Enterprise Data Breach: Causes, Challenges, Prevention, and Future Directions

Naveenraj D[1] and S Anu Priya[2]
PG Student, Department of Computer Applications[1]
Assistant Professor, Department of Computer Applications[2]
Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India
22304345@vistas.ac.in and anupriya.scs@velsuniv.ac.in

Abstract: *Cloud data is such a valuable and necessary resource, security is a major worry. Despite the widespread misperception that hackers are the source of security lapses, insiders are primarily responsible for data theft. In practically dispersed settings, critical data is routinely moved from the distributor to trustworthy parties. The stability and security of the services must be guaranteed in light of the increasing volume of user requests. When a client discloses important information, the client should be held accountable as soon as possible. Therefore, it's necessary to keep an eye on the data as it moves from the distributor to the agents. In the context of cloud computing, the project identifies data leakage detection, which examines data tampering and concludes that the information leak was caused by a particular employee in the organization*

Keywords: Data leakage Detection, Security, Authentication, Confidentiality

I. INTRODUCTION

Cloud computing is an innovative and rapidly advancing technology in the information technology domain. Nearly every IT firm is attempting to gain a foothold in this technological realm. Through cloud computing, shared software and information resources are provided on-demand to devices. One of the primary services offered by cloud computing is data storage. Employees are completely relieved from the cumbersome tasks of local data storage and maintenance, thanks to the cloud. However, it also poses a slight risk to file privacy. Data leakage is a significant concern in today's business landscape, as it must be protected against unauthorized access.Safeguarding critical data from misuse by unauthorized individuals is crucial. Information related to intellectual property rights, patents, functionalities, and other relevant details is vital. This critical organizational information has often been disseminated to various parties outside the company's boundaries. Consequently, identifying the person or entity responsible for the data leak becomes a challenge. When organizational data is leaked by an agent, the primary objective of the proposed task is to identify the culprit.

Cloud technology is entirely reliant on the internet, as it stores data in service providers data centres. One of the major challenges with cloud computing technologies is data security. Having less control over data can lead to potential data leakage, data insecurity, and data attacks by both internal and external actors. Every IT organization must focus on addressing the security challenges of protecting its data from various third parties to prevent data leaks. Since current employees tend to be the ones who occasionally carry out leaks, the security measures must be beyond their awareness, making it impossible for them to know how to breach it. Data leakage can occur at any time, and there is no set period when it will happen.

The extent of harm caused is solely determined by the quality of the private information that was disclosed by the individual. If the institution considers the disclosed material to be of critical importance, it can leave the institution vulnerable. The leakage can adversely affect sales and potentially lead to the company's downfall. Additionally, when cloud users access the self-care portal or dashboard, there is a possibility of data leakage during the authentication phase of a communication session.

Hence, the possibility of data leakage could result in security issues like a data breach on a cloud platform, which would compromise data confidentiality and legal compliance. While numerous studies have been conducted on security

DOI: 10.48175/IJARSCT-18403

risks and data leakage in the cloud, none of them have investigated data leakage and demonstrated how to detect data leakage on cloud computing platforms.

Users are not entirely confident in the cloud servers maintained by the cloud providers, as the data records stored in the cloud may contain sensitive and private information, such as employee or product details, business policies, etc. Consequently, data security in cloud computing has received significant attention. Inadequate control over data can lead to serious security problems and threats that could result in data leakage. The severity of the contamination caused by the data leak is determined by the quality of the sensitive data disclosed. If the information that was leaked is particularly important to the organization, it may leave the organization in a vulnerable state.

II. LITERATURE REVIEW

Panagiotis Papadimitriou, "presented Data allotment is a prime focus of approach"

It outlines a technique that distributors can use to carefully distribute documents to customers, increasing the chance that a dishonest customer will be exposed. The distributor creates fake items that are absent from the original data collection. To maximise the likelihood of identifying clients guilty for data breach, the objects are constructed to closely resemble real objects and supplied to clients alongside real data objects. Fake objects, however, would not always be permitted because they might impair the accuracy of the operations carried out by customers. By examining the potential that a client is responsible for the leak based on the overlap of data with the leaked data, this approach demonstrates that it is practical to do so.

AL. Jeeva, "provided comparative analysis of encryption algorithms"

Comparisons were made based on factors such key length, encryption ratio, speed, tunability, and power consumption. In conclusion, the Advanced Encryption Standard (AES) is evaluated as the most preferable alternative among the symmetric encryption algorithms due to its decreased energy consumption, buffer utilisation, and encryption and decryption times.

Simon Liu and Rick Kuhn, "constructed Data Loss Prevention"

It examines different sorts of losses, including leakage—where crucial data is no longer under the organization's control—and disappearance or loss—where the organisation is no longer in possession of an exact data copy. Data loss prevention systems were developed to address governmental, industrial, and intellectual property security needs by keeping diverse sensitive data from leaving an organization's private walls. Best practises include prioritising loss modes, providing protection without interruption, and using flexible and modular design were introduced. Loss Modes are Data at Rest, Data at the End Point, and Data in Motion.

Hector Garcia-Molinam designed "DATA LEAKAGE DETECTION"

According to this model, there is a chance to determine the propensity that a third-party agent is behind a leak based on the overlap between his data and the information that was leaked as well as the data from the other agents and the possibility that objects can be "guessed"

The algorithms used incorporate a number of data distribution strategies that can increase the chances of the distributer finding the leaker. Also, it has been found that carefully placing things can make a significant difference in identifying guilty agents, particularly in situations when the information that agents must obtain is highly comparable.

B. Sengupta and S. Ruj, 2016 "Utilizing safe cloud storage that is publicly verifiable for dynamic data"

Storage outsourcing is a service that cloud service providers offer to their customers. The integrity of the client's data is upheld through a secure cloud storage (SCS) protocol. In this work, we build a secure cloud storage system that is publicly verifiable and based on the secure network coding (SNC) protocol, allowing the client to update the outsourced data as necessary. Our protocol is the first SNC-based SCS protocol for dynamic data that is secure in the standard model and offers privacy-preserving audits in a publicly verifiable environment, as far as we are aware. Additionally, we go into great depth about the possibility of constructing a universal SCS protocol for dynamic data (DSCS protocol) from a random SNC protocol. Additionally, we alter a current DSCS scheme (DPDP).

The title suggests that the paper explores the idea of approaching text categorization tasks by framing them as graph classification problems. This implies that instead of analyzing text

DOI: 10.48175/IJARSCT-18403

III. PROPOSED METHODOLOGY

The proposed system presents a comprehensive solution to combat data leakage in cloud computing environments, integrating advanced AES encryption alongside watermarking and steganography techniques. In an era where cloud storage is ubiquitous, safeguarding sensitive data from insider threats is paramount. Our system ensures robust protection by encrypting data stored in the cloud using the AES (Advanced Encryption Standard) algorithm, safeguarding it against unauthorized access or interception.

In addition to encryption, our system enhances security through covert embedding of QR codes containing employee details within documents. Leveraging steganography, these QR codes are seamlessly integrated into the document's data, making them imperceptible to the naked eye. This hidden digital fingerprint allows for the identification of employees responsible for accessing or leaking sensitive information, facilitating swift detection and response to potential data breaches

The combination of AES encryption and covert QR code embedding provides a multi-layered defence mechanism against insider threats. AES encryption ensures that even if unauthorized individuals gain access to cloud-stored data, it remains unintelligible without the encryption key. Meanwhile, the embedded QR codes serve as a forensic tool, enabling organizations to trace and attribute data leaks to specific employees, thereby deterring malicious activities and enhancing accountability.

Moreover, our system promotes regulatory compliance by aligning with industry standards and best practices for data protection. AES encryption meets stringent security requirements mandated by regulations such as GDPR, HIPAA, and PCI DSS, helping organizations demonstrate compliance and avoid potential penalties.

Furthermore, our system minimizes overhead and complexity by seamlessly integrating encryption and watermarking/steganography techniques into existing cloud storage infrastructure. By encrypting data at rest and embedding QR codes within documents, we establish a comprehensive security framework that complements access controls and encryption mechanisms already in place.

Overall, our proposed system offers a holistic approach to data security in cloud computing, combining AES encryption with covert QR code embedding to mitigate insider threats and protect sensitive information. With the ability to encrypt stored data and trace unauthorized access or leakage back to specific employees, our system empowers organizations to maintain confidentiality, integrity, and trust in cloud-based storage solutions.

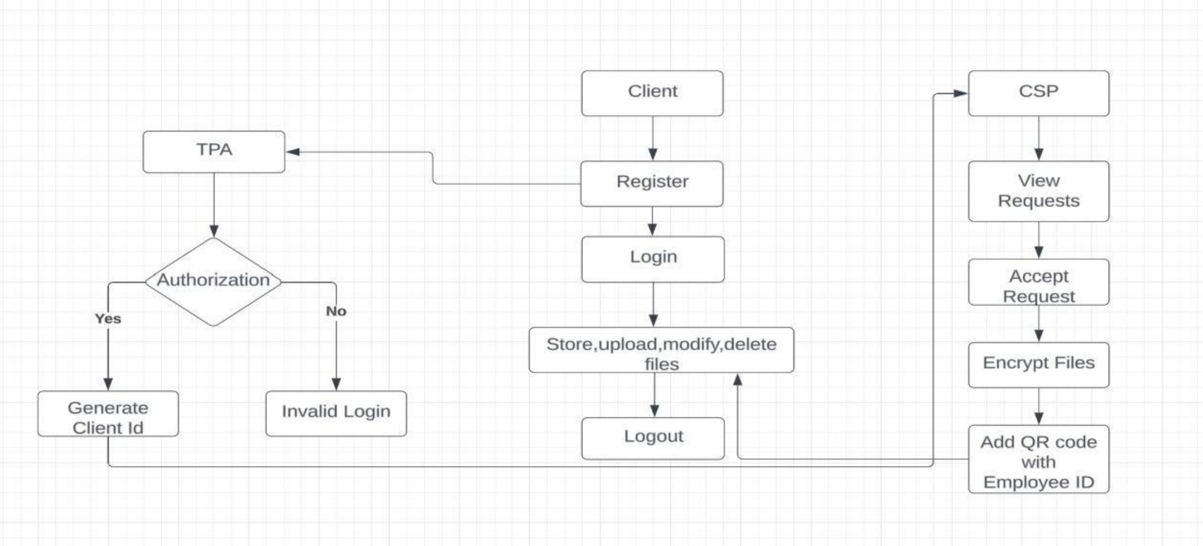


Fig1: Flow Diagram

IV. IMPLEMENTATION

The implementation of the proposed data leakage detection system in cloud computing environments involves the integration of several key components, including QR code embedding, AES encryption, and PDF manipulation. Below, we outline the implementation details of each component:

QR Code Embedding:

Utilized a Java library to read QR code images and convert them into Base64 encoded strings. Leveraged the Apache PDFBox library to access the metadata of PDF documents and embed the QR code data seamlessly. Implemented a function to concatenate the QR code data with existing metadata or create a new metadata field if necessary. Ensured that the embedding process maintains the integrity and confidentiality of the document's content, making the QR code imperceptible to the naked eye.

AES Encryption:

Employed the AES encryption algorithm to encrypt sensitive data stored in the cloud. Integrated AES encryption into the data storage and retrieval processes to ensure that data remains encrypted at rest. Generated and managed encryption keys securely to prevent unauthorized access to the encrypted data.

PDF Manipulation:

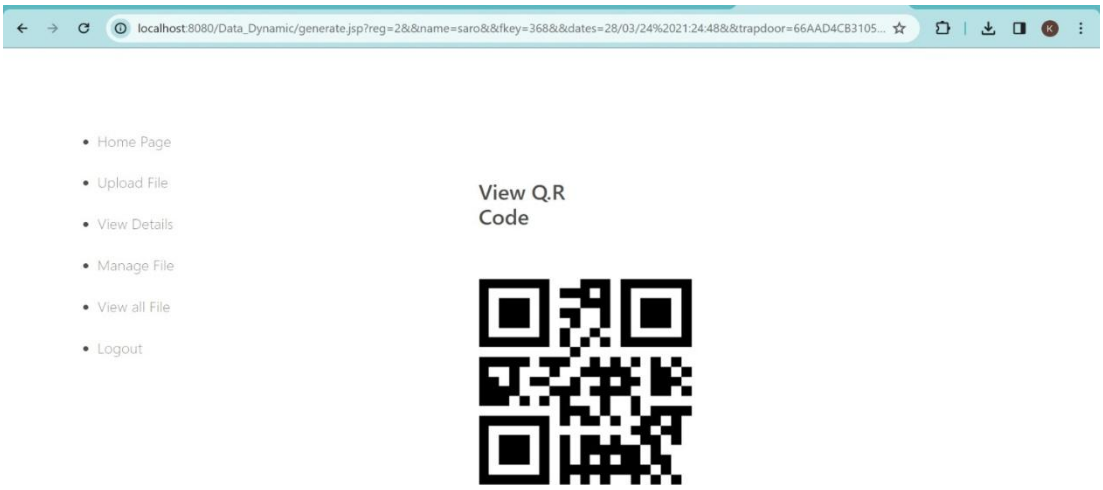
Used the Apache PDFBox library to load, modify, and save PDF documents. Developed functions to extract metadata from PDF documents and update metadata fields with the embedded QR code data. Implemented error handling to handle exceptions during PDF manipulation operations, such as file loading/saving errors or metadata extraction failures. Ensured compatibility with various PDF document formats and versions to support a wide range of use cases and scenarios.

Integration and Testing:

Integrated the QR code embedding, AES encryption, and PDF manipulation components into a cohesive system. Conducted extensive testing to validate the functionality, performance, and security of the implemented system. Tested the system with various PDF documents, QR code images, and data encryption scenarios to ensure robustness and reliability. Implemented logging and monitoring mechanisms to track system activities and detect anomalies or errors during operation

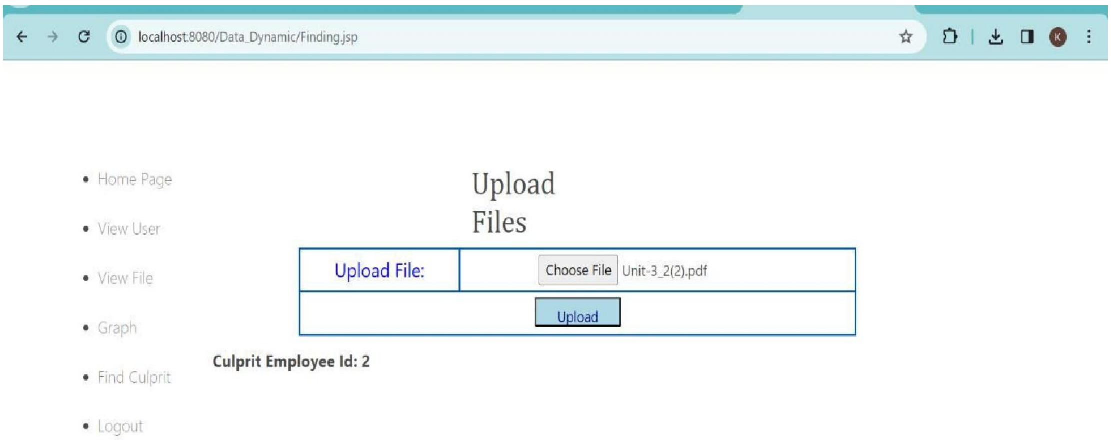
Overall, the implementation of the data leakage detection system involved the integration of QR code embedding, AES encryption, and PDF manipulation functionalities into a robust and reliable system. Through meticulous development, testing, the implemented system offers comprehensive data protection and leakage detection capabilities in cloud computing environments.

V. RESULTS



localhost:8080/Data_Dynamic/generate.jsp?reg=2&&name=saro&&fkey=368&&dates=28/03/24%2021:24:48&&trapdoor=66AAD4CB3105...

DOI: 10.48175/IJARSCT-18403

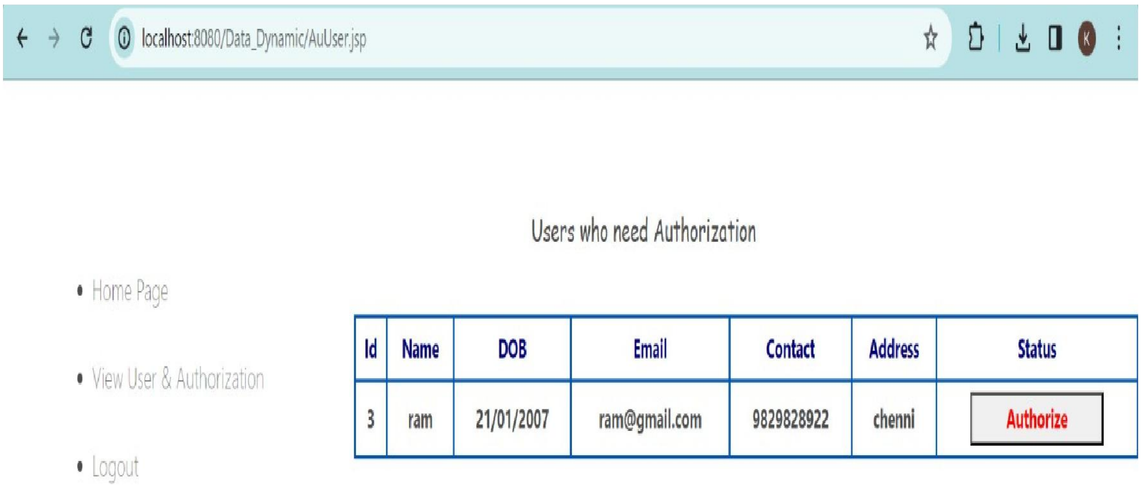


localhost:8080/Data_Dynamic/Finding.jsp

Upload Files

Upload File:	Choose File Unit-3_2(2).pdf
Upload	

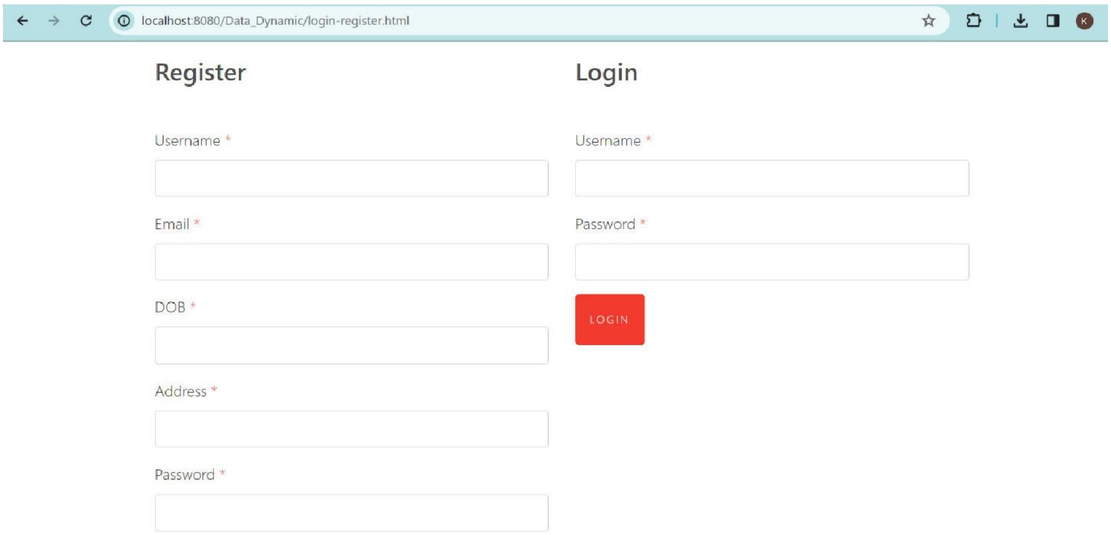
Culprit Employee Id: 2



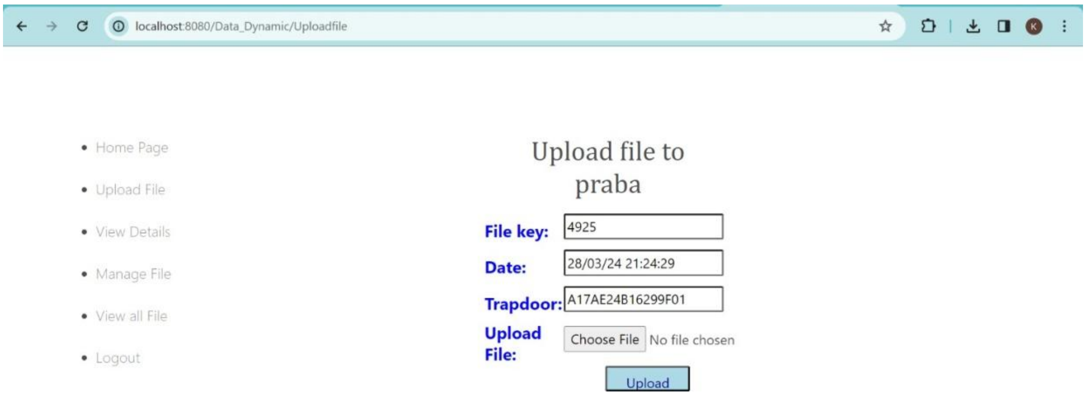
localhost:8080/Data_Dynamic/AuUser.jsp

Users who need Authorization

Id	Name	DOB	Email	Contact	Address	Status
3	ram	21/01/2007	ram@gmail.com	9829828922	chenni	Authorize



localhost:8080/Data_Dynamic/login-register.html



Upload file to praba

File key: 4925

Date: 28/03/24 21:24:29

Trapdoor: A17AE24B16299F01

Upload File: No file chosen

VI. CONCLUSION AND FUTURE WORK

In conclusion, the implemented data leakage detection system represents a significant advancement in addressing insider threats and safeguarding sensitive information in cloud computing environments. By seamlessly integrating QR code embedding, AES encryption, and PDF manipulation techniques, the system offers a comprehensive approach to data protection, ensuring confidentiality, integrity, and accountability. Through extensive testing and validation, the system demonstrates robustness and reliability in detecting and mitigating data leakage risks. Moving forward, further research and development efforts can focus on enhancing the scalability, performance, and usability of the system to meet the evolving needs of cloud-based data security. Overall, the implemented system serves as a vital tool for organizations seeking to uphold data privacy and compliance standards while leveraging the benefits of cloud computing.

Several potential future scopes can be integrated into the project to enhance its capabilities and address emerging challenges in cloud-based data security. Some of these future scopes include:

Machine Learning-based Anomaly Detection: Incorporating machine learning algorithms to analyse user behaviour patterns and detect anomalous activities indicative of data leakage. This could improve the system's ability to identify insider threats and unauthorized access more effectively.

Integration with Blockchain Technology: Exploring the integration of blockchain technology to enhance data integrity and auditability. Utilizing blockchain for recording access logs and document revisions can create an immutable audit trail, providing greater transparency and accountability.

Integration with Cloud Security Services: Leveraging existing cloud security services and APIs to enhance the system's capabilities. Integration with services such as AWS Guard Duty or Azure Security Centre can provide additional threat intelligence and proactive security measures.

REFERENCES

[1]. Rohit Pol, Vishwajeet Thakur, Ruturaj Bhise, and A Kat. Data leakage detection. International Journal of Engineering Research & Application, 2(3):404–410, 2012.

[2]. Rupesh Mishra and DK Chitre. Data leakage and detection of guilty agent. International Journal of Scientific & Engineering Research, 3(6), 2012.

[3]. Kenneth J Biba. Integrity considerations for secure computer systems. Technical report, DTIC Document, 1977.

[4]. David Elliott Bell. Bell–la padula model. Encyclopedia of Cryptography and Security, pages 74–79, 2011.

[5]. Mukesh Singhal and Niranjan G Shivaratri. Advanced concepts in operating systems. McGraw-Hill, Inc., 1994.

[6]. AL Jeeva, Dr V Palanisamy, and K Kanagaram. Comparative analysis of performance efficiency and security measures of some encryption algorithms. International Journal of Engineering Research and Applications (IJERA) ISSN, pages 2248–9622, 2012.

[7]. E Thambiraja, G Ramesh, and Dr R Umarani. A survey on various most common encryption techniques. International journal of advanced research in computer science and software engineering, 2(7):226–233, 2012.

[8]. Bruce Schneier, John Kelsey, Doug Whiting, David Wagner, Chris Hall, and Niels Ferguson. Performance comparison of the aes submissions, 1999.

[9]. Hamdan Alanazi, BB Zaidan, AA Zaidan, Hamid A Jalab, M Shabbir, Yahya Al-Nabhani, et al. New comparative study between des, 3des and aes within nine factors. arXiv preprint arXiv:1003.4085, 2010.

[10]. Aman Kumar, Sudesh Jakhar, and Sunil Makkar. Distinction between secret key and public key cryptography with existing glitches. Indian Journal of Education and Information Management, 1(9):392–395, 2012.

[11]. Hitendra GARG and Suneeta AGARWAL. A secure image based watermarking for 3d polygon mesh. SCIENCE AND TECHNOLOGY, 16(4):287–303, 2013.

[12]. Hitendra Garg and Suneeta Agrawal. Uniform repeated insertion of redundant watermark in 3d object. In Signal Processing and Integrated Networks (SPIN), 2014 International Conference on, pages 184–189. IEEE, 2014.

[13]. CISSP Susan Hansche, CISSP John Berti, and Chris Hare. Official (ISC) 2 guide to the CISSP exam. CRC Press, 2003.

[14]. D Elliott Bell and Leonard J La Padula. Secure computer system: Unified exposition and multics interpretation. Technical report, DTIC Document, 1976.

[15]. David Elliott Bell. Looking back at the bell-la padula model. In ACSAC, volume 5, pages 337–351, 2005.

[16]. Fred´ eric Deguillaume, Sviatoslav V Voloshynovskiy, and Thierry Pun. ´ Method for the estimation and recovering from general affine transforms in digital watermarking applications. In Electronic Imaging 2002, pages 313–322. International Society for Optics and Photonics, 2002.

[17]. Stallings William and William Stallings. Cryptography and Network Security, 4/E. Pearson Education India, 2006.

[18]. Achal Kumar and Vibhav Prakash Singh. Digital watermarking using color image processing using images for transmitting secret information.

[19]. JJK RUANAIDH and T PUN. Rotation, scale and translation invariant spread spectrum digital image watermarking. Signal processing, 66(3):303–317, 1998.

[20]. NIST FIPS Pub. 197. Announcing the Advanced Encryption Standard (AES), 2001

DOI: 10.48175/IJARSCT-18403