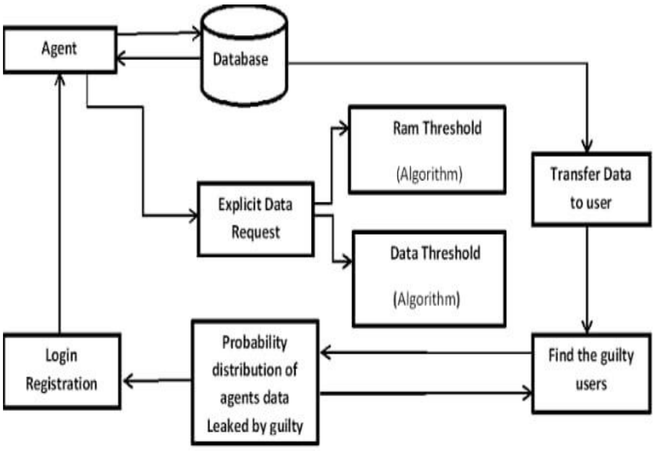

Fig 1. System Architecture

The proposed system is a three tier application that makes data access secure through a safe channel monitoring system. The resource manager makes the exact calculation of the RAM and CPU usage for the single data user and thereby predicts the usage for registered users. If the number of users are more than the registered users, data leakage is detected and the data service is stopped.

B. Algorithms Used

1) AES Encryption algorithm

Step 1: Byte Substitution (Sub Bytes)

The 16 input bytes are substituted by looking up a fixed table (S-box) given in design. The result is in a matrix of four rows and four columns.

Step 2: Shift rows

Each of the four rows of the matrix is shifted to the left. Any entries that 'fall off' are re-inserted on the right side of row. Shift is carried out as follows –

- First row is not shifted.
- Second row is shifted one (byte) position to the left.
- Third row is shifted two positions to the left.
- Fourth row is shifted three positions to the left.
- The result is a new matrix consisting of the same 16 bytes but shifted with respect to each other.

Step 3: Mix Columns

Each column of four bytes is now transformed using a special mathematical function. This function takes as input the four bytes of one column and outputs four completely new bytes, which replace the original column. The result is another new matrix consisting of 16 new bytes. It should be noted that this step is not performed in the last round.

Step 4: Add round key

The 16 bytes of the matrix are now considered as 128 bits and are XORed to the 128 bits of the round key. If this is the last round, then the output is the ciphertext. Otherwise, the resulting 128 bits are interpreted as 16 bytes and we begin another similar round.

C. Advantages of AES Algorithm:

- Symmetric key symmetric block cipher
- 128-bit data, 128/192/256-bit keys
- Stronger and faster than Triple-DES
- Provide full specification and design details.

IV. RESULT AND DISCUSSION

D. Experimental Setup

All the experimental cases are implemented in Java in congestion with Netbeans tools and MySql as backend, algorithms and strategies, and the competing classification approach along with various feature extraction technique, and run in environment with System having configuration of Intel Core i5-6200U, 2.30 GHz Windows 10 (64 bit) machine with 8GB of RAM.

E. Comparison Results

This section presents the RAM usage in system. Fig 2 and Fig 3. Shows RAM Usage Statistics of existing system and proposed system. X-axis shows Time required in seconds & Y-axis shows RAM utilized in bytes. In Existing System RAM utilized is same bytes for all seconds but proposed system initially utilize less RAM when work is less after that utilize 2251 bytes of RAM. So according to comparison results proposed system utilizes less RAM as compared to existing system.

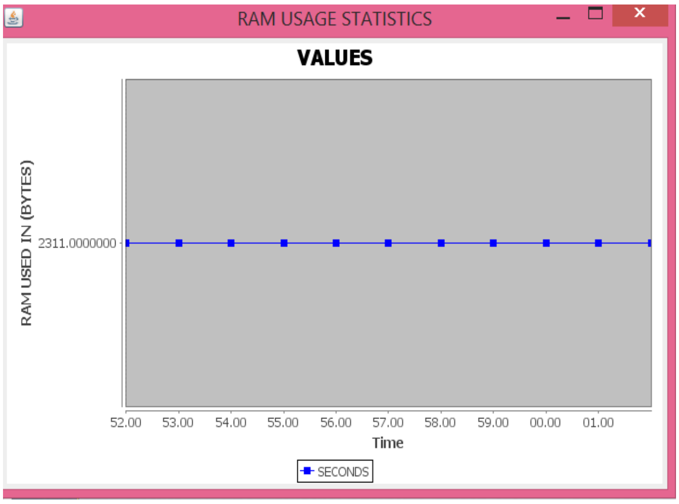

Fig. 2: RAM Usage Statistics in Existing System

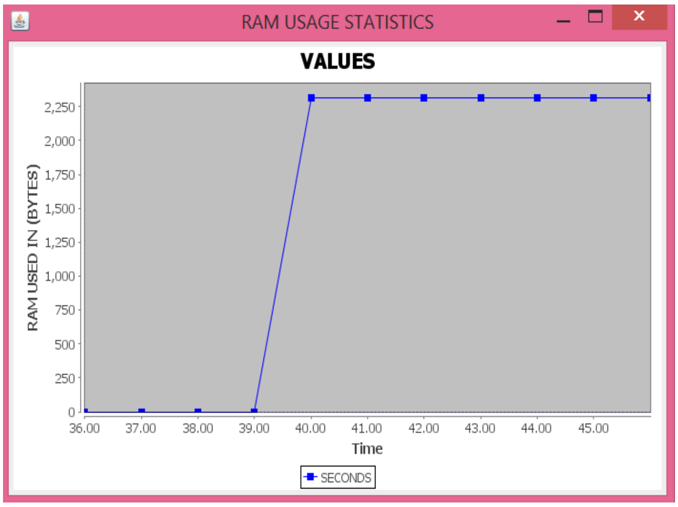

Fig. 3: RAM Usage Statistics of Proposed System

V. 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 and various offices can rely on this security detection model.

VI. REFERENCES

[1]. Data Leakage Detection, Panagiotis Papadimitriou, Student Member, IEEE, and Hector Garcia-Molina, Member, IEEE,

[2]. A novel data leakage detection, Priyanka Barge, Pratibha Dhawale, Namrata Kolashetti3 Ass. Prof., Department of Computer Engineering, NIRMALA CHOUHAN International Journal of Modern Engineering Research (IJMER) Vol.3, Issue.1, Jan-Feb. 2013 pp-538-540 ISSN: 2249-6645

[3]. Data leakage detection, Sandip A. Kale, Prof. S.V.Kulkarni Department Of CSE, MIT College of Engg, Aurangabad, Dr. B. A. M. University , Aurangabad (M.S), India International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 9, November 2012

[4]. Network security using cryptographic techniques, Sumedha Kaushik, Ankur Singhal, Department of ECE & M.M. university Ambala

(Haryana) India, volume 2, issue 12, December 2012, ISSN:2277 128X

[5]. Information Hiding in Images Using Steganography Techniques Ramadhan Mstafa, Christian Bach 2013 ASEE Northeast section conference, Norwich university, reviewed paper, March 14-16, 2013

[6]. Digital forencis and Preservation, Jeremy Leighton John, DPC technology watch report 12-03 November 2012.

[7]. Steganography and its application security, Ronak Doshi, Pratik Jain, Lalit Gupta, Department of Electronics and Telecommunication, Pune University, India.

[8]. Mr.V.Malsoru, Naresh Bollam / REVIEW ON DATA LEAKAGE DETECTION , International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 1, Issue 3, pp.1088-1091 1088.

[9]. R. Sion, M. Atallah, and S. Prabhakar, •\Rights Protection for Relational Data, Proc. ACM SIGMOD, pp. 98-109, 2003.

[10]. Hartung and kutter,. Watermarking technique for multimedia data.2003.

[11]. L. Sweeney, ―Achieving K-Anonymity Privacy Protection Using Generalization and Suppression, http://en.scientificcommons.org/43196131, 2002.

[12]. Patil .R.Rashmi and S. M. Sangve, "Public auditing system: Improved remote data possession checking protocol for secure cloud storage," 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Davangere, 2015, pp. 75-80.

Cite this article as :

Monali U. Pawar, Shraddha A. Mankar, Snehal S. Mandhare, Siddhi N. More, Rashmi R. Patil, "Enhancement of Data Leakage Detection Using Encryption Technique", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 6 Issue 3, pp. 10-16, May-June 2019.

Journal URL : http://ijsrst.com/IJSRST19638