IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

A Review on Data Leakage Detection using Cloud Computing

Vijaya Balpande[1], Pranjal Kature[2], Mitesh Wadibhasme[3], Gitika Kosulkar[4], Shubhangi Bansod[5]
Professor, Department of Computer Science & Engineering[1]
Researcher, Department of Computer Science & Engineering[2,3,4,5]
Priyadarshini J. L College of Engineering, Nagpur, Maharashtra, India

Abstract: *Data is a valuable asset and the intellectual property of many different enterprises. Every organisation keeps sensitive information such as customer information, financial information, patient information, credit card information for individual customers, and other information based on the management style, institute, or industry. Information leaking is a major issue that the company must deal with in situations like this since it may be very costly. More specifically, information leakage is defined as the deliberate disclosure of any type of information to unauthorised outside parties. when sensitive information is transferred to unauthorised parties or moves to an unapproved location. This will result in a certain industry suffering direct and indirect losses in terms of money and time. Vulnerability or its alteration are results of information leakage. Information can therefore be shielded from outside leaks. An efficient and effective mechanism to prevent and protect allowed information is required to resolve this problem. The suggested model also encrypts the data before it is downloaded via the TPA, and it sends the secret key to the original user after a successful download without being intercepted by the TPA.*

Keywords: Cloud Computing, Data Leakage, Data Security, Data analysis, AES algorithm

I. INTRODUCTION

The transfer of information, or the sharing of data from one person to another, is a common occurrence in today's world. As the data communicated with the trusted third parties are confidential and very significant, the distributor's data transmissions must be secure, private, and never repeated. When data is uploaded to the cloud, it may become vulnerable to data leakage if it is accessed, read, or altered by any user. This section provides an overview of the subject and the significance of leak detection. The several types of data loss or leakage, including intentional, unintentional, disaster, failure, and crime are also covered. Types of intentional exploitation are illustrated because the research aims to identify deliberate data leakages from an inside agent. This research is intended to show how user activities can be monitored, through the creation of an application that tracks user actions, through system transaction logs to identify data leakage channels and by whom, and through user access to a computational resource. Moreover, a transaction log will be built and used to track user activity, including login times and activities.

Information security, computer security, and IT businesses are seeing an increase in cloud computing security. While considering the implementation of cloud computing, one of an organization's top concerns is avoiding unwanted access to data resources. In this study, we explore the security features of cloud computing with the goal of improving data management system accountability and safeguarding client privacy.

II. LITERATURE SURVEY

Prashant Khobragade and Chandu Vaidya are [1] mention the issues with data storage and cloud security that affect the majority of distributed systems.This paper majorly focuses on cloud data storage security for client, which has always been an most aspect of quality of service. For ensuring the correctness of cloud clients data in the cloud, in this paper propose a encryption of client data with cryptographic algorithm. And there was a problem with the location of the attack, which may be a file, a file location, or browser history. It failed to provide the proper file location. Data are gathered from many data sources and stored in databases as proof.

The Gagdeep Kaur, Dr. Sandeep Kaustish [2] was focus on security issues in the cloud storage & they also Analyse error occur or problem in server while fetching the data from data leakage detection system. He used SHA 128 bit ,but

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

it is less secure because it will perform only 8 Rounds & he get problem also in server while fetching the website for that he used SMTP server to transfer the data over server & this server response rate is slow.

Prica I Okachi and Staley Okachi[3] describe the issues based on data fetching problem.The systems back-end was created using Microsoft SQL server. Additionally, this paper were identify the efficiency problem and the data processing speed between two panel is slow. It involves some modification of data i.e. making the data less sensitive by altering attributes of the data. He used a structured approach for analyzing and designing a system. It applies object oriented concepts and develop a set of graphical system model during the development life cycle of the software.

III. METHODOLOGY

3.1 Existing Methodology

Fake Object Method

The distributor creates and adds fake objects to the data that he distributes to agents. Fake objects are objects generated by the distributor in order to increase the chances of detecting agents that leak data. The distributor may be able to add fake objects to the distributed data in order to improve his effectiveness in detecting guilty agents. Our use of fake objects is inspired by the use of "trace" records in mailing lists. In case we give the wrong secret key to download the file, the duplicate file is opened, and that fake details also send the mail. Ex: The fake object details will display.

SHA Algorithm

While downloading data using a third-party agent and giving it to the actual cloud, the secure SHA algorithm provides a hash function. To increase the security of sensitive data transfer via the Internet, this security feature has been included. Because the SHA algorithm uses message digests to guarantee data integrity and confidentiality, it is widely utilized.

3.2 Proposed System

With the AES method, data is protected against illegal access and extraction while also allowing for the identification of leaks. AES is an encoding system that converts plain text data into cypher text, which neither humans nor machines can decipher without the use of an encryption key—a password. Phases like the data gathering phase and the analysis phase are included in the data leakage detection procedure. The data collection is designed to give a user or agent the authority to access the other project modules.This system alerts the user about the leakage occurs during transmission of data.It also helps to know the date, time, file name, user mail and file name of the leaked data. The advanced encryption standard (AES) uses an expansion process to generate a key schedule. AES-256 encryption uses the 256-bit key length to encrypt as well as decrypt a block of messages. There are 14 rounds of 256-bit keys, with each round consisting of processing steps that entail substitution, transposition, and mixing plain-text to transform it into cipher text.

The AES-256 algorithm is composed of three main parts: Cipher, Inverse Cipher and Key Expansion. Cipher converts data to an unintelligible form called cipher text while Inverse Cipher converts data back into its original form called plain-text. Key Expansion generates a Key Schedule that is used in Cipher and Inverse Cipher procedure. Cipher and Inverse Cipher are composed of specific number of rounds. The encryption algorithm encrypts one block of data at a time to produce the encrypted data block with the use of a secret key. The decryption is simply the reverse process of the encryption, and each operation is the inverse of the corresponding one in encryption. The data block length is fixed to 128 bits, while the key length can be 128, 192, or 256 bits. Each data block is rearranged in a matrix form. AES algorithm is an iterative algorithm and each iteration is called a round. Each round is iterated 14 times for 256 bit key with 8 bytes in a row of key lengths. Each round uses four transformations and inverses but final round excludes Mix Columntransformations.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

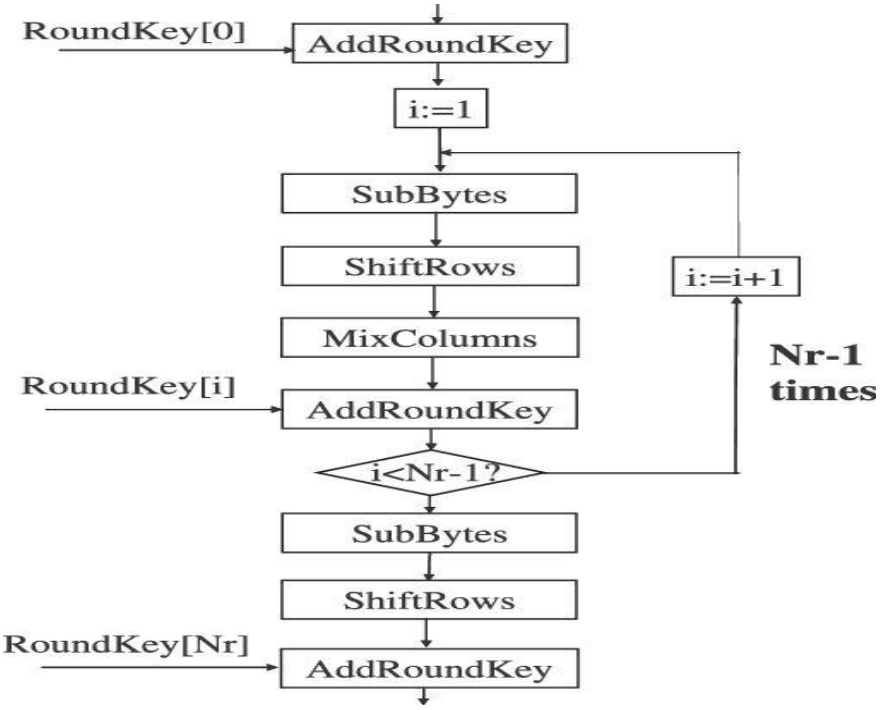

Fig1. Implementation of AES 256 Algorithm

1. Byte Substitution (Sub Bytes)

In this step, it converts each byte of the state array into hexadecimal, divided into two equal parts. These parts are the rows and columns, mapped with a substitution box (S-Box) to generate new values for the final state array.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.

2. Shiftrows

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.

3. MixColumns

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.

4. Addroundkey

The respective key for the round is XOR'd with the state array is obtained in the previous step. If this is the last round, the resultant state array becomes the cipher-text for the specific block; else, it passes as the new state array input for the next round

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

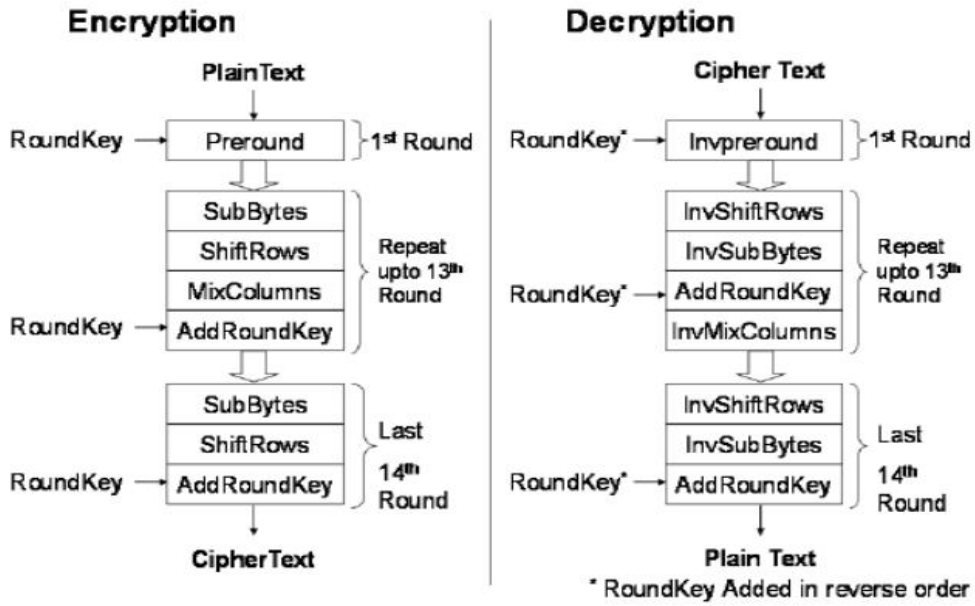

Fig 2. Encryption and Decryption of the data

IV. CONCLUSION

Organizations in the real world struggle with data leaking. We draw the conclusion from this study that the data leakage detection system is very helpful for preventing the unauthorized usage of data across various industries. So, it is necessary to create a content inspection approach that can find crucial data breaches in file or network traffic content. This data leakage system will be effective at stopping data breaches, securing the data with encryption and notifying data owners when a threat of a data leak is present. Data protection from leaks of any form is crucial. With the help of AES algorithm, data leakage can be detected in a more flexible manner than with current systems

REFERENCES

[1]. Chandu Vaidya and Prashant Khobragade, "Data Security in Cloud Computing", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 5 June 2020

[2]. Gagdeep Kaur and Dr. Sandeep Kautish "Data Lekage Deteaction Using Cloud Computing", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2454-6615, Volume: 06 Issue: 03, Mar 2019.

[3]. Prisca I. Okochi, Stanley A. Okolie 1 and Juliet N. Odii 1,"An improved data leakage detection system in a cloud computing environment", World Journal of Advanced Research and Reviews, 24 August2021, 11(02), 321–328

[4]. International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653;Volume 10 Issue IV April 2022.

[5]. R. Naik and M. N. Gaonkar, "Data Leakage Detection in cloud using Watermarking Technique," 2019 International Conference on Computer Communication and Informatics (ICCCI), 2019, pp. 1-6.

[6]. Lord, N. (2019). Data Protection: Data In transit vs. Data at Rest. Retrieved 24 2019, from https://digitalguardian.com/blog/data-protection-data-in-transit-vs-data-at-rest.

[7]. El Sharif Karrar, D., & Idris Fadl, M. (2020). Security Protocol for Data Transmission in Cloud Computing. International Journal of Advanced Trends in Computer Science an Engineering, 7(1), 1-5. doi: 10.3053/ijatcse/(2018)

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

[8]. N.Sandhya , K. Bhima , G. Haricharan Sharma," Data Leakge Detection using cloud computing".International Journal of Electronics Communication and Computer Engineering2020, Volume 3, Issue 1

[9]. Yasmin R., Memarian, M.R., Hosseinzadeh, S., Conti, M., & Leppänen, V. (2018). Investigating the Possibility of Data Leakage in Time of Live VM Migration. In: Dehghantanha A., Conti M., Dargahi T. (eds) Cyber Threat Intelligence. Advances in Information Security, vol 70. Springer, Cham.

[10]. Herrera Montano, I., García Aranda, J.J., Ramos Diaz, J. et al. Survey of Techniques on Data Leakage Protection and Methods to address the Insider threat. Cluster Comput 25, 4289–4302 (2022).