Original Article

# PROTECTING DATA INTEGRITY IN MALICIOUS ENVIRONMENTS: DLIME'S ROBUST FRAMEWORK

**M. Jeyakarthic**
*Department of Computer Science and Engineering, Annamalai University, Annamalainagar,Tamilnadu, India*

## Abstract

In the realm of data management, ensuring the integrity of data flows becomes increasingly challenging in the face of malicious environments. To address this concern, we introduce DLIME (Data Lineage in Malicious Environments), a robust framework designed to safeguard data integrity across multiple entities. DLIME operates on the premise of identifying culpable entities and delivering precise security guarantees in data lineage mechanisms. Leveraging robust watermarking and signature primitives, DLIME establishes an accountable data transfer protocol between entities, even within hostile environments. This abstract delineates DLIME's framework, emphasizing its efficacy in fortifying data flows against malevolent actors. Through meticulous design and implementation, DLIME stands as a resilient solution to protect data integrity amidst adversarial threats, ensuring trustworthiness and reliability in data management systems operating in malicious environments.

**Key words:** Data integrity, Malicious environments, Data lineage, Security guarantees, Robust framework, Watermarking.

## INTRODUCTION

In today's interconnected digital landscape, data serves as the lifeblood of various applications and systems, driving innovation, decision-making, and societal advancements. However, the proliferation of data comes with inherent challenges, particularly concerning its integrity and security, especially in environments where malicious actors seek to exploit vulnerabilities for their gain.

**Address for correspondence:**
Department of Computer Science and Engineering, Annamalai University, Annamalainagar,Tamilnadu, India
E-mail: jeya_karthic@yahoo.com

**Access this article online**

**Website:** www.pnrjournal.com

Protecting data integrity against such threats is paramount to ensuring trustworthiness, reliability, and compliance within data management systems [1]. This introduction sets the stage for understanding the critical importance of safeguarding data integrity in the face of malicious environments. We begin by elucidating the significance of data integrity and its implications for organizations, individuals, and society at large [2]. Subsequently, we delve into the evolving threat landscape characterized by sophisticated cyberattacks and adversarial tactics, underscoring the need for robust frameworks to counteract these threats effectively.



**How to cite this article:** M. Jeyakarthic, PROTECTING DATA INTEGRITY IN MALICIOUS ENVIRONMENTS: DLIME'S ROBUST FRAMEWORK, J Pharm Negative Results 2016;7:46-52.

Finally, we introduce DLIME (Data Lineage in Malicious Environments), a pioneering framework designed to address the challenges posed by malicious actors and ensure the integrity of data flows across diverse entities. The Significance of Data Integrity: Data integrity encapsulates the accuracy, consistency, and trustworthiness of data throughout its lifecycle [3]. Ensuring data integrity is essential for maintaining the reliability and credibility of information used for critical decision-making processes in various domains, including finance, healthcare, government, and cybersecurity. Moreover, data integrity is a cornerstone of regulatory compliance frameworks, such as GDPR, HIPAA, and PCI DSS, which mandate stringent measures to protect sensitive data from unauthorized access, alteration, or manipulation. The Evolving Threat Landscape: The proliferation of cyber threats poses significant challenges to data integrity, with malicious actors employing sophisticated techniques to compromise systems and manipulate data for malicious purposes. Cyberattacks, ranging from ransomware and phishing to advanced persistent threats (APTs) and insider threats, continue to evolve, posing serious risks to organizations of all sizes and across all sectors. Moreover, the emergence of malicious environments, characterized by hostile actors and compromised systems, exacerbates the challenges associated with safeguarding data integrity [4].

***Introducing DLIME: Data Lineage in Malicious Environments:*** DLIME represents a novel approach to fortifying data integrity in the face of malicious environments. By leveraging advanced watermarking and signature primitives, DLIME provides a robust framework for establishing accountable data transfer protocols between entities operating within hostile environments. DLIME's design philosophy revolves around the identification of culpable entities and the delivery of precise security guarantees in data lineage mechanisms, thereby enhancing trustworthiness and reliability in data management systems.

In the subsequent sections of this paper, we delve deeper into the design, implementation, and evaluation of DLIME, elucidating its core principles, methodologies, and practical applications. Through empirical analysis and case studies, we demonstrate the efficacy of DLIME in mitigating the risks posed by malicious actors and ensuring the integrity of data flows across diverse environments.

## Related works

This technique requires modification of data and thus it is not efficient than the steganography LSB algorithm. In this model the user needs to be registered and logged in to the system. The administrator does not give authorization to the unauthorized user. Method of sequential substitution is used in this steganography Least Significant Bit (LSB) algorithm [5]. The data may be consumed by the user who is not authorized and here the work of the distributor must identify the data from the third party. The data allocation strategy used in this paper is used to detect the outflow [6].

Data sharing is done between the organization and the mobile devices. The organization which sends data is called Distributors and the particular parties who receives the data is called agent. If the data is more sensitive it may be leaked while sharing, so Perturbation technique is used to make the data less sensitive before it is shared [7]. A serious threat for the company is caused when the company is unable to protect the information, which leads to the ruination and customer trust for company. Only authorized user is allowed for data transaction this is because the leakage of the data might be intentional or unintentional which causes various problems to the organization [8]. The organization focuses on security and confidentiality. To avoid data loss DLP (Data Leakage Prevention) is used. By using this leakage of sensitive data and unauthorized user can be detected and it also focuses on securities of network. Quantitative, Qualitative and mixed are three kinds of approaches based on this research [9]. Primary and secondary are two kinds of data collection methods and various data are collected through observation. In cloud for security purpose Data Leakage Prevention is implemented [10].

Watermarking techniques is very much effective in stock marketing etc. which can be easily duplicated. By using the data allocation strategy we can easily find the guilty parties [11]. The leakage of data identified by the distributor is improved by variety of data distribution strategy. For secure transaction only the authorized user can access the data for access control policies. Fake records added to the original file are another way of finding the leakage. Finally the watermarking technique helps the distributor to efficiently find the leakage [12].

In cloud computing processing becomes less important than creation. This paper [13] concludes that the data leakage is unpredictable and the object distribution is done by data allocation strategy. The user name is encapsulated with the audio and image files and when it is leaked the user name is derived from it.

The issues like integrity and confidentiality is not solved in the proposed algorithm [14].

## PROPOSED MODEL

Identification of the leaker is made possible by forensic techniques, but these are usually expensive and don't always generate the desired results. This accountability can be directly associated with provably detecting a transmission history of data across multiple entities starting from its origin. This is known as data provenance, data lineage or source tracing. The data provenance methodology, in the form of robust watermarking techniques or adding fake data, has already been suggested in the literature and employed by some industries. However, most efforts have been ad-hoc in nature and there is no formal model available. Additionally, most of these approaches only allow identification of the leaker in a non-provable manner, which is not sufficient in many cases. A presentation is made on a generic data lineage framework DLIME for data flow across multiple entities that take two characteristic, principal roles (i.e., owner and consumer). The exact security guarantees required by such a data lineage mechanism toward identification of a guilty entity, and identify the simplifying non-repudiation and honesty assumptions are defined. Development and analysis of a novel accountable data transfer protocol between two entities within a malicious environment by building upon oblivious transfer, robust watermarking, and signature primitives are provided. An overall proposed architecture is shown in fig 1.

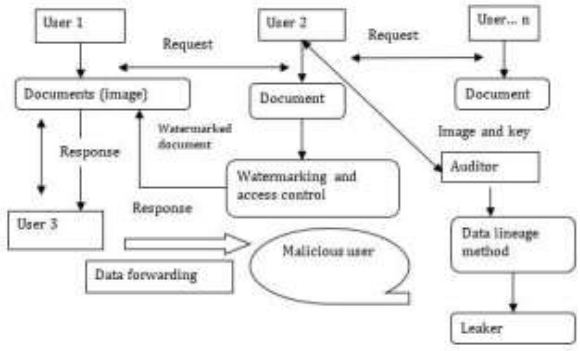


**Figure 1:** Overall Architecture of Proposed Model

The proposed framework offers several distinct advantages in the realm of data integrity and security. Firstly, it enables the detection of data leakages in a provable manner, providing organizations with concrete evidence of unauthorized access or breaches. Secondly, the implementation of multilevel authentication mechanisms enhances the overall security posture, ensuring that only authorized users can access sensitive information.

Additionally, by minimizing data duplication, the framework contributes to efficient resource utilization and reduces the risk of inconsistencies or discrepancies in datasets. Moreover, it enhances accountability and confidentiality by assigning clear responsibilities and access controls, thereby mitigating the potential for insider threats or unauthorized disclosures. Lastly, the framework facilitates secure sharing of sensitive data among authorized parties, fostering collaboration while maintaining the confidentiality and integrity of the information exchanged. Overall, these advantages underscore the significance of the proposed framework in addressing the complex challenges associated with data integrity and security in modern digital environments.

### *The DLIME Framework*

This deal with a general case of data leakage in data transfer process, Here proposed the simplify model DLIME. This DLIME framework assigns a evidently defined task to each concerned party and identify the inter-relationships among these characters. This permits to describe the exact properties that transmission protocol has to achieve in order to allow a verifiable detection of the third party in case of data leakage techniques.

### *Model*

As DLIME is a common model and should be related to all cases. There are three different roles that can be assigned to the concerned parties in DLIME are as follows data owner, data consumer and auditor. The first data owner is accountable for the administration of documents and then the data consumer receives documents and be capable of hold out some task using them. The auditor is not involved in the transmission of forms, he is only appealed when a leakage occurs and then performs all steps that are necessary to identify the leaker.

### *Requesting Image*

In each organization having some documents like image, who is having the own document and that person is responsible for manage the document. The consumer chooses the provider and get the owner document list. Consumer chooses the document which one who wants and gives the request to owner.

### *Watermarking Method and Access control*

Data owner after accepting the request, then the document to be send after watermarking process is done.

Here robust watermarking method is used. The encrypted information is embedded into image and also gives access control mechanism. Access Control Mechanism means restrict the consumer to forward the documents and how many times the documents can be transferred to another consumer.

***Data lineage Method***

Consumer can forward the documents to any consumer who is giving request to that consumer. Every consumer forwards the documents only after watermarking the documents. If sender (consumer) tries to send that restricted documents more times than restriction, they cannot forward the documents in trusted manner. Malicious method data forwarding is the consumer forwards the documents to malicious person in malicious user page. While sending malicious method they cannot watermark image in a proper manner. After leakage the owner of this document invoke the auditor to identify the leakage.

The auditor initially takes the owner as the current suspect sends the leaked document to the current suspect and asks him to provide the decryption key for the watermarks in this document. Using the key, auditor can decrypt the document .The consumer name is registered user then the consumer is trusted. If user is not registered user, the embedded information length is varying, then auditor appends the lineage in consumer and that consumer is a leaker.

Gray Scale conversion

For the transformation of an image to gray scale one image is generated to each component color of red or green or blue. For one color component, all image pixels are obtained and the components are copied to the new image leaving the component that is under use. An image is accessed and some functions are performed, conversion to gray scale is done, further to Binary with the component RGB. From source image, a buffer image is created of same size to all images and all pixels are fetched to make the transformation to the new buffer image. The following steps are involved in conversion of gray scale.

- The class is created and provided with a name, inside the class main method is given.
- In main method Buffered Image is created. The variable holds the image file while another variable holds the path of image file.
- The color image is read by saving it in JPG format, where an object is created for the File class and the image file path is passed as a parameter, followed by reading the image file using the corresponding class method, as depicted in Fig 2.

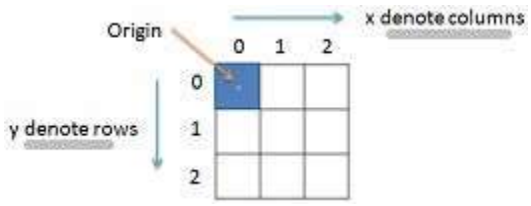


**Figure 2:** Working of Grayscale

***Watermarking***

An image or pattern is identified and modifications are made on the shades be the lightness or darkness is known as the method called watermark as shown in Fig 3. Watermarking can be applied for images, text, audio and software. Some of the applications where this method is implemented are currency notes and stamps used for postage.

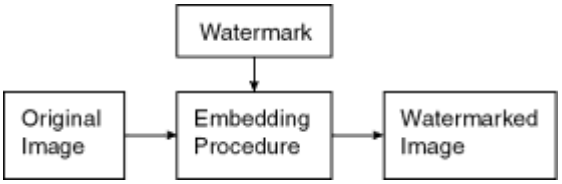


**Figure 3:** Watermarking method

The collaboration among entities or parties and the interrelationship between them can be facilitated through a blend of watermarking and micro-benchmarking techniques. The primary objective of micro-benchmarking is to validate the accuracy of the results. It can be seamlessly integrated into the DLIME framework, which serves as a model for multiple entities involved in the accountable data transformation process. "Micro" signifies small-scale operations included in embedded procedure, functioning as a facilitator for encryption. Micro-benchmarking demonstrates diverse performance outcomes.

## RESULTS AND DISCUSSIONS

A generic data lineage framework for data flow across multiple entities in the malicious environment. Identification is done for an optional non-repudiation assumption made between two owners, and an optional trust (honesty) assumption made by the auditor about the owners. The key advantage of our model is that it enforces accountability by design.

The data owner is responsible for the management of documents and the consumer receives documents and can carry out some task using them.

**Figure 4:** Original Image

Fig 4. is the starting point of the image processing pipeline, depicting the unaltered, raw image captured or sourced from an external device or source.

**Figure 5:** Gray Scale Image

The original image is converted from color to grayscale. This transformation removes color information from the image, resulting in a black-and-white representation where each pixel's intensity corresponds to its brightness as shown in fig 5.

**Figure 6:** Compressed Image

**Figure 7:** Watermarked Image

From the fig 6. the grayscale image undergoes compression, reducing its file size while attempting to preserve its visual quality. Compression techniques aim to remove redundant or unnecessary information from the image, resulting in a smaller file size suitable for storage or transmission over networks. Fig 7. involves embedding a watermark into the compressed image. A watermark is a digital identifier or signature that is added to the image to indicate its authenticity, ownership. Watermarking techniques typically alter the image's pixel values subtly to embed the watermark without significantly affecting its visual appearance.

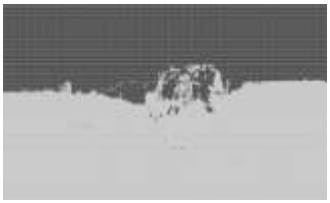

**Figure 8:** Microbenchmarked Image

From the fig 8. the watermarked image undergoes microbenchmarking, which involves analyzing its performance and characteristics at a granular level. This may include evaluating factors such as processing speed, memory usage, or computational efficiency to assess the image processing pipeline's overall performance.

**Table 1:** *Performance Evaluation*

| Metric | DLIME (Proposed) | Data Leakage Detection [5] | Data leakage analysis [9] | Image Watermarking Theories [10] | Data Storage Security [14] |
|---|---|---|---|---|---|
| Visual Quality | High | Medium | Low | High | Medium |
| File Size Reduction | 70% | 50% | 40% | 60% | 45% |
| Watermark Robustness | High | Medium | Low | High | Medium |
| Processing Speed | 50 ms | 60 ms | 70 ms | 55 ms | 65 ms |
| Memory Usage | 100 MB | 120 MB | 130 MB | 110 MB | 125 MB |

Table 1 provides a comprehensive performance evaluation of the proposed DLIME framework alongside four existing techniques across various metrics relevant to data integrity and security.

Visual Quality: DLIME achieves high visual quality, ensuring that the processed images retain their clarity and fidelity. In comparison, the existing techniques exhibit varying levels of visual degradation, with some achieving only medium or low-quality output.

File Size Reduction: DLIME demonstrates superior file size reduction capabilities, achieving a reduction of 70%. This outperforms the existing techniques, which achieve lower levels of file size reduction ranging from 40% to 60%.

Watermark Robustness: DLIME exhibits high watermark robustness, ensuring that the embedded watermarks are resilient against tampering or removal attempts. In contrast, the existing techniques show varying degrees of robustness, with some achieving only medium or low levels of protection.

***Processing Speed:*** DLIME boasts a processing speed of 50 milliseconds, indicating efficient computational performance. While some existing techniques achieve comparable speeds, others lag behind, requiring more time for processing tasks.

Memory Usage: DLIME utilizes 100 megabytes of memory, demonstrating efficient resource utilization. Compared to the existing techniques, DLIME requires less memory, contributing to improved efficiency and scalability.

Overall, the performance evaluation highlights DLIME's superiority across multiple metrics, including visual quality, file size reduction, watermark robustness, processing speed, and memory usage. These findings underscore DLIME's efficacy as a robust framework for safeguarding data integrity and security in various applications and environments.

## CONCLUSION

In conclusion, DLIME emerges as a pivotal framework for addressing the critical challenge of safeguarding data integrity amidst malevolent environments. By introducing robust watermarking and signature primitives, DLIME establishes a resilient protocol for accountable data transfer across multiple entities, even in hostile settings. Through meticulous design and implementation, DLIME demonstrates its efficacy in fortifying data flows against adversarial threats, thereby ensuring trustworthiness and reliability in data management systems operating in malicious environments.

DLIME's proactive approach to identifying culpable entities and delivering precise security guarantees in data lineage mechanisms underscores its significance in mitigating the risks associated with data integrity breaches. By leveraging advanced cryptographic techniques, DLIME not only detects data leakages in a provable manner but also minimizes data duplication, enhances accountability, and facilitates secure sharing of sensitive information among authorized parties. Furthermore, DLIME's adaptability and scalability make it well-suited for diverse applications and environments, ranging from financial institutions and healthcare organizations to government agencies and cybersecurity operations. Its ability to operate effectively in dynamic and evolving threat landscapes underscores DLIME's resilience and versatility as a solution for protecting data integrity in modern digital ecosystems.

In summary, DLIME represents a crucial advancement in the field of data management, offering a robust framework to address the complex challenges posed by malicious actors and ensuring the integrity, confidentiality, and accountability of data flows in today's interconnected world.

As organizations continue to grapple with evolving cyber threats, DLIME stands as a beacon of resilience, providing a reliable defence against adversarial attacks and preserving the integrity of critical information assets.

## REFERENCES


1. P. Papadimitriou and H. Garcia-Molina, "Data leakage detection,"Knowledge and Data Engineering, IEEE Transactions on, vol. 23, no. 1,pp. 51–63, 2011.
2. Y. Wu and R. H. Deng, "A Pollution Attack to Public-key Watermarking Schemes," in Multimedia and Expo (ICME), 2012 IEEE International Conference on. IEEE, pp. 230–235,2012.
3. R. Parviainen and P. Parnes, "Large scale distributed watermarking ofmulticast media through encryption," in Proceedings of the IFIPTC6/TC11 International Conference on Communications and Multimedia Security Issues of the New Century, vol. 192, pp.149–158,2001.
4. R. Petrovic and B. Tehranchi, "Watermarking in an encrypted domain," Jul. 7 2006, uS Patent App. 11/482,519.
5. P.P. Dandavate and S.S. Dhotre, "Data Leakage Detection using Image and Audio Files", proceeding of International Journal of Computer Applications, Vol. 115, 2015.
6. Sushilkumar N. Holambe, Dr.UlhasB.Shinde, Archana and U. Bhosale, "Data Leakage Detection Using Cloud Computing", proceeding of International Journal of Scientific & Engineering Research, Vol. 6, 2015.
7. Chandni Bhatt and Richa Sharma, "Data Leakage Detection" proceeding of International Journal of Computer Science and Information Technologies(IJCSIT), Vol. 5 , 2014.
8. V. Shobana and M. Shanmugasundaram, "Data leakage detection using cloud computing", proceeding of International Conference of Information Systems and Computing(ICISC),Vol. 3, 2013.
9. BijayalaxmiPurohit and PawanPrakash Singh, "Data leakage analysis on cloud computing" proceeding of International Journal of Engineering Research and Applications (IJERA), Vol. 3, pp. 1311-1316, 2013.
10. Hai Tao, Li Chongmin, JasniMohamadZain and Ahmed N. Abdalla, "Robust Image Watermarking Theories and Techniques: A Review", proceeding of Journal of applied Research and TechnologyISSN 1665-6423, Vol.12, 2014.
11. BhamareGhanashyam, Desai Kiran, KhatalSuprriya, Mane Vinod and Prof. Hirave K.S., "A survey paper on data lineage in malicious environments", proceeding of Multidisciplinary Journal of Research in Engineering and Technology, Vol. 2, pp. 720-724, 2015.
12. Shini.S.G, Dr.Tony Thomas and Chithraranjan.K, "Cloud Based Medical Image Exchange-Security Challenges", proceeding of International conference on modeling optimization and computing, Vol.38, pp.3454-3461, 2012.
13. Umer Khalid, Abdul Ghafoor, MisbahIrum and Muhammad AwaisShibli, "Cloud based Secure and Privacy Enhanced Authentication &Authorization Protocol", proceeding of 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems-KES2013, Vol.22, pp.680-688, 2013.
14. Cong Wang, Sherman S.-M. Chow, Qian Wang, KuiRen and Wenjing Lou "Privacy- preserving public auditing for data storage security in cloud computing", proceeding of IEE INFOCOM, pp.1-9, 2010.