Exploring NFT Validation through Digital Watermarking

Mila Dalla Preda, Masaia Francesco

ABSTRACT

Blockchain technology has brought notable advancements to diverse industries. The introduction of non-fungible tokens (NFTs) has particularly led to a lucrative market for unique digital asset ownership verification, including digital artworks. However, this trend has also given rise to concerns such as fraud, stolen works, authenticity, and copyright issues. Illicit traders exploit the market by trading unauthorized copies of digital objects as NFTs. In this study, we propose the use of digital watermarking as a means to establish the authenticity of NFTs and enhance the marketplace's credibility.

CCS CONCEPTS

• **Security and privacy → Distributed systems security**; **Digital rights management**; **Authorization**.

KEYWORDS

Digital Watermarking, non-fungible token, blockchain

ACM Reference Format:

Mila Dalla Preda, Masaia Francesco. 2023. Exploring NFT Validation through Digital Watermarking. In *The 18th International Conference on Availability, Reliability and Security (ARES 2023), August 29-September 1, 2023, Benevento, Italy*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3600160.3605063

1 INTRODUCTION

Blockchain technology is a nascent innovation technology that has been embraced by numerous applications. In finance, it has transformed payment processing, banking, and trading. It has also streamlined processes in insurance such as underwriting and claims management. In the healthcare industry, it ensures the secure handling of medical data and provides transparency in supply chain management to minimize the risk of counterfeit products. Voting systems can also benefit from blockchain technology by providing tamper-proof and highly secure voting records.

Additionally, the art industry is striving to integrate blockchain technology due to its vital role in ensuring the integrity of artwork information and its authenticity, both of which significantly influence the artwork's value. The blockchain effectively seems to support these aspects. Following the advent of non-fungible tokens (NFT) and accompanied by platforms facilitating the buying and selling of images, sequences of images, videos, and animated gifs, the realm of digital art has discovered a novel avenue for trading and promoting their artistic creations. The key breakthrough brought about by NFTs is the ability to verify ownership of digital artworks once the asset itself or a URL pointing to the asset is minted into a blockchain. An NFT is a unique digital asset minted in a blockchain that cannot be replicated or duplicated. It is created using a smart contract that takes as input a digital object (like images, videos, audio, software and more), some meta information regarding for example the ownership of the digital object, and creates the unique digital token. The value of an NFT is determined by its uniqueness, scarcity, and market demand.

The NFT market is experiencing a rapid expansion due to the groundbreaking nature of NFTs, which establish property rights in the digital realm for the first time. This distinctiveness has been a driving force behind the increasing value of NFTs. Data from DappRadar indicates that the total volume of NFT sales in the previous year came close to reaching the peak observed in 2021, approximately 24.7 billion dollars in organic trading volume throughout 2022 across various blockchain platforms and marketplaces and 25.1 billion dollars recorded in 2021 [2]. Moreover, the NFT spend value is expected to increase in the following years [11].

We have seen that the advent of NFTs has created a highly profitable market for digital artworks. However, along with this trend, concerns have emerged regarding fraud, stolen works, authenticity, and copyright issues. Unscrupulous individuals engage in the illicit trade of unauthorized copies of digital art. This unfortunate practice is made possible by the need for artistic works to be showcased in markets, where they can be easily duplicated and resold as separate NFTs. Indeed, no identity verification is required for minting an NFT on a blockchain. This means, that anyone that holds a (copy of) digital object can successfully mint an NFT declaring the ownership of the digital object. One potential solution to these concerns is the implementation of identity verification for NFT issuers. Some marketplaces have already established a blue checkmark system for verified collections and artists, which can assure consumers that they are purchasing authentic works. For instance, OpenSea [4], one of the largest NFT marketplaces, offers a blue checkmark for experts and verified collections, seeking to minimize the risk of fraud or theft. Another platform, SuperRare, also provides verified artist badges to ensure that collectors know they are investing in an original piece of work. However, the main limitation of these existing methods for identity verification is that they are performed manually by getting personal information about the creator.

As depicted in Fig.1 the illegal traders simply copy the advertised digital art available in the market and mint it as their own digital work, utilizing the same NFT technology and various blockchains, storage servers, and displaying services readily accessible [7]. This form of illegal trading poses a particular threat to inexpensive artistic works, as pursuing legal action against illicit traders may not be financially viable. The costs and challenges associated with proving ownership in a court of law are often prohibitive. Moreover, many artists who have not even created an NFT for their work fall

ARES 2023, August 29-September 1, 2023, Benevento, Italy

ACM ISBN 979-8-4007-0772-8/23/08…$15.00
https://doi.org/10.1145/3600160.3605063

victim to theft, as unscrupulous traders generate the NFT before the rightful owners have the chance to do so.

In the following, we outline the typical procedure of minting NFTs on a blockchain and put forth a more robust certification protocol that a marketplace could adopt to guarantee the authenticity of NFTs. The idea behind the proposed protocol is to combine NFT and digital watermarking [6, 9, 10, 12]. This protocol utilizes the familiar NFT minting process provided by existing marketplaces and ensures the verification of the digital object's authorship by embedding a watermark within the digital object, which can only be authenticated by the marketplace. The effectiveness of the proposed procedure is strictly related to the effectiveness of the employed watermarking algorithm. The proposed solution can allow the detection of successive NFT submissions on the same digital object. This means that we are protecting the first user that has minted the NFT.

2 DIGITAL WATERMARKING

Digital watermarking has been introduced in [12] to verify the authenticity or integrity of an image and it is then been extended to other digital objects such as videos, audio, and software [6, 9, 10]. A digital watermark refers to a concealed marker that is discreetly embedded within a signal that can tolerate noise, such as audio, video, image data, or software. Its primary purpose is to establish ownership of the copyright associated with the signal. The watermarking process entails concealing a signature, related to ownership, within a digital object, while ensuring that the functionality and usage of the digital object remain unaffected. Watermarking schemes typically involve using a secret key that enables only authorized individuals to extract the signature and provide proof of ownership. The goodness of a watermarking scheme is measured in terms of: *stealthiness* considering how much the inserted signature is transparent to the user experience, *resilience* against attacks that try to disrupt or compromise the inserted signature, *bit-rate* referring to the amount of information that can be embedded in the digital object and *credibility* measuring the probability of extracting a signature form a digital object that has not been watermarked (should be low). Watermarking schemes are widely employed for tracking copyright infringements of images and software (e.g., [6, 9, 10]). The watermarking schemes consist of two main algorithms, as depicted in Figure 2:

- An embedding process that conceals the signature in the digital object by using a secret key $K_{watermark}$. Thus, the watermarked object obj_W is computed as
$$obj_W = Embed(K_{watermark}, signature, obj)$$
- An extraction process that gives the watermarked digital object obj_W and the secret key $K_{watermark}$ recovers the inserted signature. Namely:
$$signature = Extract(K_{watermark}, obj_W)$$

3 BLOCKCHAIN AND NFT

Blockchain technology is a groundbreaking and transformative technology that has disrupted various industries, including the art one. It is a decentralized, distributed ledger that records transactions transparently and securely [3]. Its immutability feature ensures that once a transaction is recorded on the blockchain network, it cannot be altered or deleted. One of the innovative digital assets created on the blockchain is the NFT. NFTs are unique digital assets that cannot be replicated or duplicated. They are created using smart contracts, which automate contract execution and eliminate intermediaries. NFTs are minted by registering the unique digital asset on the blockchain network.

NFTs can contain a wide range of digital assets such as images, videos, audio, software, and more. The value of an NFT is determined by its uniqueness, scarcity, and market demand. The minting process involves creating a unique digital asset and registering it on the blockchain network using a digital wallet. The use of InterPlanetary File System (IPFS) is often employed to store assets related to NFTs [1]. IPFS is a distributed file system that provides a decentralized, global, and versioned filesystem that makes the storage of large amounts of data possible.

An NFT marketplace is a platform that facilitates the buying and selling of NFTs. These marketplaces are often decentralized and allow for peer-to-peer transactions without intermediaries. OpenSea is a well-known decentralized NFT marketplace that offers a broad range of NFTs such as digital art, collectibles, and virtual real estate. It provides a secure and transparent platform for buying and selling NFTs and makes the minting process easy. Additionally, some NFT marketplaces have implemented a verification system to ensure the authenticity of the NFTs being sold. This system involves a blue checkmark as shown in Figure 3, a feature that appears next to verified collections and artists, providing buyers with more confidence in their purchases. These marketplaces have improved the overall NFT user experience and increased adoption by making the process of buying and selling NFTs more streamlined and trustworthy. As the NFT market continues to expand, we can expect to see more innovative features added to NFT marketplaces that strengthen the ecosystem and attract more users.

Figure 4 gives an example of workflow to generate a new NFT starting from a digital object and some custom metadata.

4 PROTOCOL OVERVIEW

We assume that a pair of asymmetric keys are assigned to the user when she/he registers on the marketplace. Thus, the marketplace manages the generation of the keys. When the user authenticates to the marketplace, the marketplace verifies whether the pair of public and secret keys of the user is suspended. If it is found to be suspended, an error is reported, and the marketplace proceeds by verifying the identity of the user and generating a new pair of keys. In the event of compromised keys, any NFTs minted prior to the specified revocation date by the legitimate user remain valid, while those minted thereafter are no longer eligible for sale.

Recall that we have:

- A pair of public and private keys (K_{public}, K_{secret}) generated by the marketplace for each user that is registered.
- A secret key $K_{watermark}$ that is the secret key required by the watermarking scheme for embedding/extracting the signature from the digital object. This key is known only to the marketplace that uses the same $K_{watermark}$ for all the digital

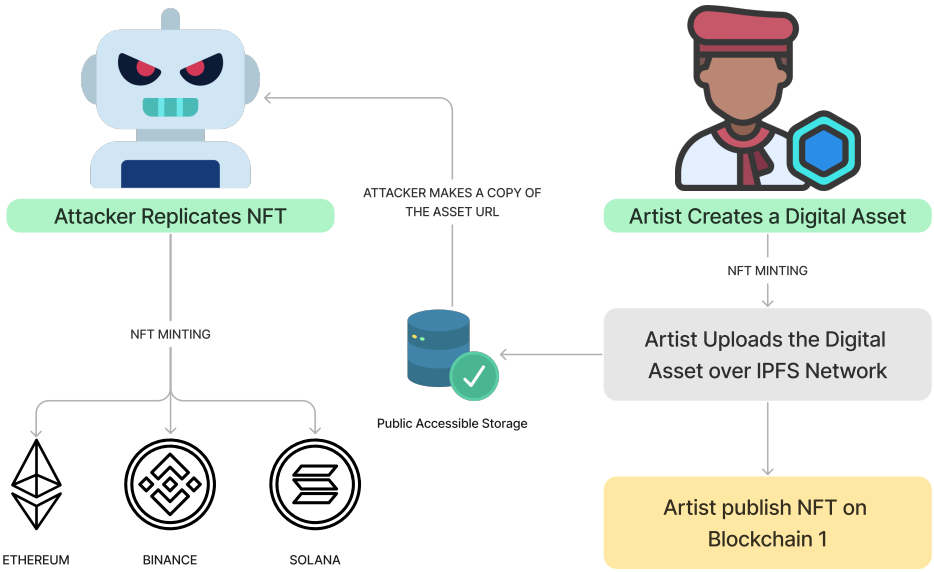

Figure 1: Illicit NFT minting

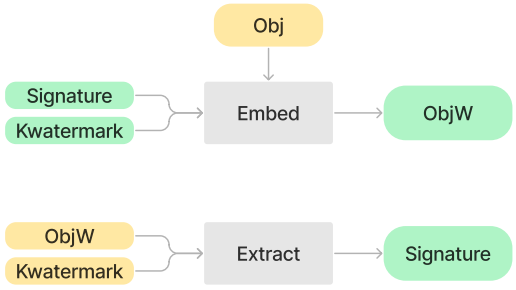

Figure 2: Watermarking scheme

Figure 3: Blue check mark verification for an NFT collection on OpenSea

objects uploaded by a given user. Thus, the marketplace associates a secret key $K_{watermark}$ for each user and stores this information in a secure way.

In Figure 5, we describe the workflow of the proposed protocol.

- The user authenticates to the marketplace that verifies whether the public key of the user is suspended. If it is found to be suspended, an error is reported and managed by the marketplace.
- If the key is valid, the user proceeds by uploading the specific digital object *obj* that she/he wishes to mint as an NFT onto the blockchain. In the following K_{secret} and K_{public} refer to the asymmetric keys of the current user.
- The marketplace recovers the secret key $K_{watermark}$ associated with the current user and attempts to extract a signature from the uploaded digital object.
- If the digital object does not contain a signature, namely if the algorithm $Extract(K_{watermark}, obj)$ does not return a signature, the marketplace embeds the public key K_{public} as

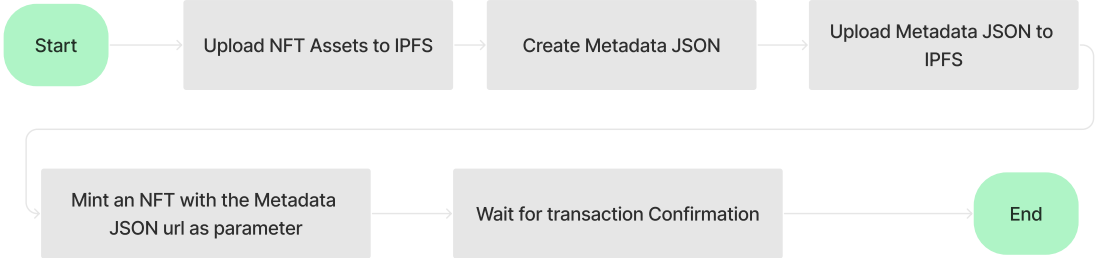

Figure 4: NFT minting process

a signature in the digital object. Thus computing the watermarked digital object

$$obj_W = Embed\{K_{watermark}, K_{public}, obj\}$$

obj_W is uploaded on a public IPFS. An NFT containing the `url` of the watermarked digital object on the IPFS, and the desired metadata containing the ownership information is minted to the blockchain.

- If the uploaded digital object *obj* already contains a signature, namely if $Extract(K_{watermark}, obj)$ returns a signature, it means that an NFT of this digital object has already been minted in the blockchain by this marketplace.
- If the extracted signature does not correspond to the public key of the current user, namely if $Extract(K_{watermark}, obj) \neq K_{public}$, it means that the digital object was minted by someone else and this user has copied, and potentially modified, the digital object form the IPFS and she/he is trying to mint it. An alarm is raised and the user is labeled as untrusted.
- If $Extract(K_{watermark}, obj) = K_{public}$, then the marketplace has to verify that the current user is indeed the one that first minted the image. A challenge is issued to verify that the user possesses the corresponding private key K_{secret}. If the challenge succeeds a warning is returned to the user saying that she/he has already minted the NFT of the uploaded digital object. If the challenge does not succeed it means that the current user is trying to impersonate the user that has originally minted the NFT by exploiting her/his public key. In this case, an alarm is raised.

5 EVALUATION AND POSSIBLE APPLICATIONS

The proposed protocol, when adopted by the marketplace, can allow the detection of illicit NFT minting. Illicit here means that there has been a previous user that has declared ownership of the digital object by minting an NFT on the blockchain. Future minting submissions are then classified as illicit. This means we are protecting the first user who minted the NFT.

The proposed protocol does not support the verification of the authenticity of the first user submitting to the marketplace a request for NFT minting on the blockchain. This first identity verification is left to the marketplace and it is difficult to automate it could be achieved by Know Your Customer (KYC), public digital identity systems, and so on.

It is worth noting that existing marketplaces typically do not verify the authenticity of both the initial user and subsequent users who submit requests to mint an NFT for a digital object. While the proposed solution provides authenticity checks for successive NFT submissions. The proposed solution mandates the marketplace to authenticate the user's identity only in the absence of previously minted NFTs for the particular digital object. Once the user clears the authentication process, an NFT is created, and any subsequent attempts to mint NFTs for the same digital object or its illicit copies are automatically rejected by the marketplace. Notably, the watermarking scheme selected demonstrates the marketplace's ability to block all types of illicit copies, even those created through transformations that maintain the inserted signature's integrity. However, it must be stressed that if the digital objects stored in IPFS are not watermarked, then the marketplace will prevent the minting of NFTs for identical digital objects. Devoid of any watermarking technique, the marketplace's ability to do so would be easily bypassed by tweaking a single bit of information. As a result, IPFS recognizes the file as entirely new, and the marketplace fails to locate it in its search for previously minted NFTs. Through the use of a watermarking scheme, the proposed solution gains resilience against various transformations that the watermarking scheme can withstand, effectively identifying instances of plagiarism or illegal copies of any objects derived from the original ones, provided the signature inserted by the watermarking technique remains intact. The importance of the resilience of the watermarking technique cannot be overstated, as malicious users may attempt to remove or diminish the embedded watermark once they copy the digital image. Therefore, ensuring the robustness of the proposed protocol is heavily reliant on the resilience of the chosen watermarking protocol and the sharing of key pairs between marketplaces. Given the right watermarking protocol and key pair sharing, this standard could greatly simplify the publication process of new NFTs for creators who need to ensure the protection of their content against such attacks.

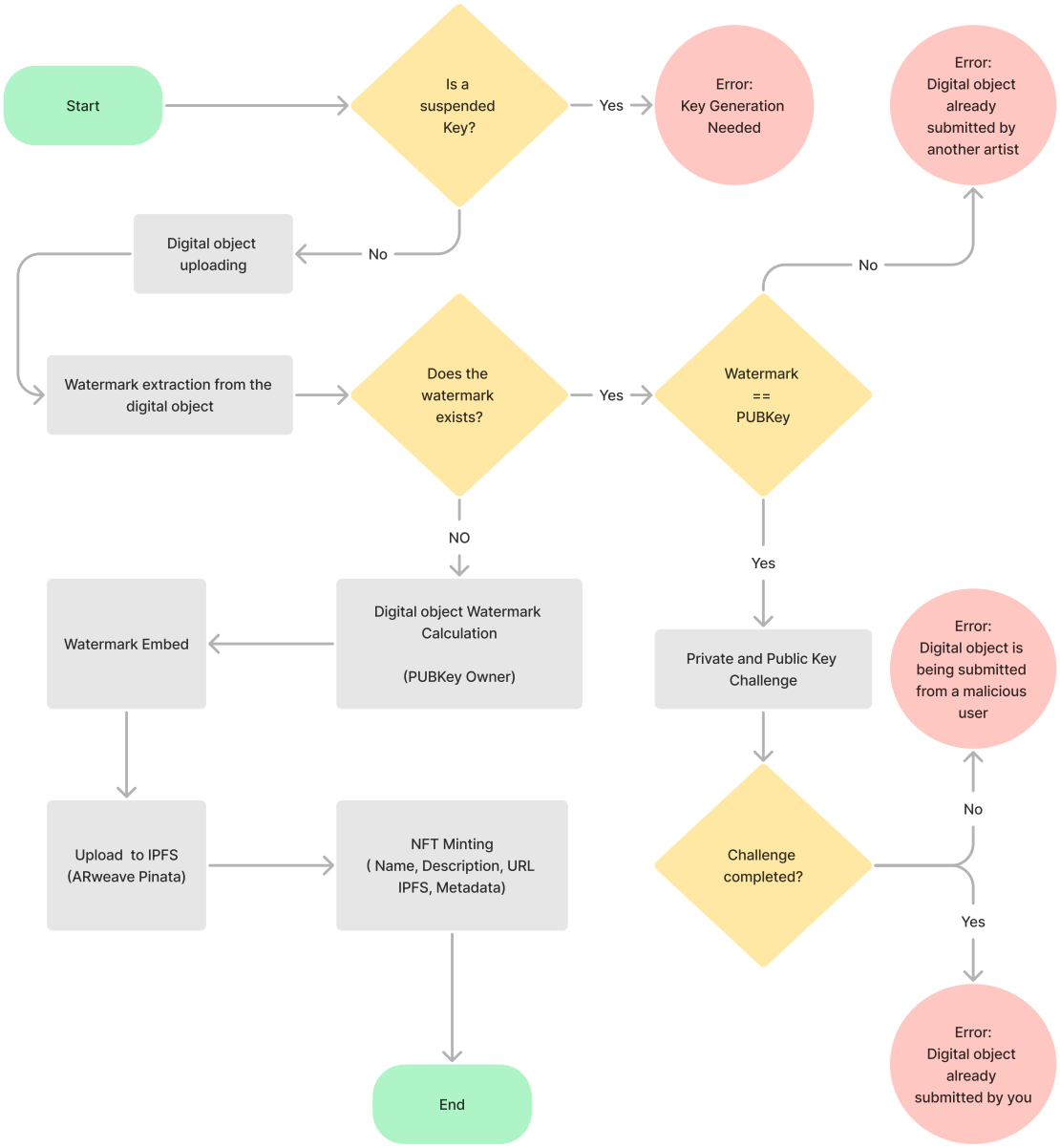

Figure 5: Workflow for platform watermark process

When evaluating the expenses associated with implementing the recommended application, it is crucial to acknowledge that the creation of NFTs will remain unaffected. The main expenditure arises during the verification phase when the marketplace must authenticate the user's NFT request. Nonetheless, the idea's significant advantage for marketplaces lies in its potential to enhance credibility and reliability, which can lead to a surge in platform users. Moreover, the expenses related to verification procedures may escalate if done manually, making the automated route proposed in the solution a more viable option where human intervention is kept to a bare minimum.

The concise and practical framework we have developed is designed to secure the integrity and authenticity of any digital object that can be subject to tampering or forgery. The context of digital art trading served as a motivating example to explore the implications of ensuring the originality of digital artifacts, but the proposed framework is portable and adaptable to any watermarking scheme applied to other digital objects such as videos, music, software, and legal documents. Digital assets are susceptible to infringement, manipulation, and unauthorized distribution, requiring rigorous protection measures tailored to specific use cases. Our framework offers a straightforward and scalable security solution for various industries, including digital art, e-commerce, intellectual property rights, electronic voting, digital identity, and supply chain management. By using robust blockchain and smart contract technologies, entities can ensure that their digital assets remain immutable and secure from any malicious attack or unauthorized access.

6 RELATED WORKS

In [7] the author presents an initial exploration of merging watermarking and NFTs using a specific image watermarking method known as the least significant bit (LSB) technique and requiring two pairs of asymmetric keys. In contrast, our proposed framework offers a more comprehensive approach wherein the marketplace can verify the authenticity of users who submit NFT minting requests for any kind of digital object. We utilize a single pair of asymmetric keys, manage the secret key for watermark extraction, and consider asymmetric key revocation.

In [5] the authors propose to verify the authenticity of off-chain NFT data using image fingerprints checked by NFT managers. The correct use of images is encouraged by economic rewards. This does not apply to our scenario since we store data in IPFS which ensure the integrity of data.

Recent relevant work is [8] where the authors propose a solution for verifying the integrity of 3d models and detecting illegal duplicates. To this end, they extract a fingerprint from 3d models by using the Fourier Fingerprint Search (FFS). This fingerprint is then inserted in the NFT of the model. The correspondence between the fingerprint and the 3d model provides proof of integrity/ownership. FFS is resilient to light modification of the model, thus enabling the legitimate proprietary to prove his ownership of slightly modified models. The main difference with our approach is that in this solution the digital object is not modified in order to contain ownership information, this information is stored directly in the NFT and not in the asset.

7 CONCLUSION

Besides blue check mark verification, adopted by the mainstream NFT Marketplace, online fraud remains a risk within the NFT market due to issues like fake NFT giveaways, copyright theft, fake stores, and NFT artist impersonation. This paper presents a general framework where digital watermarking solutions can be combined with NFT minting for mitigating illegal trading. It is important to note that while digital watermarking combined with NFT minting can help mitigate illegal trading in the NFT market, our proposed protocol does not support the verification of the authenticity of the first user submitting a request for NFT minting. This crucial identity verification is a responsibility that falls upon the marketplace and can be achieved through various means such as KYC or public digital identity systems. We believe that the adoption of the proposed solution could limit the amount of illicit trading.

REFERENCES

[1] [n. d.]. Dupres, C. (2022) IPFS, Filecoin and the Long-Term Risks of Storing NFTs. https://www.coindesk.com/layer2/2022/01/20/ipfs-filecoin-and-the-long-term-risks-of-storing-nfts/. Accessed: 24 May 2023.

[2] [n. d.]. NFT Sales in 2022 Nearly Matched the 2021 Boom, Despite Market Crash. https://decrypt.co/118438/2022-versus-2021-nft-sales. Accessed: 2023-05-14.

[3] [n. d.]. Plabon, H.R. (2023) Unlocking the Power of Blockchain: A Beginner's Guide to Decentralized Digital Ledgers, Tech Plabon. https://www.techplabon.com/2023/02/unlocking-power-of-blockchain-beginners.html. Accessed: 24 May 2023.

[4] [n. d.]. What is a verified account or badged collection? https://support.opensea.io/hc/en-us/articles/360063519133-What-is-a-verified-account-or-badged-collection- Accessed: 2023-05-24.

[5] Yulong Chen, Ziwei Wang, Xiangyu Liu, and Xuetao Wei. 2022. A New NFT Model to Enhance Copyright Traceability of the Off-chain Data. In *2022 International Conference on Culture-Oriented Science and Technology (CoST)*. 157–162. https://doi.org/10.1109/CoST57098.2022.00041

[6] C. Collberg and J. Nagra. 2009. *Surreptitious Software: Obfuscation, Watermarking, and Tamperproofing for Software Protection.* Addison-Wesley Professional.

[7] Joceli Mayer. 2022. Review on Watermarking Techniques Aiming Authentication of Digital Image Artistic Works Minted as NFTs into Blockchains. In *Information Security and Privacy in the Digital World - Some Selected Topics*, Prof. Jaydip Sen and Prof. Joceli Mayer (Eds.). IntechOpen, Rijeka, Chapter 1.

[8] Dimitris Mouris and Nektarios Georgios Tsoutsos. 2022. NFTs For 3D Models: Sustaining Ownership In Industry 4.0. *IEEE Consumer Electronics Magazine* (2022), 1–1. https://doi.org/10.1109/MCE.2022.3164221

[9] Mohammad Ali Nematollahi, Chalee Vorakulpipat, and Hamurabi Gamboa Rosales. 2017. *Digital watermarking.* Springer.

[10] Prabhishek Singh and Ramneet Singh Chadha. 2013. A survey of digital watermarking techniques, applications and attacks. *International Journal of Engineering and Innovative Technology (IJEIT)* 2, 9 (2013), 165–175.

[11] TechInsight360. [n. d.]. Global NFT market intelligence and future growth dynamics databook - 50+ kips on NFT investments by key assets, currency, sales channels - Q2 2022.

[12] Anatol Z Tirkel, GA Rankin, RM Van Schyndel, WJ Ho, NRA Mee, and Charles F Osborne. 1993. Electronic watermark. *Digital Image Computing, Technology and Applications (DICTA'93)* (1993), 666–673.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009