
Proceedings Paper:
Rainey, J., Elawady, M., Abhayartne, C. et al. (1 more author) (2023) TRAIT: A trusted media distribution framework. In: 2023 24th International Conference on Digital Signal Processing (DSP). 2023 24th International Conference on Digital Signal Processing, 11-13 Jun 2023, Island of Rhodes, Greece. Institute of Electrical and Electronics Engineers (IEEE) . ISBN 979-8-3503-3959-8

https://doi.org/10.1109/dsp58604.2023.10167909

TRAIT: A Trusted Media Distribution Framework

James Rainey*, Mohamed Elawady†, Charith Abhayaratne‡ and Deepayan Bhowmik*
*School of Computing, Newcastle University, Newcastle upon Tyne, UK
†Computer and Information Sciences, University of Strathclyde, Glasgow, UK
‡Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, UK
{*james.rainey, deepayan.bhowmik*}*@newcastle.ac.uk, mohamed.elawady@strath.ac.uk, c.abhayaratne@sheffield.ac.uk*

***Abstract*—Trusted distribution and consumption of media content has become a challenging issue, especially with the advancement of machine learning-based techniques such as deep fake. To address such challenges, this paper proposes a new metadata schema which is embedded within a larger framework that facilitates trusted media distribution. This schema is realised through a distributed media blockchain core in conjunction with algorithms to detect media modifications. Such a framework is expected to improve trust in media consumption, ensuring media integrity, authenticity and provenance.**

I. Introduction

The emergence of machine learning-based image manipulation opens up new possibilities in the creative sector, as it is now possible to produce near-realistic media assets. A recent example includes 'For All Mankind', which uses deepfake technology to bring back well-known characters such as Johnny Carson, John Lennon, and Ronald Reagan [1]. However, it equally causes the creation of fake media that spreads misinformation and plays havoc in many aspects of human society, from political unrest to financial harm [2]. As a result, public trust is being diminished in any media they consume. Therefore, it is necessary to create a solution, even better, a framework that can help media distribution to be transparent and trusted. In an attempt to address such gaps, this paper proposes a **TR**usted Medi**A** DI**s**Tribution (TRAIT) framework that provides a metadata schema in Extensible Markup Language (XML) and its implementation using Extensible Metadata Platform (XMP). An example of image manipulation and detection using the TRAIT framework is shown in Fig. 1.

In achieving trust in media distribution, information such as ownership, copyright, intellectual property rights (IPR), provenance, integrity, authenticity *etc*. are crucial. Literature suggests individual tools, techniques and software are available both in academia we well as the industry. For example, ownership, copyright, and IPR are available through 1) XIF (eXtended Image Format), JPEG Systems Part 4: Privacy and Security [3] or other metadata formats; 2) digital watermarking [4] and 3) Digital Right Management software, *e.g.*, Imagen[1], VdoCipher[2]. Tracing back the previous history or even the origin, in short, the provenance of an image is important, especially in relation to media distribution and has been a topic of recent interest [5], [6]. The integrity of an image is commonly verified using content hashing[3] and file hashing. Recently, content authenticity has generated significant interest

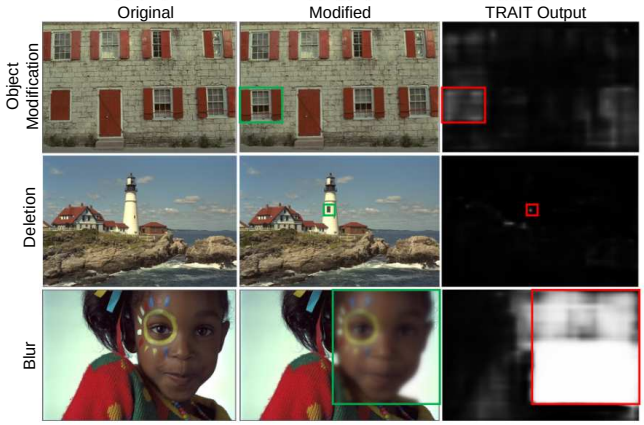

Fig. 1: Examples of fake media (image manipulation): object modification, cropping and blur. 1st and 2nd columns are the original and the modified versions of images, and 3rd column represents the output of TRAIT framework. The modified regions have been highlighted in green (2nd column) and red colors (3rd column).

in the community, especially due to the advancement of photo editing tool sets and machine learning-based models such as *deep fake*. A significant number of detection algorithms are available in the literature, and we briefly discuss the notable ones here.

File-based cryptographic/hashing algorithms are commonly used to verify the integrity and authenticity of a file by generating a unique digital fingerprint. The most popular hashing algorithms are SHA1 (1995), MD5 (1995), SHA256 (2002) and Blake3 (2020), which are different in security levels and computation time. However, file hashing is not always suitable for media files as often media contains remain the same, but file hashes change due to file format conversion or compression. Considering image-based contextual information, perceptual hashing algorithms are introduced to focus on pixel changes inside an image by looking into visual features (*i.e.*, color, texture *etc*.) in order to generate a unique hash value. In the pre-deep learning era, there were traditional/hand-crafted approaches for perceptual hashing: Average Hash (A-Hash) [7], Perceptual Hash (P-Hash) [8], Singular Value Decomposition Hash (SVD-Hash) [9], Wavelet Hash (W-Hash) [10] and Laplace-based Hash (L-Hash) [11]. More recently, the learning-based (deep) hashing approaches have become popular [12]–[14].

Although the perceptual hashing is robust against clearly visible image manipulation (*i.e.*, cropping, rotation *etc*.), it

[1]https://imagen.io/
[2]https://www.vdocipher.com/
[3]https://iscc.codes/

is vulnerable to barely visible / illusion-based attacks (*i.e.*, splicing, local region removal/editing). In overcoming that, literature proposed recent deep-based approaches for fake media/tampering detection such as ManTra-Net [15], SPAN [16], MVSS [17], [18] and EMT-Net [19].

While techniques for individual components are generally available, there is no/very little existing mechanism that holistically provides means to transparently communicate such information to the end user in a comprehensive manner. To the best knowledge of the authors, the closest and current attempt is made by an industry consortium C2PA (Coalition for Content Provenance and Authenticity)[4] led by Adobe Inc.[5]. C2PA provides a metadata schema and an implementation mechanism (through JPEG universal metadata box format (JUMBF) [20]). However, notable exclusions in C2PA include the ability to detect tampering or other modifications in the media and the availability of metadata in a centralised or decentralised repository. To address these gaps, TRAIT framework is proposed (see Section III) in this paper. Main contributions are:

- Proposition of a file format agnostic new metadata schema for trusted media distribution. This is achieved through an XML-based schema description and an XMP-based (widely adopted to various file formats) implementation for metadata embedding.
- Development of a modular framework that uses blockchain and IPFS (InterPlanetary File System) as its core components to facilitate secure and transparent means to manage and share relevant information and store the assets in a distributed file server.
- Demonstration of TRAIT framework capability through two real-life inspired use cases.

In this context, it is worth noting that JPEG[6] is making a substantial effort to address the challenges of *media trust* through an upcoming international standard. Among others, the proposed TRAIT framework is actively engaged in this process and envisages contributing in a significant manner.

II. JPEG STANDARDISATION ON FAKE MEDIA

As part of the international standardisation body Joint Photographic Experts Group (JPEG) committee (ISO/IEC JTC 1/SC 29/WG 1) initiated a standardisation effort called JPEG Fake Media in order to provide a mechanism that facilitates a secure and reliable annotation of media asset creation. The standardisation aims to support both use cases that are in good faith, *e.g.*, media creation for entertainment or marketing, as well as those with malicious intent, *e.g.*, spreading misinformation. As part of this initiative, JPEG published a *Use Cases and Requirements* document [21] and issued a call for proposals [22] in April 2022. In response to this call, the proposed TRAIT framework was submitted, which complies with most of the requirements outlined in the call issued by JPEG. The TRAIT framework builds upon and extends the utility of the JPEG standard by proposing a metadata schema, its implementation, and example use cases that use this framework.

[4]https://c2pa.org/

[5]https://www.adobe.com/

[6]https://jpeg.org/

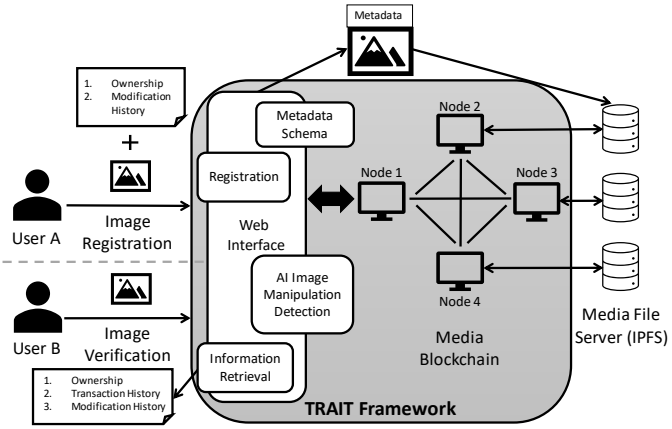

Fig. 2: Overview of the TRAIT Framework.

III. THE FRAMEWORK

A. Overview

The proposed framework consists of four components as depicted in Fig. 2: 1) Graphical Web Interface, 2) Media Blockchain, 3) Manipulation Detection engine and 4) Distributed File Server.

1) Graphical Web Interface provides an interface by which users can register new images, verify existing images and add modified versions of images to the TRAIT repository. For the implementation, REACT[7] (a front-end JavaScript library) is used for the interface and backend connectivity.

2) Media Blockchain is the core of the TRAIT framework which uses Hyperledger Fabric [23], [24] as its core technology. This is an extension of our previous work on *Multimedia Blockchain* [25] and is responsible for recording the transactions of media assets and verifying the integrity of the assets. Each transaction registered on the TRAIT framework consists of a number of fields as defined in the TRAIT schema shown in Fig. 3. Newly registered images are given a Media Unique ID (MUID), which is generated from a hash of the asset. Basic details such as *Name* and *Location* are entered by the user on the registration page of the web interface.

Details of any modifications are added automatically, including the probability, the method used, regions of interest, the type and category of the modifications along with a user-defined text field to provide a purpose for modification. Each uploaded image is assigned an ImageID and a Parent ID. The ImageID is unique to that image, the ParentID is the ImageID of the parent image, and both IDs are distinct from the MUID, except for the root image in which all 3 IDs are identical.

The inclusion of the Image ID and ParentID allows multiple distinct modified versions of an image to exist in parallel, even if they have the same parent image. This also allows the full transaction history of an image, including all modified child images, to be retrieved using a single ID. The metadata is embedded to the image file head using a new set of XMP tags as defined in the TRAIT metadata schema (Fig. 3).

[7]https://react.dev/

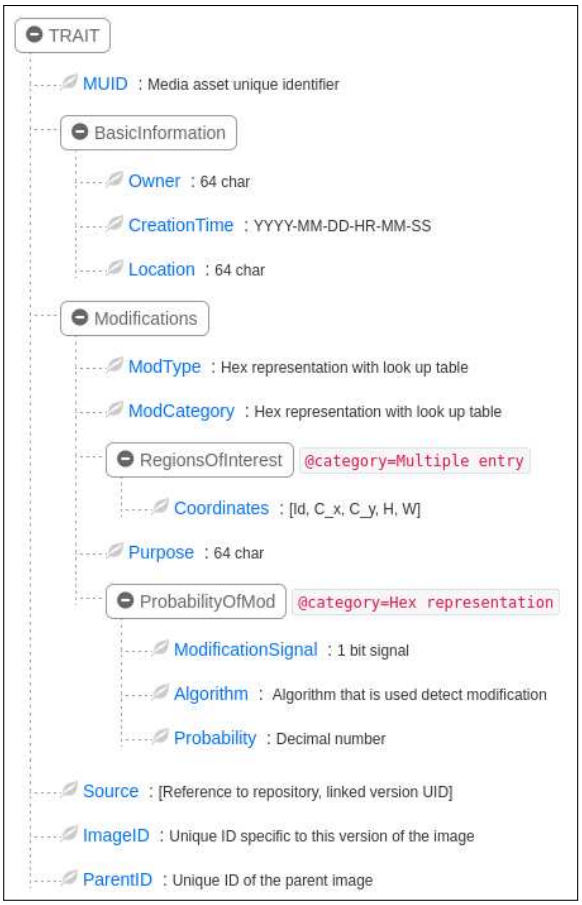

Fig. 3: TRAIT Metadata Schema.

3) Manipulation Detection Engine accepts input by uploading images with existing metadata, which results in a search using the embedded ID, after which images that have previously been registered are scanned for modifications using an image manipulation detection algorithm such as MVSS [18]. A probability (currently derived through MVSS) that the image contains modification is produced as well as a binary mask showing any modified areas detected in the image.

4) Distributed File Server manages the storage of registered images via the Inter Planetary File System [26]. Each subsequent modified version of an image is also stored on the file server and a link to the file is retained within the metadata stored in the image and on the related blockchain transaction.

B. Image Registration

Image registration is performed when a user uploads a new image that does not contain TRAIT metadata or does not have a hash matching any previously registered image. To successfully register a new image the user is also required to input their name and location. A new unique ID will be generated for the image from a hash of the image, using SHA256, and will be used for the MUID and ImageID. The modification data is initialised, with all fields set to ”None”. For a modified version of a previously registered image, the existing information from the previous transaction is used to initialise many of the data fields, other than the ImageID which is calculated for that specific image and the purpose field which is provided by the user. The metadata is then embedded into the XMP tags of the image file and a new transaction is created on the blockchain, both using the schema defined in figure 3. The image is now registered and ready for verification.

C. Image Verification

Image verification is performed when a user searches with, or tries to register, an image containing existing TRAIT metadata. The image hash is calculated and compared to the stored hash, if there is a difference the image is passed through the modification detection algorithm. The modification detection algorithm returns a probability that the image has been modified as well as a binary mask showing the likely areas in which the modifications occur. Alongside the modification results there is a list of each transaction that has occurred for the selected image, this allows the user to trace the asset modification history and to verify the authenticity and provenance of the image.

An image that has a transaction history and has produced a probability of modification over a threshold is flagged and can be registered as a new modified version of the image. Only the original creator of the image will have the ability to add this new transaction.

IV. Use Cases

We provide two real-life inspired use cases that demonstrate the application of the TRAIT framework. Use case 1 presents an example of a photographer using the framework to prevent copyright violations of their work. Use case 2 presents an example of an art collector verifying the authenticity and provenance of a physical artwork before purchasing it. Both of these use cases make use of the proposed TRAIT framework workflow consisting of the web interface, media blockchain, IPFS distributed file system and REST (Representational state transfer) API (to link the blockchain and IPFS) along with the choice of MVSS [18] algorithm for image modification detection.

A. Use Case 1: Photography copyright infringement

Photographers own the copyright on any original photographs that they create. Having a copyright gives them the exclusive right to distribute and sell their images; users of the images will need permission from the creator to use them in any capacity. Nowadays, throughout the digital world, copyright infringement cases have been prominent since the last decade [27] (*i.e.*, Shepard Fairey vs The Associated Press, Cariou vs Prince).

In this use case, User A, a photographer, takes a photograph with their camera and becomes the copyright owner for the image produced. They register the image using the TRAIT framework; a new set of metadata is generated containing a unique ID. The metadata is embedded in the image, a transaction containing the metadata is added to the blockchain and a copy of the image is stored on IPFS.

User B obtains the image and modifies it, applying a median filter to the image. They try to register the image claiming

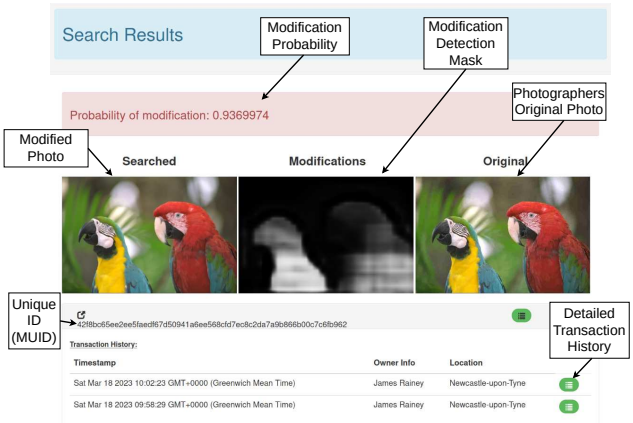

Fig. 4: Use Case 1 search result, using kodim23 image from Kodak dataset [28].

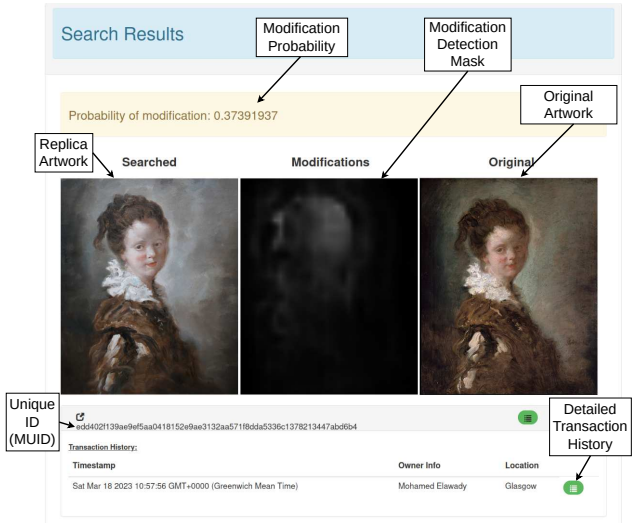

Fig. 5: Use Case 2 search result, using Fragonard's Young Woman [31].

to be the original photographer and infringing on the original creator's copyright. When they attempt to register the modified image on the TRAIT framework, the existing metadata is detected and the unique ID is found. This is used to look up the transaction history of the original image on the blockchain. The image is scanned by the image manipulation detection algorithm, which highlights the modifications to the image and subsequently registration of the image as a new media asset is stopped. Infringement of the original photographer's copyright is prevented. This is shown in Fig. 4.

However, with permission of the copyright holder, this modified image can be registered as a modification of the existing asset which would create a new transaction in the existing transaction tree and a new modification to the history of the asset.

B. *Use Case 2: Art forgery*

Buying and selling art can be a very lucrative business, however, this makes it a target for forgery. Tracking the authenticity and provenance of an artwork is an important stage in preventing forgeries and reproductions of artworks being maliciously or accidentally mistaken for original pieces. Several incidents of art forgery have occurred over the last century and have been sold for significant sums of money [29], [30], including Han van Meegeren's Vermeer and Elmyr de Hory's Matisse & Modigliani.

In this use case, User A, an art collector, takes a digital photograph of an artwork and registers it on the TRAIT framework. New metadata is generated for the image and is recorded in both the XMP tags of the image and a blockchain transaction. This allows the authenticity and provenance of the artwork to be verified if the collector wishes to sell it.

User B, a second art collector, wishes to buy the same artwork owned by user A that is being displayed in an art gallery. However, before making the purchase they want to verify the authenticity of the artwork. A digital photograph of the artwork is taken and a search is performed on the TRAIT framework. Unfortunately, User B received a *forged artwork* and the results show distinct differences and areas of modification compared to the original, as shown in Fig. 5, and the transaction history presented does not match the history suggested by the seller.

The differences between the images and the discrepancy in the provenance of the artwork indicate that it may not be authentic. As the collector is unable to verify that the artwork that they wanted to acquire is the original and not a reproduction, they do not complete the purchase.

V. Conclusions

We have presented a new framework which promotes improved trust and transparency in media distribution by ensuring the integrity, authenticity and provenance of media assets. The TRAIT framework provides a metadata schema which is implemented using a media blockchain and integrates media manipulation detection to allow the verification of the integrity of media assets. The use of file hashing combined with fake media detection facilitates the recognition of local changes between images. We provide two use cases to show the capabilities of the current TRAIT framework.

TRAIT framework expects to be expanded significantly in future to accommodate various needs in different industries, including but not limited to the creative industry, news syndicates, GLAM sector (galleries, libraries, archives, and museums), and insurance services. The framework will be made open-sourced to support the community and to build upon usable applications.

References

[1] B. Lindbergh, "How they made it: The deeply real deepfakes of 'for all mankind'," https://www.theringer.com/tv/2021/3/5/22314809/for-all-mankind-season-2-deepfakes-ronald-reagan-john-lennon-johnny-carson, March 2021, [Online; posted 5-March-2021].

[2] J. Botha and H. Pieterse, "Fake news and deepfakes: A dangerous threat for 21st century information security," in *ICCWS 2020 15th International Conference on Cyber Warfare and Security. Academic Conferences and publishing limited*, 2020, p. 57.

[3] "JPEG systems — Part 4: Privacy and security," International Standard ISO/IEC 19566-4:2020, 2020.

[4] D. Bhowmik and C. Abhayaratne, "Quality scalability aware watermarking for visual content," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5158–5172, 2016.

[5] R. Caldelli, R. Becarelli, and I. Amerini, "Image origin classification based on social network provenance," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 6, pp. 1299–1308, 2017.

[6] D. Moreira, A. Bharati, J. Brogan, A. Pinto, M. Parowski, K. W. Bowyer, P. J. Flynn, A. Rocha, and W. J. Scheirer, "Image provenance analysis at scale," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 6109–6123, 2018.

[7] S. F. C. Haviana, D. Kurniadi *et al.*, "Average hashing for perceptual image similarity in mobile phone application," *Journal of Telematics and Informatics*, vol. 4, no. 1, pp. 12–18, 2016.

[8] V. Monga and B. L. Evans, "Perceptual image hashing via feature points: performance evaluation and tradeoffs," *IEEE transactions on Image Processing*, vol. 15, no. 11, pp. 3452–3465, 2006.

[9] S. S. Kozat, R. Venkatesan, and M. K. Mihçak, "Robust perceptual image hashing via matrix invariants," in *2004 International Conference on Image Processing, 2004. ICIP'04.*, vol. 5. IEEE, 2004, pp. 3443–3446.

[10] R. Venkatesan, S.-M. Koon, M. H. Jakubowski, and P. Moulin, "Robust image hashing," in *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)*, vol. 3. IEEE, 2000, pp. 664–666.

[11] M. Fei, Z. Ju, X. Zhen, and J. Li, "Real-time visual tracking based on improved perceptual hashing," *Multimedia Tools and Applications*, vol. 76, pp. 4617–4634, 2017.

[12] H. Liu, R. Wang, S. Shan, and X. Chen, "Deep supervised hashing for fast image retrieval," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2064–2072.

[13] H. Zhu, M. Long, J. Wang, and Y. Cao, "Deep hashing network for efficient similarity retrieval," in *Proceedings of the AAAI conference on Artificial Intelligence*, vol. 30, no. 1, 2016.

[14] R. Xia, Y. Pan, H. Lai, C. Liu, and S. Yan, "Supervised hashing for image retrieval via image representation learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 28, no. 1, 2014.

[15] Y. Wu, W. AbdAlmageed, and P. Natarajan, "Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9543–9552.

[16] X. Hu, Z. Zhang, Z. Jiang, S. Chaudhuri, Z. Yang, and R. Nevatia, "Span: Spatial pyramid attention network for image manipulation localization," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*. Springer, 2020, pp. 312–328.

[17] X. Chen, C. Dong, J. Ji, J. Cao, and X. Li, "Image manipulation detection by multi-view multi-scale supervision," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 14 185–14 193.

[18] C. Dong, X. Chen, R. Hu, J. Cao, and X. Li, "Mvss-net: Multi-view multi-scale supervised networks for image manipulation detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[19] X. Lin, S. Wang, J. Deng, Y. Fu, X. Bai, X. Chen, X. Qu, and W. Tang, "Image manipulation detection by multiple tampering traces and edge artifact enhancement," *Pattern Recognition*, vol. 133, p. 109026, 2023.

[20] "JPEG systems — Part 5: JPEG universal metadata box format (JUMBF)," International Standard ISO/IEC 19566-5:2019, 2019.

[21] "Use Cases and Requirements for JPEG Fake Media," Techincal Report ISO/IEC JTC 1/SC29/WG1N100156, 2022.

[22] "Final Call for Proposals for JPEG Fake Media," Techincal Report ISO/IEC JTC 1/SC29/WG1N100157, 2022.

[23] C. Cachin *et al.*, "Architecture of the hyperledger blockchain fabric," in *Workshop on distributed cryptocurrencies and consensus ledgers*, vol. 310, no. 4. Chicago, IL, 2016, pp. 1–4.

[24] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich *et al.*, "Hyperledger fabric: a distributed operating system for permissioned blockchains," in *Proceedings of the thirteenth EuroSys conference*, 2018, pp. 1–15.

[25] D. Bhowmik and T. Feng, "The multimedia blockchain: A distributed and tamper-proof media transaction framework," in *2017 22nd International conference on digital signal processing (DSP)*. IEEE, 2017, pp. 1–5.

[26] "Inter Planetary File System (IPFS)," https://ipfs.tech/, accessed: 2023-03-17.

[27] A. Adler, "Fair use and the future of art," *NYUL Rev.*, vol. 91, p. 559, 2016.

[28] "Kodak lossless true color image suite," http://r0k.us/graphics/kodak/, accessed: 2023-03-17.

[29] J. M. Bonner, "Let them authenticate: Deterring art fraud," *UCLA Ent. L. Rev.*, vol. 24, p. 19, 2017.

[30] D. Becker, A. Fischer, and Y. Schmitz, *Faking, forging, counterfeiting: discredited practices at the margins of mimesis*. transcript Verlag, 2018.

[31] "Fragonard's young woman revealed as replica in made in china project — dulwich picture gallery," https://www.dulwichpicturegallery.org.uk/about/press-media/press-releases/fragonards-young-woman-revealed-as-replica-in-made-in-china-project/, accessed: 2023-03-17.