IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

Data Leakage Tracer System for Document Sharing

Rakshitha B[1], Sona A C[2], Prof. Thejaswini M N[3]
Students, Department of MCA[1,2]
Assistant Professor, Department of MCA[3]
MITM, Belawadi, SR Patna, Mandya

Abstract: *The increasing use of cloud computing and digital platforms has made secure document sharing essential for organizations, educational institutions, healthcare systems, and government agencies. Traditional security mechanisms such as authentication, encryption, and access control protect documents during storage and transmission but provide limited protection after a document is downloaded. This survey reviews recent techniques for preventing data leakage, including Digital Rights Management (DRM), Data Loss Prevention (DLP), invisible watermarking, digital fingerprinting, blockchain technology, and secure hashing. Their strengths, limitations, and challenges are analyzed with respect to document security and traceability. Based on the identified research gaps, a secure document-sharing framework integrating invisible watermarking, digital fingerprinting, and secure hash generation is presented to improve document accountability and leak tracing. The survey also discusses future research directions involving artificial intelligence, blockchain, and cloud-based security..*

Keywords: Secure Document Sharing, Data Leakage Prevention, Invisible Watermarking, Digital Fingerprinting, Secure Hashing, Document Traceability, Information Security.

1. INTRODUCTION

The widespread adoption of cloud computing and digital platforms has transformed the way organizations create, store, and share confidential documents. Business enterprises, educational institutions, healthcare organizations, and government agencies regularly exchange sensitive information through online systems. While digital document sharing improves accessibility and collaboration, it also increases the risk of unauthorized access, insider threats, and data leakage.

Conventional security mechanisms such as authentication, encryption, and role-based access control effectively protect documents during storage and transmission. However, once a document is downloaded by an authorized user, these mechanisms provide limited control over its further distribution. As a result, copied or shared documents become difficult to trace, making forensic investigation and accountability challenging.

To overcome these limitations, researchers have explored several techniques, including Digital Rights Management (DRM), Data Loss Prevention (DLP), invisible watermarking, digital fingerprinting, blockchain technology, and secure hashing. Among these, invisible watermarking and digital fingerprinting have gained significant attention because they enable unique identification of document copies without affecting document quality. These techniques improve document traceability and assist in identifying the source of leaked documents.

This survey reviews recent research on secure document sharing and data leakage prevention. It analyzes existing techniques, identifies their strengths and limitations, and highlights current research gaps. In addition, a secure document-sharing framework based on invisible watermarking, digital fingerprinting, and secure hash generation is presented to improve document traceability, accountability, and forensic investigation. The remaining sections discuss the literature review, comparative analysis, proposed framework, future research directions, and conclusion.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, July 2026 Impact Factor: 8.2

CONTRIBUTIONS OF THIS SURVEY

The main contributions of this survey are as follows:

- Reviews recent research on secure document sharing and data leakage prevention techniques.
- Examines the strengths and limitations of Digital Rights Management (DRM), Data Loss Prevention (DLP), invisible watermarking, digital fingerprinting, blockchain, and secure hashing.
- Compares existing approaches based on security, traceability, scalability, and forensic capabilities.
- Identifies key research gaps in document traceability and leak detection after document distribution.
- Presents a secure document-sharing framework integrating invisible watermarking, digital fingerprinting, and secure hash generation.
- Highlights future research directions involving artificial intelligence, blockchain, cloud security, and advanced watermarking techniques.

II. LITERATURE SURVEY

The increasing use of cloud computing and digital collaboration has made secure document sharing an important research area. Researchers have proposed several techniques, including Digital Rights Management (DRM), Data Loss Prevention (DLP), invisible watermarking, digital fingerprinting, blockchain, artificial intelligence, and secure hashing to reduce document leakage. This section summarizes these approaches and discusses their strengths and limitations.

A. Digital Rights Management (DRM)

Digital Rights Management (DRM) controls document access by restricting operations such as copying, printing, editing, and forwarding. It protects documents while they remain within the authorized environment. However, DRM cannot effectively trace leaked documents once security restrictions are bypassed. Its high deployment cost also limits adoption in small organizations.

B. Data Loss Prevention (DLP)

Data Loss Prevention (DLP) monitors networks and endpoints to prevent confidential information from leaving an organization. It helps detect unauthorized file transfers and policy violations. However, DLP becomes less effective after a document has been downloaded by an authorized user, making post-distribution traceability difficult.

C. Invisible Watermarking

Invisible watermarking embeds hidden information into documents without affecting their appearance or readability. Each distributed document contains a unique identifier that can be extracted to trace the source of a leaked copy. Although it improves accountability, watermark robustness may decrease when documents are modified or converted into different formats.

D. Digital Fingerprinting

Digital fingerprinting assigns a unique identifier to every distributed document using user and document information. It enhances traceability and supports forensic investigations by identifying the source of leaked documents. However, managing fingerprints becomes challenging in large-scale systems with numerous users and documents.

E. Blockchain-Based Document Security

Blockchain technology provides a secure and tamper-resistant record of document transactions. It improves transparency, integrity, and auditability by maintaining immutable logs of document access and sharing. Despite these advantages, blockchain introduces higher computational costs and integration challenges.

F. Artificial Intelligence-Based Security

Artificial Intelligence (AI) and Machine Learning (ML) analyze user behavior to identify suspicious document access patterns and insider threats. These techniques enable proactive security monitoring but require large datasets and may produce false alerts. Most AI-based solutions focus on threat detection rather than document traceability.

G. Secure Hashing Techniques

Secure hashing algorithms such as SHA-256 generate unique fingerprints for documents to verify integrity and authenticity. Any modification to a document changes its hash value, making tampering detectable. However, secure

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

hashing alone cannot identify the source of leaked documents and is most effective when combined with watermarking or digital fingerprinting.

Comparative Analysis of Existing Techniques

Technique	Advantages	Limitations
Digital Rights Management (DRM)	Restricts copying, printing and editing	Cannot identify leaked document source
Data Loss Prevention (DLP)	Monitors data movement inside organization	Ineffective after document download
Invisible Watermarking	Unique hidden identification for each document	Vulnerable to document modification
Digital Fingerprinting	Strong document traceability	Complex management for large organizations
Blockchain	Immutable audit trail and transparency	High implementation complexity
Artificial Intelligence	Detects abnormal user behavior	Requires large datasets and training
Secure Hashing	Ensures document integrity	Cannot trace leaks independently

Summary of Literature Survey

The literature review shows that existing document security techniques mainly focus on preventing unauthorized access and protecting data during storage and transmission. While Digital Rights Management (DRM) and Data Loss Prevention (DLP) improve document security, they provide limited support for tracing leaked documents after distribution. Blockchain enhances transparency and maintains secure audit records but increases implementation complexity. Similarly, Artificial Intelligence (AI) helps detect suspicious user activities, but it does not directly identify the source of leaked documents.

Among the reviewed techniques, the combination of invisible watermarking, digital fingerprinting, and secure hashing offers a more effective solution for document traceability. These methods enable organizations to uniquely identify distributed document copies, support forensic investigations, and improve accountability without affecting document quality.

III. COMPARATIVE ANALYSIS

Several approaches have been proposed to improve document security and prevent data leakage. Although each technique offers unique advantages, they differ in terms of security, traceability, forensic support, and implementation complexity. Table I compares the major document protection techniques and highlights their strengths and limitations.

Technique	Advantages	Limitations
DRM	Restricts copying and printing	Cannot trace leaked copies
DLP	Monitors confidential data	Ineffective after download
Invisible Watermarking	Unique hidden identifier	Sensitive to document modification
Digital Fingerprinting	Strong traceability	Difficult to manage at large scale
Blockchain	Secure audit trail	High computational overhead
AI	Detects abnormal behavior	Requires large datasets
Secure Hashing	Ensures document integrity	Needs watermarking for tracing

Table I. Comparative Analysis of Existing Techniques

The comparison shows that combining invisible watermarking, digital fingerprinting, and secure hashing provides a more effective solution for document traceability and forensic investigation than traditional security techniques.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

IV.RESEARCH GAP

The literature review highlights several limitations in existing document-sharing techniques. Most current solutions focus on protecting documents during storage and transmission but provide limited support after documents have been downloaded. Traditional methods such as authentication, encryption, Digital Rights Management (DRM), and Data Loss Prevention (DLP) help prevent unauthorized access but cannot effectively identify the source of leaked documents.

Although blockchain improves data integrity and auditability, it increases implementation complexity and computational overhead. Similarly, Artificial Intelligence (AI) can detect suspicious user behavior but does not provide reliable document traceability or identify leaked document copies.

These limitations indicate the need for a lightweight and efficient document-sharing framework that combines invisible watermarking, digital fingerprinting, and secure hashing. Such a framework can improve document traceability, strengthen accountability, and support forensic investigations while maintaining document quality and usability.

V. PROPOSED FRAMEWORK

The proposed framework integrates invisible watermarking, digital fingerprinting, secure hashing, and forensic analysis to improve document security and traceability. Its primary objective is to uniquely identify every distributed document and enable organizations to trace the source of leaked copies.

The framework includes the following components:

- User Authentication Module
- Document Repository
- Watermark Generation Engine
- Secure Hash Generator
- Download History Database
- Watermark Database
- Leak Detection Module
- Forensic Investigation Engine
- Administrative Dashboard

When an authenticated user downloads a document, the system generates a unique invisible watermark using user and document information. A secure hash is then created and stored along with the watermark details in the database. If a document is leaked, the system extracts the embedded watermark, compares it with the stored records, and identifies the responsible user. Finally, a forensic investigation report is generated to support document tracing and accountability.

The proposed framework enhances document security while preserving document quality, usability, and efficient leak detection.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

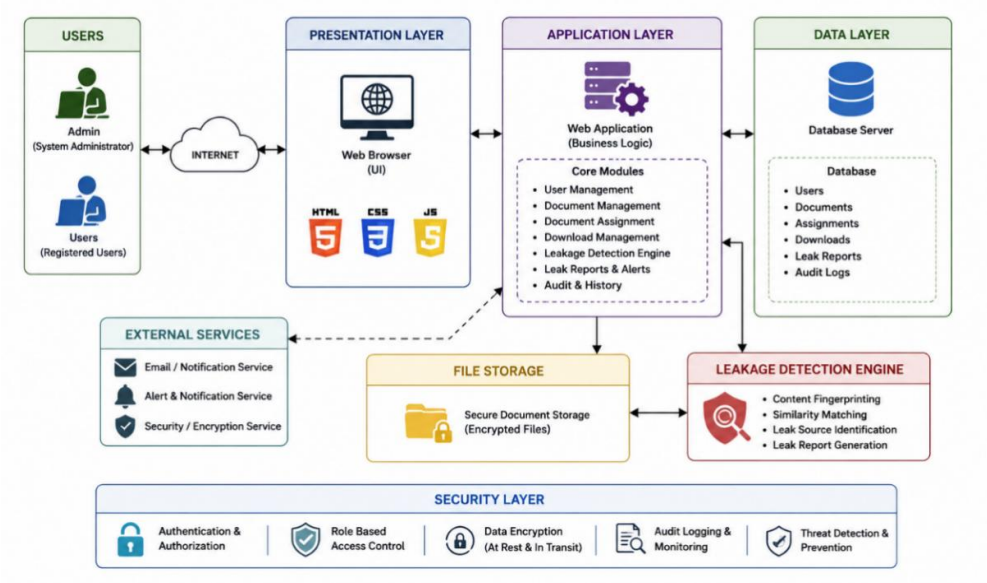

Figure 1. Proposed Framework for Secure Document Sharing and Data Leakage Tracing

VI. WORKING OF THE PROPOSED FRAMEWORK

The proposed framework follows a systematic process to secure document sharing and trace leaked documents. The workflow consists of the following steps:

Step 1: User Authentication

The user logs into the system using valid credentials. The authentication module verifies the user's identity before granting access.

Step 2: Document Selection

After successful authentication, the user selects the required document from the secure repository.

Step 3: Watermark Generation

A unique invisible watermark is generated using user-specific information, including the User ID, Document ID, download timestamp, and device details.

Step 4: Secure Hash Generation

A secure cryptographic hash is created from the watermark and document metadata to ensure data integrity and uniqueness.

Step 5: Watermark Embedding

The generated watermark is embedded into the document without affecting its appearance, readability, or content.

Step 6: Document Download

The watermarked document is made available for download to the authenticated user.

Step 7: Activity Logging

The system records the download details, watermark information, secure hash value, and user information in the database for future verification.

Step 8: Leak Detection

If a leaked document is discovered, the administrator uploads the suspected document to the leak detection module for analysis.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, July 2026 Impact Factor: 8.2

Step 9: Watermark Extraction and Verification

The system extracts the embedded watermark, verifies the secure hash, and compares the information with the stored records.

Step 10: User Identification and Report Generation

Based on the verification results, the system identifies the user associated with the leaked document and generates a forensic investigation report.

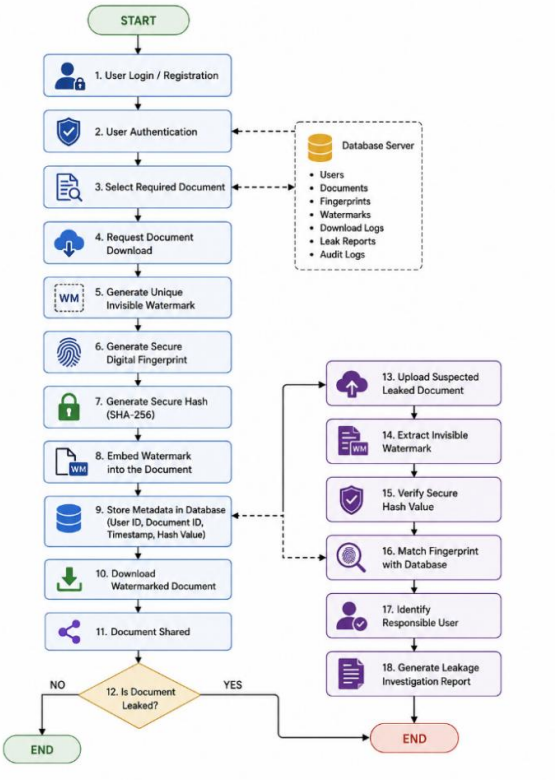

Figure 2. Working of the Proposed Secure Document Sharing and Data Leakage Tracing Framework

VII. ADVANTAGES OF THE PROPOSED FRAMEWORK

The proposed framework offers several advantages over existing document protection techniques by improving document security, traceability, and accountability. Its key benefits are as follows:

- Provides a unique identity for every distributed document.
- Improves accountability by enabling the identification of the document recipient.
- Preserves document quality through invisible watermark embedding.
- Supports accurate leak tracing and forensic investigation.
- Ensures document integrity using secure hash verification.
- Can be integrated with existing document management systems.
- Offers a scalable solution suitable for both small and large organizations.
- Reduces implementation cost by supporting open-source technologies.
- Supports future enhancements through AI and blockchain integration.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, July 2026 Impact Factor: 8.2

VIII. FUTURE RESEARCH DIRECTIONS

Future research can further improve secure document-sharing systems by incorporating advanced technologies that enhance security, traceability, and system efficiency. Some promising research directions include:

- Artificial Intelligence (AI): AI and machine learning can analyze user behavior to detect suspicious activities and potential data leakage at an early stage.
- Blockchain Integration: Blockchain can provide secure and tamper-resistant audit logs, improving transparency and trust in document-sharing systems.
- OCR-Based Watermark Detection: Optical Character Recognition (OCR) can enable watermark extraction from scanned or printed document copies, improving leak detection.
- Cloud-Based Security: Cloud-native security solutions can provide scalable document protection and secure access across distributed environments.
- Multimedia Watermarking: Future research can extend watermarking techniques to images, videos, audio files, and other multimedia content.
- Mobile Document Security: Secure watermarking and traceability mechanisms can be integrated into mobile applications to protect documents accessed through smartphones and tablets.

IX. CONCLUSION

Secure document sharing has become essential as organizations increasingly exchange confidential information through digital platforms. Although traditional security mechanisms such as authentication, encryption, and access control protect documents during storage and transmission, they provide limited support for tracing leaked documents after distribution.

This survey reviewed major document protection techniques, including Digital Rights Management (DRM), Data Loss Prevention (DLP), invisible watermarking, digital fingerprinting, blockchain, secure hashing, and artificial intelligence. Their strengths, limitations, and applications were analyzed to identify current challenges in document security and traceability.

Based on the identified research gaps, a secure document-sharing framework integrating invisible watermarking, digital fingerprinting, and secure hashing was presented. The proposed framework enhances document traceability, strengthens accountability, and supports effective forensic investigation while preserving document quality.

Future research should focus on integrating emerging technologies such as artificial intelligence, blockchain, cloud security, and advanced watermarking techniques to develop more secure, scalable, and intelligent document-sharing systems.

REFERENCES

[1] Z. Wang, O. Byrnes, H. Wang, R. Sun, C. Ma, H. Chen, Q. Wu, and M. Xue, "Data Hiding with Deep Learning: A Survey Unifying Digital Watermarking and Steganography," IEEE Access, 2021.
[2] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Transactions on Image Processing, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.
[3] J. Nin and S. Ricciardi, "Digital Watermarking Techniques and Security Issues," IEEE AINA Workshops, 2013.
[4] M. Fatima, A. Akhunzada, M. A. Khan, and S. M. Abbas, "A Novel Fingerprinting Technique for Data Storing and Sharing in Cloud Environment," IEEE Access, 2021.
[5] T. Ji, E. Ayday, E. Yilmaz, and P. Li, "Privacy-Preserving Database Fingerprinting," 2021.
[6] National Institute of Standards and Technology (NIST), Secure Hash Standard (SHS), FIPS PUB 180-4, Aug. 2015.
[7] National Institute of Standards and Technology (NIST), Digital Signature Standard (DSS), FIPS PUB 186-4.
[8] ISO/IEC 27001:2022, Information Security, Cybersecurity and Privacy Protection — Information Security Management Systems — Requirements.
[9] V. Attasena, J. Darmont, and N. Harbi, "Secret Sharing for Cloud Data Security," 2017.
[10] H. G. Schaathun, "On Watermarking/Fingerprinting for Copyright Protection," 2006.