Ensuring privacy in provenance information for images

Nikolaos Fotos
Computer Architecture Department
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
nikolaos.fotos@upc.edu

Jaime Delgado
Computer Architecture Department
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
jaime.delgado@upc.edu

***Abstract*—Current technologies facilitate the generation of falsified images that spread misinformation and endanger trust between users and media. Technological and academic institutions have contributed towards building frameworks that try to combat the aforementioned issue by establishing verifiable mechanisms to annotate metadata regarding the creation and modifications of an image (i.e., provenance). However, this type of metadata might conceal information about a specific person or place which threatens the privacy of individuals and locations. To address this problem, we propose the extension of existing provenance schemas in order to provide privacy-enabling features that allow users to gain control over the verifiable data that they intend to share. The proposal examines the provenance schema of the Coalition for Content Authenticity and Provenance (C2PA) specification and extends it using the JPEG Privacy and Security standard (ISO/IEC 19566-4) in order to support the protection of privacy-related information in images. Part of this work has been submitted as a contribution to the standardization activities of the ISO/IEC JTC 1/SC 29/WG 1 "JPEG Coding of digital representations of images" committee.**

Keywords — privacy, security, provenance, images, authenticity, JPEG

I. Introduction

Recent advances in technology have facilitated the development of easy-to-use tools which can be used to generate falsified information. Specifically, in the domain of media sharing, the abundance of fake news has posed a tremendous peril in the way societies are shaped. Inevitably, tools like machine learning have been used for both adversarial and benign use. Thus, trying to develop the best model to address the erosion of information is an everlasting challenge.

To address this issue, an alternative approach has been proposed by various organizations, suggesting that misinformation could be mitigated by providing reliable and secure tools to annotate the entire history of a media asset. This history eventually traces back to the origin of the asset, thus, allowing a media consumer to inspect the steps that have been applied to an image from its initial to its final form. According to the definition of the World Wide Web Consortium (W3C) [1]: "[...] provenance is defined as a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing". By describing provenance information for media in a tamper-proof and secure way, it is possible to establish a certain level of trust among media consumers.

Ensuring privacy in provenance data for images is a critical feature that provenance data models should support for various use cases. Provenance data might disclose information about the person and the location that are involved in an image. For example, a journalist might prefer to hide her identity when capturing an image that proves a case of infringement of human rights. The provenance data of that image would still trace back to the origin of the image (i.e., the pixel data that were captured via the journalist's camera), but the journalist is protected from being victimized. Similarly, an artist wants to distribute an image with her artwork while protecting the provenance data that describe the exact steps that have led to the generation of the art work as it constitutes her intellectual property. Even more, she might want to define a set of access rules that will grant access only to a set of media consumers to whom she decides to disclose these metadata.

Overall, there have been many initiatives which have focused on establishing a chain of provenance consisting of all the metadata and modifications that accompany an asset [1], [2], [3], [4], [9]. However, this might pose tremendous privacy risk as provenance might disclose sensitive information. Therefore, this paper defines the Multimedia Information Protection And Management System (MIPAMS) Provenance specification which sets the basis for identifying how current provenance schemes could support the privacy-related use cases. It also defines the mechanisms to reconstruct the protected data in order to be consumed by authorized users, thus offering lossless protection of sensitive information. The specification focuses on providing privacy features for provenance data related to images but it could be extended to other content such as video or even health information.

The remainder of this paper is structured as follows: Section 2 presents the background information, followed by Section 3 which presents the MIPAMS Provenance specification. Next, Section 4 demonstrates how this specification could be applied through an existing application. Subsequently, in Section 5 the information on how the proposed specification has contributed to the activities of the

979-8-3503-3959-8/23/$31.00

JPEG working group is described. Finally, Section 6 summarizes conclusions and future work.

II. BACKGROUND

Initiatives from the industry have tried to address the problem of misinformation using various approaches. Some of those were focusing on finding a way to express provenance information in a secure and reliable way which allows the description and characterisation of all the events that have taken place in a given digital asset. Microsoft's Project Origin [2] and Adobe's Content Authenticity Initiative (CAI) [3] were two impactful initiatives from the industry which specified specific architectures that could provide authenticity in media. Both approaches specified the means to express and serialize provenance information. Furthermore, Project Origin defined the mechanism to record all possible transactions into a permissioned ledger to ensure transparency and integrity. To promote interoperability and to increase the adoption of such technologies, the Coalition for Content Provenance and Authenticity (C2PA) was created unifying existing efforts. C2PA has published a specification [4] for modeling provenance information and providing various recommendations on how such systems should be built. The technical specification provides an explicit model to express provenance information. Its schema is presented in Figure 1. In brief, all provenance statements (modifications of digital asset, metadata, etc.) are called "Assertions", and, along with the digital signature and certificate of the actor who asserts the statements, are grouped into a structure called C2PA Manifest. Based on this model, the entire history of a digital asset can be expressed as a tree-like structure of multiple C2PA Manifest structures.

The serialization of the proposed provenance schema is achieved by adopting a container format which is standardized by the ISO/IEC JTC 1/SC 29/WG 1 "Joint Photographic Experts Group (JPEG) technical committee [5]. This container format is called JPEG Universal Metadata Box Format (JUMBF, ISO/IEC 19566-5) [6] and it is based on the Base Media File Format (BMFF) specified in ISO/IEC 14496-12 [7]. This allows JUMBF to be compatible with many common image and video file formats as well.

In addition, the JPEG technical committee has published a standard aiming to address the protection of JUMBF metadata. Specifically, JPEG Privacy & Security standard (ISO/IEC 19566-4) [8] provides the means to protect part of the image and/or part of the metadata which are embedded in an image codestream. This is achieved by defining two extensions of the JUMBF serialization, namely the Protection and the Replacement Content type JUMBF boxes - which allow the signaling of various parameters related to the protection (i.e., encryption and application of access control rules) of the protected resource and its replacement with placeholder data.

Finally, JPEG has identified the need for standardization in the area of defining clear annotations which can describe the modifications and the metadata of a digital asset. Therefore, a new activity titled as "JPEG Fake Media" [9] was launched in order to explore user cases & requirements, evaluate existing contributions and develop new standards towards enabling

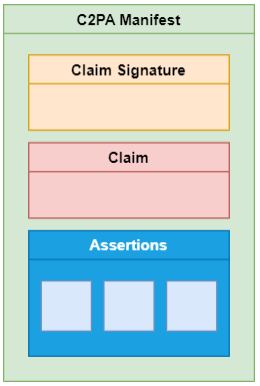

Fig. 1. *C2PA Base Model.*

trust in media. Part of the identified requirements focused on specifying the means to ensure privacy of users and locations, as it was a missing point of the current technologies that have been proposed.

III. ARCHITECTURE

In general, provenance information could be conceived as a set of all the events that characterize, with respect to an asset, its generation, its modifications and its associated metadata. Each event might consist of one or more actions that are related to the asset and may be generated by multiple entities as well. To ensure that this information is not tampered with, it is essential that a provenance model specifies means to apply integrity measures (e.g., calculate the SHA-256 [10] of a provenance scheme or a subset of it). However, some of the actions of a provenance event might disclose information about a human or a location that could constitute a leak of privacy information.

Therefore, this paper proposes the MIPAMS Provenance specification that defines the extension of existing provenance schemes in order to support the protection of part of the information that a user identifies as confidential. This is achieved using cryptographic tools to protect the privacy-sensitive actions. Specifically, this is achieved by encrypting the aforementioned metadata and allowing the ability to provide access control rules to regulate the access on parts of the provenance information. This specification could be applied to provenance specifications that can identify when part of the model is encrypted or not. In addition, a provenance scheme should be able to provide a data model that is modular and define a referencing scheme (e.g. URI [11]) where parts of the model could be located outside of the provenance structure. Additionally, it is essential that protection capabilities are applicable only to the provenance actions that could disclose sensitive information. This is essential in order to keep a balance between the right to privacy and also the right to know.

On a more technical level, the paper provides a specification on how privacy-related provenance information could be protected in the case of digital assets and, specifically, for JPEG encoded images. Existing provenance schemes for digital assets support various technologies for serialization (e.g. XML [12], JSON [13], CBOR [14]). Regardless of how a provenance scheme serializes its data, it

can be embedded in a JPEG encoded image by encapsulating the payload using JUMBF, a generic, container, box-like format that is supported by all JPEG encoded images.

MIPAMS Provenance specification takes advantage of the JUMBF data model capabilities and proposes the signaling and serialization of encrypted, privacy-related provenance actions using the Privacy & Security standard. Hence, it is possible to serialize the provenance information in a newly defined JUMBF Box but also store each protected piece of information in a separate JUMBF Box, specifically, of type Protection Content type JUMBF box. The fact that JUMBF defines its own referencing scheme, allows a provenance data model to reference protected serialized information by providing the corresponding JUMBF URI.

Figure 2 illustrates how a provenance data model is serialized using JUMBF and how the corresponding protected content could be serialized with a Protection Content type JUMBF box as specified in JPEG Privacy and Security standard [8]. Specifically, depending on how provenance information is expressed, the serialized payload could be embedded in a JPEG encoded image using the corresponding JUMBF Content Type. For instance, if a provenance data model is expressed in XML, then the XML Content type JUMBF box is used. Using the JUMBF reference scheme, the provenance scheme annotates the existence of protected information by specifying the JUMBF URI which points to the label of the Description Box located in the corresponding Protection Content type JUMBF box. The latter box contains information about the encryption procedure in the Protection Description box while the ciphertext - which corresponds to the provenance data - is serialized in the Binary Data box.

The key attribute of MIPAMS Provenance specification is that it provides privacy functionalities to provenance information of JPEG encoded images, taking advantage of some features of the JUMBF and Privacy and Security JPEG standards. Thus, it requires a single software which is compliant to JPEG Systems specification. This is of paramount importance as it facilitates the adoption of a provenance specification by the community.

IV. Application

To showcase the applicability of the MIPAMS provenance specification, an existing provenance scheme is considered. Specifically, the paper illustrates how the C2PA provenance scheme could be adopted in order to comply with the proposed specification. To further support the aforementioned use case, an implementation has been developed demonstrating how an end-user could consume provenance information. The application is open source and the link to the repository can be found here: https://github.com/dmag-upc/mipams-fake-media-demonstrator.

Regarding the C2PA provenance scheme, each provenance event is translated into a C2PA Manifest consisting of - possibly - multiple actions which are called C2PA Assertions. The integrity of all C2PA Assertions is recorded into the C2PA Claim structure. Each of these structures is serialized using JUMBF. To extend the C2PA

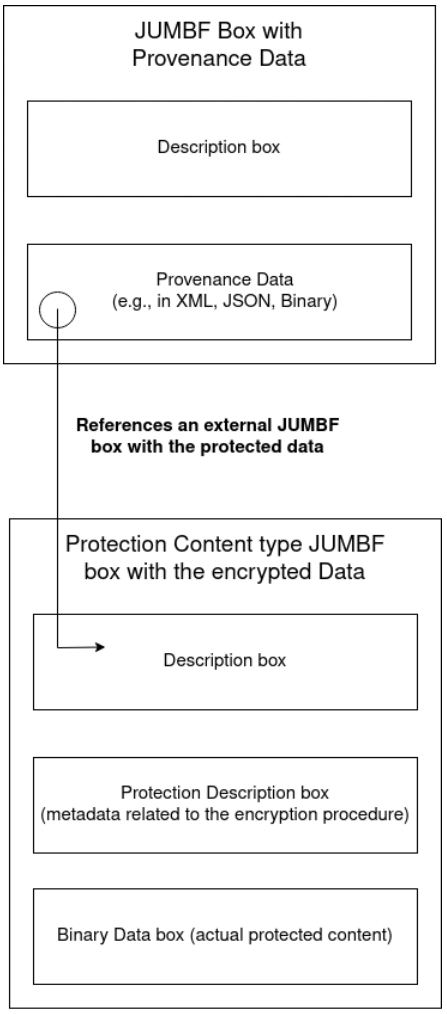

Fig. 2. *Supporting privacy functionalities in provenance data models serialized with JUMBF.*

specification according to the MIPAMS Provenance one, it is suggested that the conversion of C2PA Assertion JUMBF boxes into Protection Content type JUMBF boxes should be supported. This would allow the provenance scheme to support the protection of privacy-sensitive information.

In the scope of the MIPAMS Provenance specification, a web application has been developed which showcases the end-user interaction with provenance information, including protected pieces of data. The web application requires the functionality to handle JUMBF data in order to serialize/deserialize provenance information. As a result, a library is developed and included as a dependency to the web application. This library is called MIPAMS JPEG Systems and the link to the repository can be found here: https://github.com/dmag-upc/mipams-jpeg-systems.

To demonstrate how image provenance data are consumed, a motivation example is defined. Initially, it is assumed that a user, say a journalist, is authenticated and authorized in our application and possesses a digital certificate which allows her to create and digitally sign provenance information. Given that the user has already taken an image, she uploads it to the web application which allows her to perform a set of modifications

such as cropping, filtering, blurring etc. All the modifications are recorded as provenance actions. Subsequently, the application identifies that the JPEG encoded image has embedded metadata in Exchangeable image file format (EXIF) [15], that disclose location and user information. She decides to protect this information by encrypting it. The provenance data are generated and embedded to the JPEG encoded image which is, now, ready for distribution.

A consumer user can now download the image from any digital repository and inspect its provenance metadata. In Figure 3, the user interface of the web application is depicted. It is assumed that a JPEG encoded image has been created and manipulated using an editing tool. All the events have been recorded in C2PA Manifest Store structure and embedded inside the image. By default, only the latest C2PA Manifest data (i.e., latest provenance actions) are shown to the user. However, the user has the ability to audit the entire provenance history of the image by selecting the "FULL INSPECTION" button.

The user who has made the modifications decides to include privacy sensitive information, but allow access only to authenticated users of this web application. Consequently, the sensitive information is encrypted and serialized with a Protection Content type JUMBF box while the access control rules are stored on a separate JUMBF Box - based on the Privacy & Security standard recommendations. The end user that tries to access the provenance information is assumed not to be authenticated. As it is clearly depicted in Figure 3, the provenance data are validated for their integrity and authenticity according to C2PA specification. However, since the user is not authorized to access the protected resource, a corresponding message is shown to the interface, notifying her that although the information is valid, there is a piece of information that cannot be accessed based on her current privileges.

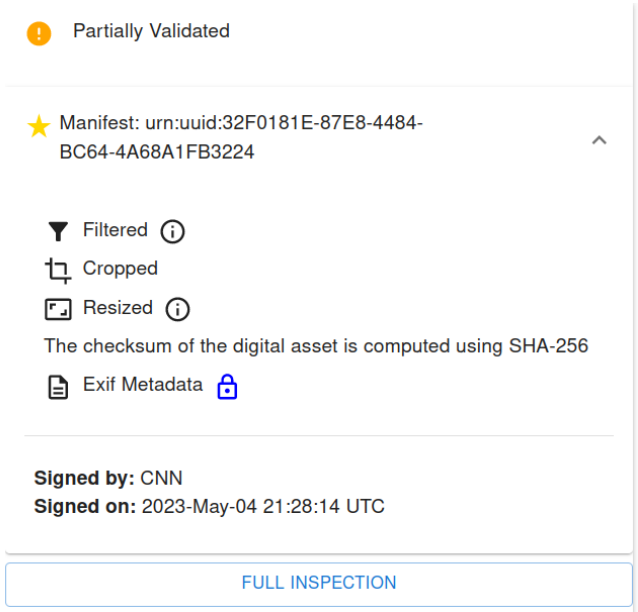

Fig. 3. *Inspection of Provenance information.*

V. STANDARDIZATION

In April 2022, a Call for Proposals [9] was issued by the JPEG working group concerning the topic of combating Fake Media. The requirements of this call [16] sought proposals to provide a set of tools that could allow the secure and reliable annotation of provenance information including the respect to privacy concerns. The goal of these tools is to facilitate the annotation of provenance information to address use cases that describe both malicious modifications but also modifications of good intent.

The MIPAMS Provenance specification - presented in this paper - was submitted as a contribution [17] to the JPEG Fake Media activity and it showcased how it is possible to use existing JPEG standards to ensure privacy on top of provenance information for JPEG encoded images. One of the main requirements of JPEG Fake Media is that the developed provenance schema should be serialized using the JPEG Universal Metadata Box Format. Therefore, by using the JUMBF specification, a provenance data schema could leverage the entire ecosystem that is built around JUMBF such as the standard related to Privacy & Security or its reference scheme.

After October 2022, the six submitted proposals were evaluated for their relation to the requirements [18]. MIPAMS Provenance specification was the only proposal that provided conditional access to privacy sensitive information. The next step was to identify the foundational provenance data model that can set the basis for the development of a new standard for establishing trust in media throughout its lifecycle.

VI. CONCLUSION

In recent years, there has been an abundance of high-end, easy-to-use tools which allow users to generate images that are almost indistinguishable from the original ones. Organizations have developed provenance data models to help annotate the origin and modification history of a piece of data. Even though this is crucial towards establishing trust in media, current provenance data model approaches have not adequately dealt with the privacy concerns that such an effort might raise. To that extent, this paper has presented how it is possible to build a privacy-oriented layer on top of existing provenance data models. MIPAMS Provenance specification covers use cases where privacy needs to be preserved in images.

As future work, it is crucial to assess the generalization of the proposed specification by evaluating the ability to be applicable to other existing provenance schemes like the W3C Provenance scheme [1]. In addition, it would be interesting to investigate new use cases with respect to privacy preserving functionalities. One of them is the protection of provenance events describing modifications before the final version of the content.

ACKNOWLEDGMENT

This work is partly supported by the Spanish Government (GenClinLab-Sec, PID2020-114394RB-C31) and by the Generalitat de Catalunya (2017 SGR 1749, IMP - Information Modeling and Processing Research Group).

References

[1] P. Missier, K. Belhajjame, and J. Cheney, “The W3C PROV family of specifications for modelling provenance metadata,” in ACM International Conference Proceeding Series, 2013, pp. 773–776. doi: 10.1145/2452376.2452478.

[2] P. England et al., “AMP: Authentication of media via provenance,” in MMSys 2021 - Proceedings of the 2021 Multimedia Systems Conference, 2021, pp. 109–121. doi: 10.1145/3458305.3459599.

[3] L. Rosenthol et al., “The Content Authenticity Initiative, Setting the Standard for Digital Content Attribution” Content Authenticity Initiative https://acrobat.adobe.com/link/track?uri=urn%3Aaaid%3Ascds%3AUS%3A2c6361d5-b8da-4aca-89bd-1ed66cd22d19&viewer%21megaVerb=group-discover (accessed Mar. 17, 2023)

[4] L. Rosenthol, "C2PA: the world’s first industry standard for content provenance," Applications of Digital Image Processing XLV. Vol. 12226. SPIE, 2022.

[5] JPEG. https://www.jpeg.org/. (accessed Mar. 17 2023)

[6] ISO/IEC JTC 1/SC 27, ISO/IEC 19566-5:2019/Amd 1:2021 – Information Technology – JPEG systems – Part 5: JPEG universal metadata box format (JUMBF) (2021). ISO/IEC JTC 1/SC 27

[7] ISO/IEC JTC 1/SC 29, ISO/IEC 14496-12:2022 – Information Technology – Coding of audio-visual objects – Part 12: ISO base media file format (2022).

[8] ISO/IEC JTC 1/SC 29, ISO/IEC 19566-4:2020 – Information Technology – JPEG systems – Part 4: Privacy and security (2020).

[9] ISO/IEC JTC 1/SC29/WG1 N100157, REQ "Final Call for Proposal for JPEG Fake Media", 95th Meeting, Online, April 2022. https://ds.jpeg.org/documents/jpegfakemedia/wg1n100157-095-REQ-Final_Call_for_Proposals_for_JPEG_Fake_Media.zip/ (accessed Mar. 11, 2023)

[10] H. Krawczyk, M. Bellare, and R. Canetti, “RFC2104: HMAC: Keyed-Hashing for Message Authentication.” RFC Editor, 1997.

[11] T. Berners-Lee, R. Fielding, and L. Masinter, “RFC3986: Uniform resource identifier (URI): Generic syntax.” RFC Editor, 2005.

[12] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, F. Yergeau, “Extensible Markup Language (XML) 1.0 (Fifth Edition),” W3C, https://www.w3.org/TR/2008/REC-xml-20081126/ (accessed Mar. 17 2023)

[13] T. Bray, “The JavaScript Object Notation (JSON) Data Interchange Format,” RFC Editor, 2017.

[14] C. Bormann, P. Hoffman, “RFC7049: Concise Binary Object Representation (CBOR),” RFC Editor, 2013.

[15] "Exchangeable image file format for digital still cameras: Exif Version 2.3," Tech. Rep. JEITA CP-3451, 2010.

[16] ISO/IEC JTC 1/SC29/WG1 N100156, REQ "Use Cases and Requirements for JPEG Fake Media", 95th Meeting, Online, April 2022. https://ds.jpeg.org/documents/jpegfakemedia/wg1n100156-095-REQ-Use_Cases_and_Requirements_for_JPEG_Fake_Media.pdf/ (accessed Mar. 12, 2023)

[17] ISO/IEC JTC 1/SC29/WG1 M97036 REQ “UPC’s answer to the Call for Proposals for JPEG Fake Media: MIPAMS Provenance module”, 96th meeting, Online, October 2022. https://github.com/DMAG-UPC/mipams-fake-media-demonstrator/docs/ (accessed Mar. 17, 2023)

[18] ISO/IEC JTC 1/SC29/WG1 N100388, REQ "Updated report on the JPEG Fake Media Call for Proposals", 98th Meeting, Sydney, Australia, January 2023. https://ds.jpeg.org/documents/jpegfakemedia/wg1n100388-098-REQ-Updated_Report_on_the_JPEG_Fake_Media_Call_for_Proposals/ (accessed Mar. 11, 2023)