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: 7.53
Volume 4, Issue 4, March 2024

Utilizing Watermarking Technique for Detecting Data Leakage in Cloud Environment

Sibi Sharan S[1] and Haripriya V[2]
PG Student, Department of MSc CS-IT[1]
Assistant Professor, School of CS & IT[2]
Jain (Deemed-to-be University), Bangalore, India
[1]isibisharan@gmail.com and [2]v.haripriya@jainuniversity.ac.in

Abstract: *Ensuring safety in dossier management is paramount due to the immense value of stored information. While hackers are often attributed to security breaches, the reality is that a significant portion of data loss stems from insiders. In traditional setups, the transfer of critical data from suppliers to trusted entities is a frequent occurrence. Preserving the security and integrity of these transactions is crucial to meeting the increasing demands of consumers. Any leakage of sensitive data exposes customers to potential risks from the outset. Therefore, establishing secure channels for data transfer between suppliers and recipients is imperative. This project proposes a solution for detecting data leaks using watermarking technology, which detects tampering attempts and identifies the source of leaked information. The system operates within a cloud environment, ensuring accessibility and scalability.*

Keywords: Watermark, Data leakage, Tampering, Steganography, Cloud, AES, QR code, DCT, DWT, SVD.

I. INTRODUCTION

Distributed computing is quickly arising as a progressive innovation in the field of information handling, with essentially all IT associations endeavoring to embrace it. In distributed computing, shared assets, for example, information and programming are given to clients on request. One of the key benefits presented by the cloud is information stockpiling and access. By utilizing the cloud, clients are liberated from the limitations of neighborhood information capacity and recovery. Notwithstanding, this comfort likewise represents a critical danger to information security. While cloud servers oversaw by suppliers may not be completely trusted by clients, the information put away in the cloud can be delicate and private, like monetary examinations, business techniques, and so forth. Subsequently, guaranteeing information security and protection in distributed computing has turned into an issue of foremost significance. The absence of command over information can prompt different security issues and dangers, including information breaks. The seriousness of the harm brought about by an information break relies upon the responsiveness of the spilled information. On the off chance that the spilled information is exceptionally important to the association, it can bring about huge misfortunes and reputational harm. To address this test, a few strategies for information spillage recognition have been created, including peculiarity location, interruption discovery, and so on. Every strategy plan to identify information spillage in view of explicit examples or peculiarities in the information. The proposed project centers around recognizing information spillage in cloud conditions with a lot of information. The current framework uses a procedure known as watermarking, where remarkable identifiers are implanted in each sent duplicate of touchy information. In the event that a duplicate is viewed as gotten to by unapproved faculty, the leaker can be recognized. Furthermore, the information is inspected for indications of altering or control. Generally, the projected venture stresses the recognition of information leakers and altering involving watermarking procedures to guarantee information uprightness and privacy in distributed computing conditions

II. LITERATURESURVEY

Panagiotis Papadimitriou [1] proposes a process that involves data allocation. Distributors strategically deliver papers to clients in order to maximize the possibility of finding a client who is guilty. The distributor creates fake items, which

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: 7.53
Volume 4, Issue 4, March 2024

don't exist in the original dataset. To increase the likelihood of identifying clients accountable for data leaks, these objects—which are made to resemble real items—are sent to clients alongside real data objects. However, adding fictitious objects could affect how accurately client action's function, thus it might not always be acceptable. This method shows that it is possible to determine a client's liability for a leak by looking at the data overlap with the information that was disclosed.

A technique that highlights the significance of watermarking sensitive data before distribution is presented by Abhijit Singh et al. [2], making it possible to pinpoint the source of the data with complete certainty. By making data less accessible and making it easier to identify dishonest customers through the use of phony objects—watermarks positioned at various points throughout the data—watermarking improves data security. The method investigates watermarking strategies like the least significant bit (LSB) [3], in which the watermark is placed in an audio file by using the LSB algorithm and encoded using the RSA [4] algorithm. This encryption technique ensures notable robustness by making it extremely difficult to remove the watermark.

The Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS), as described by S. Geetha et al. [5], entails splitting up an owner's data into different portions and replicating them throughout the cloud. With this approach, the file is divided into several parts and dispersed over various nodes rather than being stored in its whole in the cloud. Because a fragment's similarity to the spilled data allows for identification of the uninvited person responsible for a leak, fragmentation makes this possible.

Neeraj Kumar et al. [6] propose a technique where the server adds an image logo to the stored documents to embed confidential messages effectively. ASCII code is added to documents and AES is applied with SHA-512 authentication hash algorithm. The focus is on limiting access to crucial documents to approved clients only. The method relies on symmetric key encryption, making it unsuitable for web-based scenarios where multiple parties may access the data.

Rupesh Mishra [7] suggests data allocation strategies that can increase the likelihood of detecting a guilty client. A random number of data objects are distributed among clients. Analysis reveals that fake objects have a significant impact on identifying corrupt clients, measurable based on the similarity between leaked data and client data.

AL. Jeeva [8] provides a comparative analysis of encryption algorithms based on parameters such as encryption ratio, speed, tunability, power consumption, hardware-software implementations, and key length. The Advanced Encryption Standard (AES) is favored among symmetric encryption techniques due to its lower energy consumption, reduced buffer usage, and shorter encryption and decryption times.

Weijun Zhang [9] introduces a mechanism where data to be embedded as a watermark is translated into a QR code prior to embedding.

III. OBJECTIVES

1. To make sense of the significance of Analyzable man-made intelligence (XAI) and interpretability in improving dependability and straightforwardness inside computerized reasoning frameworks.
2. Objective: To introduce a synopsis of current XAI and interpretability procedures, zeroing in on their application in AI models, especially profound learning models.
3. To investigate techniques, for example, highlight representation, saliency maps, choice trees, and model refining, evaluating their viability in accomplishing XAI and interpretability.
4. To examine the advantages and downsides of each XAI procedure, taking into account their reasonableness for various application necessities and settings.
5. To distinguish remaining difficulties in the field, including the requirement for normalized measurements to assess model interpretability and guarantee the dependability of clarifications gave.
6. To advocate for the headway of XAI and interpretability strategies as significant parts in laying out trust in computer based intelligence frameworks and driving advancement in the field of man-made consciousness.
7. To give bits of knowledge and suggestions for future examination headings focused on additional working on the unwavering quality, receptiveness, and interpretability of man-made intelligence frameworks.

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: 7.53
Volume 4, Issue 4, March 2024

IV. EXISTING MODEL

Existing models in the space of Information Spillage Location in cloud utilizing Watermarking Strategy display an expansive range of approaches and procedures pointed toward strengthening information security and safeguarding secrecy. These models epitomize a multi-layered scene of procedures, going from installing exceptional identifiers inside delicate information to work with the following of spilled data sources, to utilizing division and replication techniques for ideal execution and security improvement. Inside this far-reaching collection, a few models dig into many-sided watermarking techniques, like least huge piece (LSB) encoding and RSA encryption, to invigorate information against unapproved access and altering endeavors, consequently supporting the flexibility of information respectability systems. Besides, headways in picture watermarking procedures have opened roads for implanting private messages straightforwardly into reports, in this manner enlarging the shields against information breaks. These models complicatedly wind around together cryptographic standards, like symmetric key encryption, with creative watermarking procedures to lay out powerful information assurance systems. By interlacing encryption calculations like the High-level Encryption Standard (AES) with modern watermark installing systems, these models guarantee that delicate data stays protected from unapproved access while saving the productivity and versatility essential for cloud-based conditions. Besides, information portion procedures, described by the dissemination of randomized information objects among clients, elevate the adequacy of recognizing guilty gatherings, consequently enlarging the accuracy of information spillage discovery instruments.

V. DISADVANTAGES OF EXISTING MODEL

Although the models now in use for cloud-based data leakage detection utilizing the watermarking technique present encouraging ways to improve data security, they are not without drawbacks. The possible influence on resource usage and system performance is one of the main drawbacks. Particularly in large-scale cloud systems, adding watermarks or encrypting data can result in computational overhead that increases processing times and resource consumption. This may lead to problems with latency and a reduction in system responsiveness, which may eventually affect productivity and user experience. Furthermore, it can be difficult for current models to strike a compromise between security and usability. Although watermarking and strong encryption might improve data security, they can also make data access and retrieval more difficult. This complexity may make it more difficult for authorized users to collaborate and share data easily, requiring more time and money to properly maintain access permissions and decryption keys. The possible vulnerability to hostile attacks and evasive strategies is another disadvantage. Even though encryption and watermarking are intended to protect data from manipulation and unwanted access, these security measures can still be circumvented or manipulated by clever attackers. Encryption and watermarking algorithms can include flaws that adversaries could use to remove watermarks or decrypt protected data, jeopardizing the integrity and secrecy of sensitive data. Furthermore, the scalability of existing models may pose a challenge, particularly in dynamic cloud environments with fluctuating data volumes and user demands. Scaling watermarking and encryption techniques to accommodate growing data sets and user populations can be complex and resource-intensive, potentially leading to scalability bottlenecks and performance degradation. Overall, while existing models offer valuable insights and solutions for data leakage detection in the cloud, addressing these disadvantages will be crucial for ensuring the effectiveness, efficiency, and scalability of data security measures in evolving cloud computing ecosystems.

VI. PROPOSED MODEL

The proposed model for Data Leakage Detection in the cloud using Watermarking Technique endeavors to address the shortcomings of existing approaches while advancing the frontier of data security and integrity in cloud computing environments. At its core, the proposed model integrates cutting-edge watermarking methodologies with advanced encryption techniques to fortify data against unauthorized access, tampering, and leakage. Unlike conventional watermarking techniques, which may introduce computational overhead and usability challenges, the proposed model leverages innovative algorithms and optimizations to minimize performance impact while maximizing security efficacy.

The creation of resilient watermarking systems that include distinct IDs or markers into sensitive data without sacrificing system performance or user experience is essential to the suggested approach. These watermarks are

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: 7.53
Volume 4, Issue 4, March 2024

purposefully made to be resistant to evasion tactics and hostile attacks, guaranteeing the integrity and traceability of data even in the face of complex dangers. The suggested architecture provides a multi-layered protection against data breaches and leakage occurrences by fusing watermarking with cutting-edge encryption techniques including homomorphic encryption and multi-party computation.In addition, the suggested approach places a strong emphasis on adaptability and scalability to satisfy the changing requirements of cloud computing environments. Scalable watermarking and encryption algorithms enable seamless integration with dynamic cloud infrastructures, accommodating growing data volumes and user populations without sacrificing security or performance. Additionally, the model incorporates mechanisms for efficient key management and access control, streamlining data sharing and collaboration while preserving confidentiality and privacy.Beyond technical innovations, the proposed model advocates for a holistic approach to data security that encompasses not only technological advancements but also organizational policies, regulatory compliance, and user awareness. By promoting a culture of security and accountability, the proposed model aims to foster trust and confidence in cloud computing systems, empowering organizations to leverage the full potential of cloud technologies while mitigating the risks of data leakage and unauthorized access.

VII. ADVANTAGES OF PROPOSED MODEL

The proposed model for Data Leakage Detection in cloud using Watermarking Technique offers a range of significant advantages, positioning it as an innovative solution in the domain of data security within cloud computing environments. Foremost among these benefits is the robustness and effectiveness of the model's watermarking and encryption techniques. By employing advanced algorithms and optimizations, the model ensures seamless embedding of watermarks into sensitive data while minimizing computational overhead and performance impact. This capability guarantees the security and traceability of data, even against sophisticated attacks and evasion strategies.Furthermore, the model boasts exceptional scalability and adaptability, seamlessly integrating with dynamic cloud infrastructures and meeting the evolving demands of modern organizations. Large data volumes may be securely stored and sent thanks to its scalable watermarking and encryption algorithms, and its effective key management mechanisms make access control and data sharing procedures easier to carry out. The model's efficacy is ensured by its scalability in a variety of application contexts, ranging from small-scale deployments to cloud systems at the enterprise level. The paradigm balances security and accessibility while giving priority to user experience and usability in addition to technological capabilities. User productivity and efficiency are not compromised when it comes to safely accessing and sharing sensitive data thanks to user-friendly interfaces and efficient procedures. By fostering a culture of security awareness and accountability inside businesses, this user-centric approach makes it possible to effectively enforce compliance standards and data protection rules.Moreover, the model includes comprehensive governance and regulatory compliance features, ensuring adherence to industry-specific regulations and data protection laws. The model incorporates governance systems pertaining to audits, monitoring, and reporting, which empowers firms to exhibit adherence to rigorous regulatory requirements and minimize the potential legal liabilities that may arise from data breaches. All things considered, the suggested model offers a comprehensive and reliable framework for protecting sensitive data in cloud computing settings, which is a major development in data leakage detection and prevention. The model sets a new standard for data security and integrity by fusing cutting-edge encryption and watermarking techniques with scalable infrastructure and user-centric design principles. This allows organizations to fully utilize cloud technologies while reducing the risks of data leakage and unauthorized access.

VIII. TECHNIQUES

The suggested model for Data Leakage Detection in the Cloud Using Watermarking Technique incorporates a number of important strategies, all of which add to the thorough framework for protecting sensitive data. Watermarking is the most advanced of these methods; it is the process of adding distinct markers or identifiers to data in order to aid in authentication and traceability. In order to protect data from unwanted access and manipulation efforts, advanced watermarking methods are used, such as least significant bit (LSB) encoding and RSA encryption, guaranteeing the integrity and confidentiality of information. To improve data security and privacy, encryption techniques are used in conjunction with watermarking. Data is protected both in transit and at rest by using cutting-edge encryption techniques like multi-party computing and homomorphic encryption. The concept guarantees the security of sensitive information

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: 7.53
Volume 4, Issue 4, March 2024

by encrypting it using strong cryptographic keys, even in the event of illegal access or interception. Furthermore, the model incorporates segmentation and replication strategies to optimize data storage and access in cloud environments. By segmenting owner's files into various chunks and replicating them across distributed cloud nodes, the model enhances data availability and resilience while mitigating the risk of data loss or corruption. This fragmentation of data enables efficient distribution among clients while facilitating the detection of unauthorized access or leakage incidents based on the similarity of data fragments with leaked information.

Furthermore, the suggested approach uses sophisticated authentication techniques to impose strict access controls and governance standards, like digital signatures and access control lists. The strategy lowers the risk of insider threats and unauthorized disclosures by ensuring that only authorized individuals can access sensitive data through extensive authentication and authorization mechanisms.

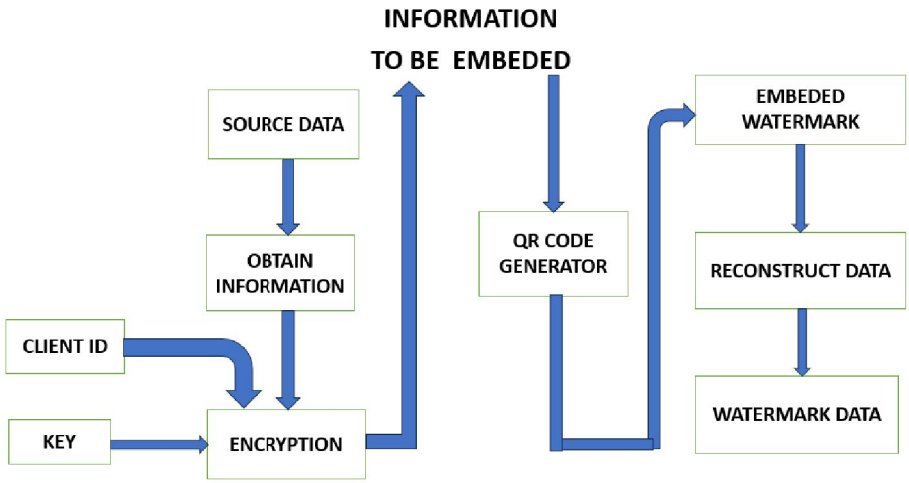

Figure1

Overall, the integration of these techniques within the proposed model establishes a robust framework for detecting and preventing data leakage in cloud environments. By combining watermarking, encryption, segmentation, replication, and authentication mechanisms, the model provides a multi-layered defense against evolving cyber threats while safeguarding the confidentiality, integrity, and availability of sensitive information.

IX. CONCLUSION

A significant development in a variety of fields, including businesses, associations, academic communities, and establishments that share data online with outside parties, is the data leakage detection model. Thanks to a sophisticated watermarking method, its architecture provides a reliable way to detect instances of data leaks and tampering inside datasets saved in the cloud. This novel method uses image data as the information carrier and embeds a Quick Response (QR) code produced by a hybrid watermarking algorithm. The model extracts the watermark and cross-references it with the client's data to enable accurate detection of possible data leakers by integrating recipient client specifics into the watermark. Additionally, the system subjects the transmitted data to stringent integrity tests, comparing its attributes to those of the original dataset in order to look for any indications of tampering or unauthorized adjustments. This data leakage detection system is important because it provides complete end-to-end security during the data transfer process, hence providing excellent protection against any breaches and leaks. In contrast to traditional security methods that merely use encryption algorithms, the suggested approach provides strong detection capabilities in addition to guaranteeing data secrecy, improving overall data protection. The use of a hybrid watermarking technique, which has been carefully designed to strike a balance between resilience and imperceptibility, is essential to the model's effectiveness since it protects it from illegal access attempts and other types of data manipulation. Adopting this novel concept, in short, is a critical first step toward bolstering data security and privacy in the digital sphere and addressing the urgent need for all-encompassing solutions that can protect confidential data in online settings. The methodology provides organizations and entities with the ability to proactively limit the risks associated with data leakage and

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: 7.53
Volume 4, Issue 4, March 2024

tampering by utilizing advanced watermarking techniques and careful integrity checks. This approach promotes higher trust and confidence in online data sharing procedures.

REFERENCES

[1] PanagiotisPapadimitriou, HectorGarcia-Molin"Data Leakage Detection" IEEE Transactions on Knowledge and Data Engineering, 2011, Volume 23, Issue 1.
[2] Abhijeet Singh, Abhineet Anand, "Data Leakage Detection Using Cloud Computing" International Journal of Engineering and Computer Science, Volume 6, Issue 4, April 2017.
[3] Abdullah Bamatraf, Rosziati Ibrahim and Mohd, Najib Mohd Salleh, "A New Digital Watermarking Algorithm Using Combination of Least Significant Bit (LSB) and Inverse Bit", International Journal of computing, volume 3, Issue 4, April 2011.
[4] Xin Zhou, Xiaofei Tang, "Reasearch and Implementation of RSA Algorithm for Encryption and Decryption", Interational Forum on Strategic Technology, 2011 IEEE.
[5] S.Geetha, M.Nishanthini, G.Shanthi, K.Sivabharathi, M.Suganya "Data Leakage Detection and Security Using Cloud Computing", International Journal of Engineering Research and Applications, Volume 6, Issue 3, March 2016.
[6] Neeraj Kumar, Vijay Katta, Himanshu Mishra, Hitendra Garg, "Detection of Data Leakage in Cloud Computing Environment", International Conference on Computational Intelligence and Communication Networks, 2014 IEEE.
[7] Rupesh Mishra, D.K Chitre, "Data Leakage and Detection of Guilty Agent", International Journal of Scientific & Engineering Research, Volume 3, Issue 6, 2012.
[8] AL.Jeeva, Dr.V.Palanisamy, K.Kanagaram, " Comparative Analysis of Performance and Security Measuresof Some Encryption Algorithm", International Journal of Engineering Research and Applications, Volume 2, Issue 3.
[9] Sumit Tiwari, " An introduction to QR Code Technology" internatinal conference on information technology, 2016 IEEE.
[10] Yanqun Zhang, "Digital Watermarking technology: A Review" Internation Conference on Future Computer and Communication, 2009 IEEE.
[11] Gu Tianming ,Wang Yanjie, "DWT-based Digital Image Watermarking Algorithm", The Tenth International Conference on Electronic Measurement & Instruments, 2011 IEEE.
[12] Syed Ali Khayam, "The Discrete Cosine Transform (DCT): Theory and Application", March 2003.