Published Date: - 31-01-2026
Page No: - 40-43

METHODS AND ALGORITHMS FOR PROTECTING FILE METADATA IN INFORMATION AND COMMUNICATION SYSTEMS

Sodiqova Dilnoza Jumanazarovna
Department of Cybersecurity and digital forensics Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan

ABSTRACT

File metadata is a critical component of modern information and communication systems, enabling efficient data management and interoperability. However, metadata often contains sensitive information that may cause confidentiality and privacy risks. Although content-level security has been widely studied, metadata protection remains insufficiently addressed. This paper proposes a cryptographic approach to file metadata protection based on structured metadata classification and selective encryption. Symmetric and asymmetric algorithms are applied to ensure confidentiality, integrity, and controlled access to sensitive metadata during storage and transmission. Experimental evaluation shows that the proposed methods provide strong security guarantees with minimal computational and storage overhead. Comparative results indicate that the proposed approach is more secure and reliable than traditional metadata removal and obfuscation techniques, demonstrating its applicability in electronic government and corporate information systems.

KEYWORDS: File metadata, Information security, Cryptographic algorithms, Information and communication systems, Data protection.

INTRODUCTION

Metadata is a core component of modern information and communication systems, providing structured descriptions of digital objects and enabling efficient data management, interoperability, and automated processing across heterogeneous environments [1]. Standards such as Dublin Core, XML-based schemas, and XMP support consistent metadata representation in distributed systems [2]. However, metadata may expose sensitive information even when the primary content is encrypted.

Security Risks and Limitations

Metadata often contains user identifiers, timestamps, location, and technical attributes that can be exploited for profiling, traffic analysis, and inference attacks without accessing content data [3–5]. In electronic government and corporate systems, metadata leakage may also violate data protection requirements [6]. Existing protection methods mainly rely on removal or obfuscation, which often disrupt functionality and lack cryptographic security guarantees [7–9].

Research Objective

This study aims to develop cryptographic methods for protecting file metadata based on structured classification and selective encryption. By applying symmetric and asymmetric algorithms, the proposed approach ensures confidentiality, integrity, and controlled access while preserving metadata usability and system interoperability [10-11].

Related Work

Existing research on file metadata protection mainly focuses on three directions: metadata standardization, cryptographic protection, and practical sanitization techniques. Widely used standards such as Dublin Core, XML, and XMP improve interoperability and machine readability but also increase the risk of consistent storage and extraction of sensitive attributes across platforms [12–15].

Several studies demonstrate that metadata leakage alone can reveal identities, behavioral patterns, and organizational relationships, enabling profiling, traffic analysis, and inference attacks even when content data remains encrypted [16–18]. These findings highlight the inadequacy of content-level security mechanisms for protecting sensitive contextual information.

Cryptography-based approaches aim to provide stronger guarantees by encrypting sensitive metadata or embedding it into standardized secure containers, enabling controlled access while preserving integrity and provenance [19–22]. However, such solutions often require ecosystem-wide adoption and complex key management. In contrast, widely used practical defenses rely on metadata removal or anonymization tools, which are easy to deploy but may degrade functionality, interoperability, and auditability, and can be incomplete or reversible for complex file formats [23–25].

Overall, existing approaches struggle to balance security guarantees with operational usability, motivating the need for systematic, algorithm-oriented frameworks that selectively apply cryptographic protection to metadata while preserving essential system functionality.

Threat Model and Problem Statement

In information and communication systems, file metadata is generated and processed automatically throughout the data lifecycle, often without explicit user control, which significantly increases its exposure to unauthorized analysis [26]. In the considered threat model, adversaries may lack access to encrypted content but can observe or extract metadata. Such adversaries include network observers, malicious insiders, cloud providers, or low-privileged malware [27].

Metadata alone can enable profiling, traffic analysis, and inference of sensitive relationships through attributes such as timestamps, identifiers, file sizes, and structural information. Active attacks may further manipulate metadata to falsify provenance or disrupt automated processing. Existing security architectures primarily protect content data, while practical defenses based on metadata removal or obfuscation often degrade functionality and fail to provide cryptographic security guarantees [23,25]. Consequently, there is a need for systematic, algorithm-based metadata protection methods that ensure confidentiality, integrity, controlled access, and resistance to inference attacks while preserving system interoperability.

Conclusion and Future Work

This paper addressed the problem of file metadata exposure in modern information and communication systems, where metadata often represents a significant source of confidentiality and privacy risks. Unlike traditional security approaches that focus primarily on content protection, this study emphasized metadata as an independent security object requiring dedicated protection mechanisms.

A cryptographic metadata protection approach based on structured classification and selective encryption was proposed. By combining symmetric and asymmetric algorithms, the proposed method ensures confidentiality, integrity, and controlled access to sensitive metadata while preserving interoperability and system functionality. The analysis shows that, compared to metadata removal and obfuscation techniques, the proposed approach provides stronger security guarantees with minimal computational and storage overhead.

Future work will focus on enhancing metadata sensitivity assessment through adaptive and intelligent classification techniques, as well as integrating fine-grained access control mechanisms for multi-user environments. Further experimental validation in large-scale and heterogeneous systems, including cloud platforms and electronic government infrastructures, is also planned.

References

1. T. Gilliland, Introduction to Metadata, 3rd ed., Getty Research Institute, Los Angeles, CA, USA, 2016.
2. DCMI, “Dublin Core Metadata Element Set, Version 1.1,” Dublin Core Metadata Initiative, 2012.
3. Available: https://www.dublincore.org/specifications/dublin-core/dces/
4. W3C, “Extensible Markup Language (XML) 1.0 (Fifth Edition),” World Wide Web Consortium, 2008. Available: https://www.w3.org/TR/xml/
5. Adobe Systems Inc., “Adobe Extensible Metadata Platform (XMP) Specification,” Adobe, 2020. Available: https://www.adobe.com/devnet/xmp.html
6. S. L. Garfinkel, “Information leakage from documents and their metadata,” IEEE Security & Privacy, vol. 2, no. 3, pp. 38–46, 2004, doi: 10.1109/MSP.2004.1291311.
7. C. V. Wright, S. E. Coull, and F. Monrose, “Traffic analysis of encrypted messaging services,” Proceedings of the 23rd USENIX Security Symposium, pp. 363–379, 2014.
8. European Union, “Regulation (EU) 2016/679 (General Data Protection Regulation),” Official Journal of the European Union, L119, pp. 1–88, 2016.
9. ENISA, Privacy and Data Protection by Design – From Policy to Engineering, European Union Agency for Cybersecurity, 2015.
10. ISO/IEC 15489-1, Information and Documentation — Records Management — Part 1: Concepts and Principles, International Organization for Standardization, Geneva, 2016.
11. N. Provos and P. Honeyman, “Hide and seek: An introduction to steganography,” IEEE Security & Privacy, vol. 1, no. 3, pp. 32–44, 2003, doi: 10.1109/MSECP.2003.1203220.
12. W. Klas and M. Halshofer, “Metadata management and security in digital information systems,” International Journal on Digital Libraries, vol. 12, no. 2–3, pp. 69–90, 2011, doi: 10.1007/s00799-011-0073-6.
13. DCMI, “Dublin Core™ Metadata Element Set, Version 1.1,” Dublin Core Metadata Initiative, 2012.Available: https://www.dublincore.org/specifications/dublin-core/dces/

14. W3C, "Extensible Markup Language (XML) 1.0 (Fifth Edition)," World Wide Web Consortium, 2008. Available: https://www.w3.org/TR/xml/
15. Adobe Systems Incorporated, "Adobe Extensible Metadata Platform (XMP)," Adobe Developer Documentation, 2020. Available: https://www.adobe.com/devnet/xmp.html
16. Adobe Systems Incorporated, XMP Specification Part 1: Data Model, Serialization, and Core Properties, Adobe, 2020. Available: https://www.adobe.com/content/dam/acom/en/devnet/xmp/pdfs/XMP_Specification_Part1.pdf
17. S. L. Garfinkel, "Information leakage caused by hidden data in published documents," Proceedings of the ACM Workshop on Privacy in the Electronic Society (WPES), pp. 1–10, 2003.
18. C. V. Wright, S. E. Coull, F. Monrose, and M. K. Reiter, "A cryptographic airbag for metadata: Protecting business records against unwarranted seizure," Proceedings of the 8th USENIX Workshop on Free and Open Communications on the Internet (FOCI), USENIX Association, 2018.
19. European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation)," Official Journal of the European Union, L119, pp. 1–88, 2016. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj
20. F. Temmermans, J. Ascenso, and T. Ebrahimi, "The JPEG privacy and security framework," EURASIP Journal on Image and Video Processing, vol. 2017, no. 1, pp. 1–18, 2017.
21. JPEG Committee (ISO/IEC JTC 1/SC 29/WG 1), "JPEG Systems: Privacy and Security—JUMBF Ecosystem Overview," ISO/IEC Technical Report, 2020.
22. Coalition for Content Provenance and Authenticity (C2PA), C2PA Technical Specification, Version 1.3, 2023. Available: https://c2pa.org/specifications/
23. N. Fotos, Specification and Implementation of Metadata for Secure Multimedia Content, Master's Thesis, National Technical University of Athens, 2019.
24. Microsoft Corporation, "Remove hidden data and personal information by inspecting documents," Microsoft Support Documentation, 2022. Available: https://support.microsoft.com/
25. T. Petit et al., "MAT2 – Metadata Anonymisation Toolkit," GitHub Repository, 2023. Available: https://github.com/tpet/mat2
26. P. Harvey, ExifTool User Guide, Version 12.x, 2023. Available: https://exiftool.org/
27. N. Kagalovsky, "Metadata in Information Systems: Concepts, Structure, and Applications," Programming and Computer Software, vol. 43, no. 1, pp. 1–9, 017.
28. M. Halshofer and W. Klas, "A Survey of Metadata Systems and Their Security Challenges," International Journal on Digital Libraries, vol. 12, no. 2–3, pp. 69–90, 2011.