Chapter

Review on Watermarking Techniques Aiming Authentication of Digital Image Artistic Works Minted as NFTs into Blockchains

Joceli Mayer

Abstract

The recent creation of Non Fungible Tokens (NFTs) has enabled a multibillionaire market for digital artistic works including images or sequence of images, videos, and animated gifs. With this new trend issues regarding fraud, stolen works, authenticity, and copyright came along. The goal of this chapter is to provide an overview of the watermarking techniques that can be employed to mitigate those issues. We will discuss transparency, robustness, and payload of watermarking techniques aiming to educate the artists, researchers, and developers about the many approaches that watermarking techniques provide and the resulting trade-offs. We focus on fragile watermarking techniques due to their high transparency for embedding into artistic works. We discuss the spread spectrum and Least Significant Bit techniques. We describe the usual process of NFT minting into a blockchain and propose a more secure certification protocol with watermarking which employs the same usual NFT minting offered by current marketplaces. The proposed certification protocol mints a checksum string into a blockchain, ensuring the validity of the watermark and the information embedded into this watermark. This proposed protocol validates the date of creation and author identification which are transparently embedded in the artistic work, thus, increasing the security and confidence of markets for artistic works transactions.

Keywords: image watermarking, non fungible tokens, blockchain, reversible watermarking, visible watermarking, transparent watermarking

1. Introduction

The world of digital art has found an innovative way to trade and/or advertise their artistic image works after the recent creation of Non Fungible Tokens (NFTs) associated with a blockchain and some service to sell and buy the images or sequence of images, videos, and animated gifs. The main innovation conferred by NFTs is that the ownership of the digital artistic work is verifiable after the digital asset or a link to the asset with a URL (Universal Resource Locator) is minted into a blockchain.

After the first NFT work was created in 2014, named “quantum”, a multibillionaire business has grown around NFTs and blockchains. The market cap of trading NFTs totaled over 23 billion dollars last year. Along with this surge of lucrative trading digital art through NFT, markets came also another black market with players that trade unauthorized copies of digital art disposed of at the markets. As a result, many artists started to include visible and invisible watermarks in their works in the hope that it would prevent stealing or provide additional legal evidence about the authorship to be disputed in a court of law. Moreover, protocols including the watermarked NFTs and the embedded data in the watermarks are being designed to provide the buyers some extra confidence that the work is actually original and created or owned by the seller, avoiding or mitigating a problem created in the market of NFTs: unauthorized copies sold to unwise buyers.

A complication issue is that the artistic work needs to be shown in the markets and is easily copied and re-sold as another NFT. The illegal trader just copies the advertised digital art in the market and mints the digital work (or minting its URL as is the usual practice) as his or her own using the same NFT technology and one of the many available blockchains and storage/displaying servers and services. This illegal trading is particularly damaging to low-cost artistic works which would not be worthwhile to start a legal prosecution against the illegal trader. The costs and difficulties to prove ownership in a court of law are prohibitive. Moreover, many artists that do not even create an NFT for their work are being stolen as illegal traders create the NFT before the actual owner.

The goal of this chapter is to provide an overview of the watermarking techniques that can be employed to mitigate the problem of authentication in this multibillionaire market of NFTs. We will discuss transparency, robustness, and payload of watermarking techniques divided into three categories: transparent with low impact in the artistic work, very robust with high impact in the artistic work and transparent and reversible watermarks. The discussion aims to educate the artists, researchers, and developers about the many approaches that watermarking techniques provide and the trade-offs that each watermarking technique imposes. As the technology of digital art trading evolves, these watermarking technologies and trading protocols will take place to provide a safer and more lucrative environment to the sellers and buyers in this innovative market.

2. Security of authorship for minted NFTs

The process of registering digital data (coin, NFT, image, video, etc) into a blockchain, due to a somewhat complex and secure cryptographic protocol employed, provides a very high probability that the digital data can be securely assigned to a owner along with some extra data such as a URL, date and other information about the transaction. The process of registering the data into a blockchain is named minting, due to its similarity to printing (minting) fiduciary money. This process is considered very secure, publicly accessed, and it is verifiable in a noncentralized way by many participants in the process. The decentralized finance (DeFi) approach is based on blockchain to assure proper secure transactions (digital coins or smart contracts) without the need for a unique institutional agent, such as a bank or government [1].

2.1 Minting Process for NFTs

Regardless of the blockchain chosen for minting, there is a cost associated with the computing energy spent to process and validate the transactions in the blockchain, usually referred to as “gas” fees. For this reason, the minting of NFTs usually requires an associated storage server to upload the actual image, video, or animated GIF image. Otherwise, the “gas” fees become prohibitive high due to a large amount of data (bytes) required to be verified by the computing servers. Therefore, due to registration costs, in practice, only some data related to the NFT (URL, author, date, or some small information) is actually minted into the blockchain. This raises some issues regarding the security of the digital asset since it is stored in external servers and not registered into the blockchain. Currently, some services are provided for that external storage, however, they do not use blockchain technology and the security is left to the service provider's considerations. Recently, it has been reported that US$ 1,7 million of NFTs was stolen by a hacker from a very popular NFT service name OpenSea.

Therefore, additional technologies need to be provided to digital art creators in order to enable more confidence in the transactions. Besides cryptographic protocols, watermarking techniques are being employed by artists aiming protection for copyright, authentication, and mitigation of frauds in the NFT market.

3. Watermarking techniques applied to NFTs

There are a variety of watermarking techniques such as visible, fragile, semi-fragile, strong, and reversible watermark. These techniques may be used to achieve different goals of authentication, copyright protection, tracking, or fraud detection. Some techniques properties, namely, robustness, transparency, and payload are required depending on the desired goal.

3.1 Watermarking properties and tradeoffs

3.1.1 Robustness

Robustness is a desirable property for a technique in the sense that the watermark, which may contain copyright or authentication information, is able to survive a given attack. There are two types of attacks: malicious and nonmalicious attacks. Nonmalicious attacks refer to normal transformations that one digital work may suffer during transmission or processing as a change of image format, from a JPEG to PNG for instance, or a mild filtering or histogram equalization. On the other hand, malicious attacks are designed to either remove the watermark and/or to substitute it with another watermark for fraudulent purposes. Some malicious attacks may include geometric (shearing, horizontal flipping, collage) and volumetric transformations (noise addition, color map modification, filtering, JPEG compression) [2].

3.1.2 Transparency

Transparency is a very desired property in the context of artistic digital works. The watermarking technique should be as invisible as possible in order to not affect the image quality since the work is presented by a given site or application for potential buyers. However, many artists use available software to insert very visible watermarks over the original work. This approach intends to provide a sample for the digital work either for advertising the author's artistic qualities and/or for indicating that a watermark-free can be purchased after by contacting the author. The approach aims to mitigate possible stealing of the work and re-selling under other authors' names. As a result, the

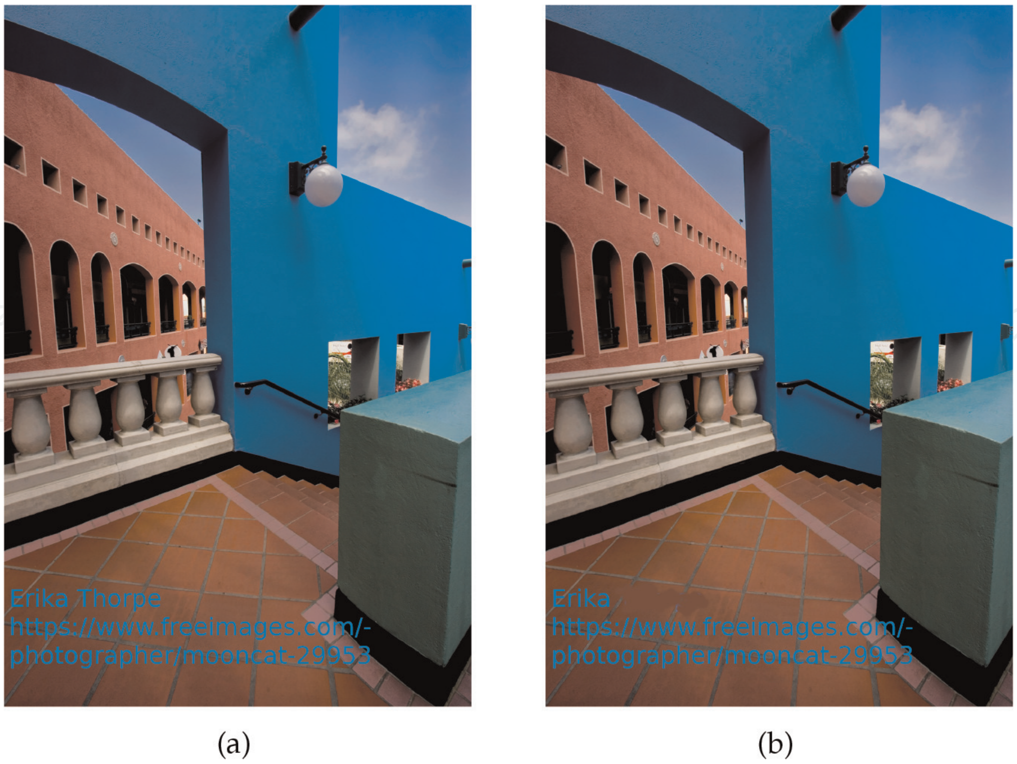
Erika Thorpe
https://www.freeimages.com/-
photographer/mooncat-29953
Erika
https://www.freeimages.com/-
photographer/mooncat-29953
(a)
(b)

Figure 1.
(a) A copyright-free image from [3] had the author's name and URL included at the bottom. (b) By using image tools, the surname of the author has been removed.

watermark location is visible, and usually damages the image quality and presentation to a certain degree and additionally, operations of collage with image processing can be used to remove the visible watermarks creating a watermark-free similar version in order to re-sell the stolen art in the same site or another similar service for NFT trading.

In order to illustrate the point, in **Figure 1a**, a copyright-free image from [3] had the author name (found in the metadata image information) and the original URL drawn over the image bottom, and in **Figure 1b**, using image tools, the surname of the author has been removed, indicating how easily visible watermarks can be tampered with.

Moreover, in this scenario, a buyer has no guarantee that the received digital work is indeed original as it was created by the seller or it is a stolen edited copy. Therefore, for this NFT trading scenario, very transparent watermarks can be employed to convey authentication information along with some certification protocol provided by a trusted organization. Although invisible watermarks are more complicated and there is no available standard protocol for the artists, the need for a more secure market has been recognized and some corporations are building a trusting environment and friendly applications using watermarking techniques and blockchains that enable certification for the authors along with their digital works.

3.1.3 Payload

Payload is the amount of information measured in bytes that a watermarking technique is able to carry into the artistic image work. The required amount depends on the security protocol used and, on the need to convey some particular information such as author ID, URL, date of minting, and so on. For each given watermarking technique there is a trade-off among robustness, transparency, and payload. A technique with high robustness usually provides a relatively low transparency and a small payload. Conversely, a very transparent technique usually has low robustness and low payload. However, low robustness might be desired in some authentication applications. In such applications, the goal is to keep the digital work authenticated only if it has not been tampered with, thus very transparent and low robustness are proper for the NFT scenario where scarcity and authenticity are essential.

4. Semi-fragile and reversible watermarks

Robust watermarking techniques usually produce very low transparency and as stated before, transparency is a very important property when dealing with artistic works, therefore robust watermarking may not be a good choice for NFT authentication. On the other hand, very transparent watermarking can be achieved with fragile, semi-fragile, and also reversible techniques. Among these techniques, some are based on the spacial domain approach using Spread Spectrum (SS) [4] or Least Significant Bits (LSB) techniques. Others techniques rely on the transform domain approach using either Discrete Cosine Transform (DCT) as it is a basis for JPEG compression or Discrete Wavelet Transform (DWT) [2].

The semi-fragile approach allows a small degree of distortion imposed by nonmalicious attach such as image format transcoding, i.e., converting the digital image work from JPEG format to PNG format. However, large distortions usually meant for fraudulent purposes such as image horizontal flipping will result in losing the watermark and the digital work will not be authenticated anymore. In the next section, we describe an authentication protocol aiming to help to provide a more secure market for NFTs.

Reversible watermarking is designed to be able to remove the watermark with a proper secret key in order to restore the original artistic work. This approach is quite interesting for the NFT scenario where image quality is highly desirable. We describe how this feature can be achieved in the next sections.

4.1 Spatial domain techniques

Spread spectrum and LSB-based techniques are widely used and are able to provide a transparent watermarking for authentication purposes. These techniques can be designed as region based in order to locate which regions have been tampered with.

4.1.1 Spread spectrum techniques

The spread spectrum approach [5] is an additive operation on the spatial domain resulting in the watermarked image:

$$Im_W = Im + \alpha b W, \tag{1}$$

where Im is the original image (or frame of a video), α is the scaling parameter designed according to a desired robustness and transparency, b is an antipodal bit $\in \{-1, +1\}$ and W is a watermarking image of same size as Im. Usually, this watermarking image is built as white noise, generating a signal with a large spread spectrum in the frequency domain. The antipodal bit b is used to convey one bit of information along with the watermark signal authentication, in some cases it can be discarded, remaining only the weighted watermark signal, i.e., $Im_W = Im + \alpha W$. In

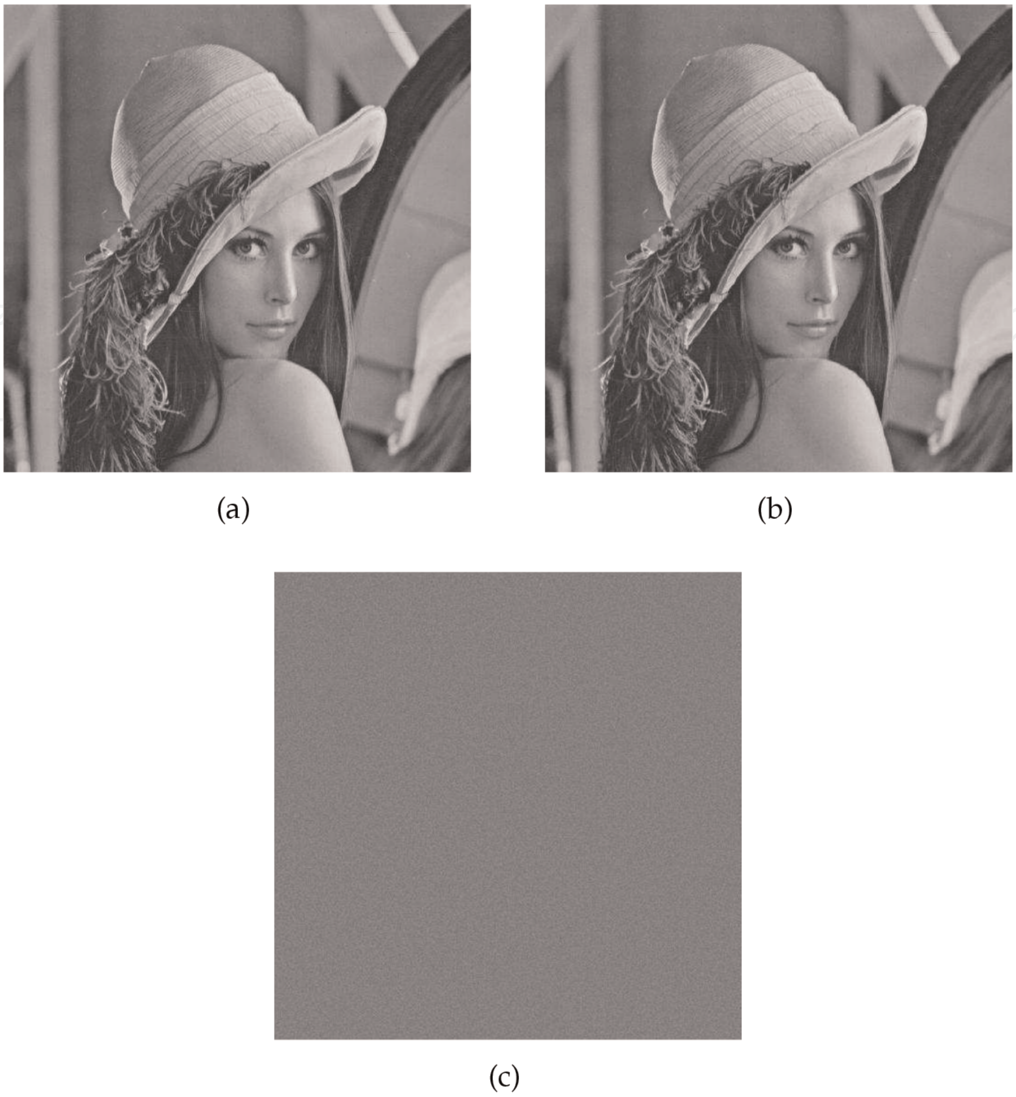

Figure 2.
(a) Original Lenna Image. (b) Watermarked Lenna Image embedded with 10 bits presents very low perceptual impact using the multibit technique in [4]. (c) Difference among the images scaled for visibility.

other cases, more bits can be inserted with a more complex watermark signal composed of a weighted sum of pseudorandom sequences of the same dimension as the original image. These pseudorandom sequences can be further optimized for their orthogonality [6]. In either case, the weight α can be computed to satisfy a given tradeoff between robustness and transparency as defined in [4]. **Figure 2** illustrates the very high transparency achieved using an elaborated multibit spread spectrum technique.

4.1.2 LSB techniques and reversible approach

The watermarking embedding is performed by changing the last K least significant bits of image pixels. The resulting impact for the last 2 bits, for instance, is usually very small and results in a very transparent embedding. Moreover, the approach of LSB embedding can also be reversible using a property of the binary operation XOR (exclusive OR). To understand how it works for the case of LSB embedding in the last bit of each pixel, assume a secret key, named Im_K as one 1-bit image of the same dimension as the original image, Im. Using an XOR ($\oplus$) operation for each last bit i, the embedding watermark is given as:

$$W(i) = Im_K(i) \oplus Im(i) \tag{2}$$

Next, the last bit of each pixel of the image, $Im(i)$, is replaced by the watermark $W(i)$, resulting in the watermarked image Im_W. The process allows the authentication of the digital image using a given protocol. Moreover, given the secret key Im_K, the last bits, changed previously, of the original image can be restored:

$$Im(i) = W(i) \oplus Im_K(i) \tag{3}$$

By replacing the last bit of each pixel of the watermarked image, $Im_W(i)$ by $Im(i)$, all bits of the original image are properly restored. This property can be used to improve the security of the NFT market. The LSB can be applied to more than the last bit, decreasing the transparency and increasing the payload. Notice that the LSB approach is a very fragile technique where any image modification will damage the watermark. This fragility is acceptable for authentication purposes within a certification protocol and services associated in order to improve the NFT market security and acceptance (**Figure 3**).

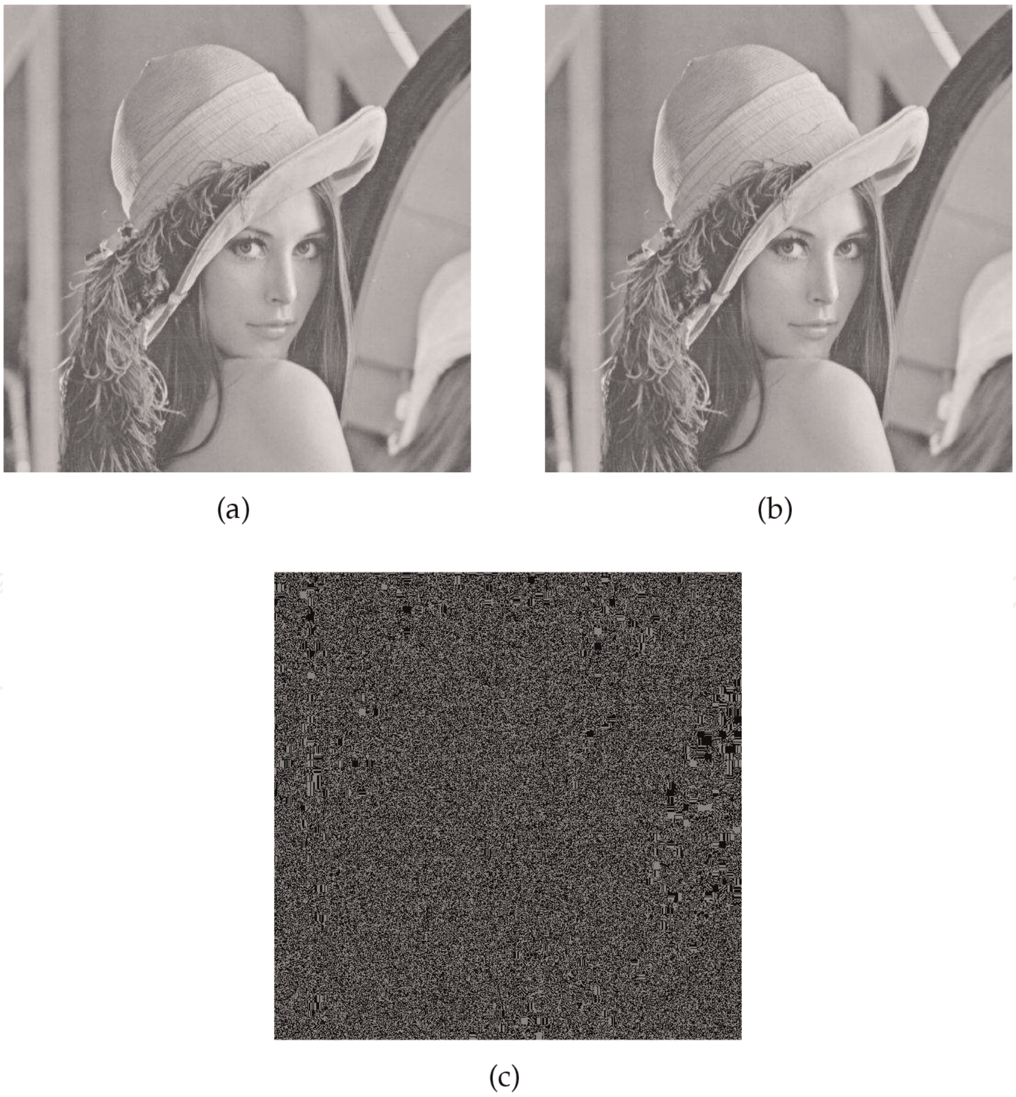

Figure 3.
(a) Original Lenna Image. (b) Watermarked Lenna Image embedded with 1 bits (all zeros in this example) presenting very low perceptual impact using the LSB. (c) Difference among the images scaled by 100 times for visibility.

4.2 Frequency domain techniques

In the frequency domain, it is possible to embed a waterkmark considering some model of human perception in relation to the frequency. In this way the technique can be properly adjusted to transparency according to such a human perception model and, as consequence, to reduce the visual impact of the embedding as compared to spatial domain techniques. The most used transforms for this purpose are the DCT and the DWT. There are a variety of approaches that modify some coefficients in the frequency domain in order to embed a watermark, a review on these advanced techniques is found in [2].

5. Certification protocol for watermarking NFTs

The process of minting an NFT into a blockchain is complex and a detailed example of the minting process for an NFT into the Ethereum blockchain is found in [7]. However, services provided by NFT marketplaces can mitigate the complexity for the users. Some of the top NFT marketplaces include OpenSea, Axie Marketplace, Larva Labs/CryptoPunks, NBA Top Shot Marketplace, Rarible, SuperRare, Foundation, Nifty Gateway, Mintable, and ThetaDrop. Using these services the process is simplified to a minimal number of steps which are explained in [8]. Moreover, the Rarible NFT marketplace offers a feature called “Lazy Minting”: all fees are charged to the buyer, only after buying the Work is actually minted, the seller receives the Work price amount minus the fees, including the minting “gas”. This feature is very interesting to incentive artists and creators [9].

As explained before due to the costs of mining, usually mentioned as “gas” fees, only the URL to the artistic work is actually minted into the blockchain. Usually, the data (representing the image, video, or another Work format) is stored in an Interplanetary File System (IPFS) which is a decentralized protocol and peer-to-peer network for storing and sharing data in a distributed file system. For example, the Pinata [10] system provide a convenient IPFS API and toolkit, to store NFT asset and metadata to ensure that the NFT is truly decentralized.

The minting process validates in a blockchain the transaction associated with the URL of the data (image, video, music, work) stored in an IPFS. Notice that the data itself is not minted into the blockchain. Watermarking is another verification layer of the authentication process along with procedures and evaluations provided by the marketplaces to verify for frauds of many types. As stated before, many artists are employing visible and invisible watermarking to reduce the number of frauds or even to help to detect when an artistic Work is stolen. Other approaches, out of the scope of this work, can be used to help to detect frauds, such as techniques to investigate image similarities and image forensics [11].

The third entity for certification purposes of the transaction can be implemented to help to validate the watermarking process using an Rivest–Shamir–Adleman (RSA) cryptographic protocol. The RSA is a public-key cryptosystem that is widely used for secure data transmission. The acronym “RSA” comes from the surnames of Ron Rivest, Adi Shamir and Leonard Adleman, who publicly described the algorithm in 1977 [12]. Both the work owner and the certification entity can use their private and public keys to improve the authenticity of the Work, the creator (authorship), and the certification entity by using the extra signature (watermark) embedded into the artistic Work. The checksum of this extra validation signature (watermark) can be also minted into a blockchain to register the transaction for extra security, keeping the decentralized approach for NFTs and digital coins.

Using the RSA cryptography process one can generate a watermark W to embed into the image (Work) in order to certificate the creation date, $DATECREA$, the owner identification, $USERID$, and other information. Assume an RSA symmetric cryptography using an encryption process named $PUB(., Key_{PUB})$ and a decryption process named $PRIV(., Key_{PRIV})$ with corresponding public and private keys Key_{PUB} and Key_{PRIV} such that

$$W = PRIV(Key_{PRIV}, PUB(Key_{PUB}, \ W)). \tag{4}$$

These processes need these public and private keys in order to properly encrypt and decrypt messages. The public key used for encryption may be distributed publicly without compromising the security while the private key should be only known to the message sender or the Work creator/owner. In the following, we present a certification protocol that validates the authenticity of the Work and the ownership of the creator.

5.1 Proposed watermarking certification protocol

Let's assume a certification entity is used for giving better credibility to the artists by showing and dealing with the artistic works, registering the transactions into a blockchain for public auditing as well as for validation of the embedded watermark. This entity can be one of the current marketplaces that register the URL of the Work along with other information into the blockchain, which is usually the Ethereum blockchain. Other related approaches can be found in [13–15].

For a given image, Im, a watermark, W, can be embedded using one of the many watermarking techniques, including the spread spectrum and LSB techniques explained above. The Work owner (buyer or creator) can use the services of the marketplace to create private and public keys, $Key_{PRIVUSER}$ and $Key_{PUBUSER}$, the private key is kept secure under the user personal and digital wallet. The marketplace also creates those keys, $Key_{PRIVMKT}$ and Key_{PUBMKT} for this transaction. The private keys should be kept secret from the owner and the marketplace. On the other hand, the process of embedding and extracting the watermark is public. The creation of the watermark is based on the user identity given by the marketplace when the account of owner is created, $USERID$, the date of the creation of work, $DATECREA$ and date of transaction (or minting into the blockchain), $DATEMI$, which are properly combined by concatenation, | operator. The owner encrypts his part of the watermark, W_1 using the public key of the marketplace and the marketplace encrypts its part of the watermark, W_2, using the public key of the owner, such that the final watermark, W, is composed of XOR operation, ⊕, from both parts:

$$W = PUB(Key_{PUBMKT}, USERID|DATECREA) \oplus PUB(Key_{PUBUSER}, USERID|DATEMI) \tag{5}$$

The watermark $W = W_1 \oplus W_2$ is then embedded into the work before storing the watermarked Work in an IPFS server. Notice that when necessary, the part W_1 can be generated by the entity marketplace that knows the part W_2 and the extracted watermark W by using the reversible property of the XOR operation explained above in the

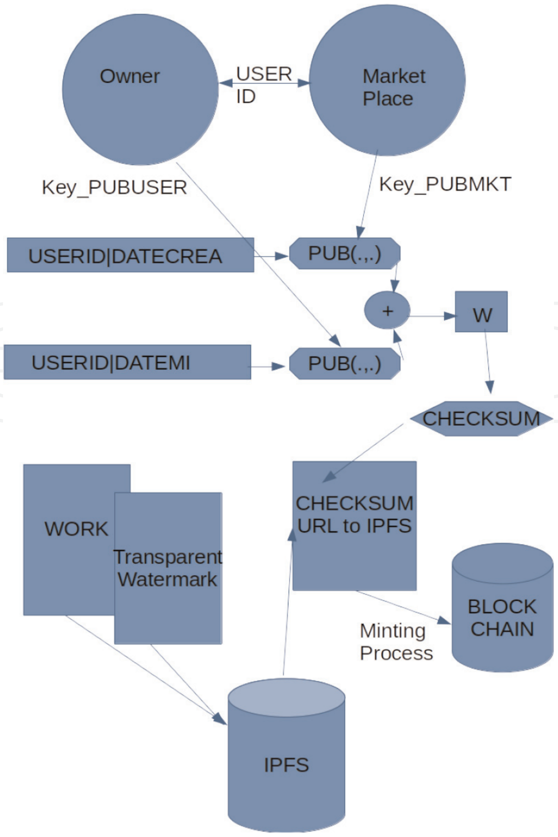

Figure 4.
Proposed Certification Protocol using the owner and a marketplace as entities to validate the ownership.

LSB technique section. The part W_2 can be generated by the owner in the same way. Moreover, a checksum, *CHKSUM*, of the watermark can be generated by one of the available algorithms [16]. This checksum is then included in the data that is going to be minted, *DATAMINT*, which includes the URL of the work in an IPFS server and other information. In order to reduce the “gas” fees, The checksum process should result in a much smaller amount of bits than the watermark itself and it aims to provide a validation of the authenticity of the embedded watermark itself. The proposed certification protocol is illustrated in **Figure 4**.

5.2 Contestant process using the embedded watermark

Consider that the marketplace system, by using forensics tools, finds out that a posted Work is duplicated or very similar. Alternatively, the owner or creator finds out that his work has been stolen and it was also minted into the blockchain after his original work was minted. By extracting the transparent watermarks from the works and using private keys to decrypt the relevant information, one can verify the authorship with the checksum that validates the watermarks into the blockchain, the creation and minting dates along with the users' identification. This information can be used as trusted legal evidence about the contested works. Therefore, the proposed process can be implemented by the marketplace providing a better and more secure service to the artist. Notice that validation depends on the marketplace and the owner's information about the transaction and the work itself. Both entities (owner and marketplace) can verify the corresponding part of the watermark $W = W_1 \oplus W_2$ and validate the ownership. Crossing these two information parts validate the entire process of mitigating a possible fraud from one of these entities. Variations of this protocol can be proposed to increase even more the trust in NFT trade and turning the art market even more valuable. Notice that visible watermarks and multiple transparent techniques can be used with advanced semi-fragile watermarking techniques.

6. Conclusions

In this chapter, we discussed how watermarking technology can be employed to increase the security of trading NFTs in this new and multimillionaire market. We propose that transparent embedded watermarks into the original work bring another level of security and do not preclude the use of visible watermarks and the traditional minting process used by current marketplaces. The additional checksum data may increase the costs of minting, however, brings a huge gain in terms of the capacity of securing the authorship of the artistic works in the market. We discuss basic transparent watermarking techniques in order to understand how to generate a watermark to employ with the proposed certification protocol. A certification protocol is discussed in detail and shown to be viable and very interesting to bring more confidence to artistic creators, owners, sellers, and buyers of artistic works.

Abbreviations

NFT	Non Fungile Token
URL	Universal Resource Locator
SS	Spread Spectrum
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
LSB	Least Significant Bit
XOR	Exclusive OR (binary operation)
RSA	Rivest–Shamir–Adleman public-key cryptosystem
IPFS	Interplanetary File System

Author details

Joceli Mayer
Eletrical and Electronics Department (EEL), Universidade Federal de Santa Catarina, UFSC, Florianopolis, Brazil

*Address all correspondence to: mayer@eel.ufsc.br

References

[1] Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System, Bitcoin White Paper. 2009. Available from: https://www.bitcoin.com/bitcoin.pdf

[2] Yu X, Wang C, Zhou X. Review on semi-fragile watermarking algorithms for content authentication of digital images. Future Internet. 2017; **9**:56

[3] Thorpe E. Colorful Architecture. Available from: https://www.freeimages.com/photo/colorful-architecture-1-1216925

[4] Mayer J. Optimization of Multibit Watermarking, Watermarking Book. London, UK: Intechopen; 2012

[5] Cox IJ, Kilian J, Leighton FT, Shamoon T. Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing. 1997; **6**(12):1673

[6] Mayer J, Bermudez JCM, Silverio AV. On the design of pattern sequences for spread spectrum image watermarking. In: IEEE International Telecommunications Symphosium, ITS'02. Natal; 2002

[7] Mudgil S. How to Write & Deploy an NFT. 2021. Available from: https://ethereum.org/en/developers/tutorials/how-to-write-and-deploy-an-nft. [Accessed: April 20, 2022]

[8] Zipmex. How To Mint NFTs On The NFT Marketplace? Available from: https://zipmex.com/learn/nft-minting-explained/. [Accessed: April 21, 2022]

[9] Rarible NFT Marketplace. Available from: https://rarible.com/. [Accessed: April 21, 2022]

[10] Pinata IPFS Site. Available from: https://www.pinata.cloud/. [Accessed: April 21, 2022]

[11] Piva A. An Overview on Image Forensics, ISRN Signal Image Forensics, ISRN Signal. London, United Kingdom: Published by Hindawi; 2013

[12] Nisha S, Farik M. RSA public key cryptography algorithm—A review. International Journal of Scientific & Technology Research. 2017;**6**(7):187-191

[13] Dedge O, Shingade R, Jagtap A, Yadav A, Kamble A. Image copyright protection system using blockchain. International Journal of Future Generation and Communication Networking. 2020;**13**(3s):37-43

[14] Gountia D. Towards scalability trade-off and security issues in state-of-the-art blockchain. EAI Endorsed Transactions on Security and Safety. 2019;**5**(18):e4-e4

[15] Joshi A, Mishra V, Patrikar RM. Real time implementation of digital watermarking algorithm for image and video application. Watermarking. 2012; **2**:65

[16] Available from: https://en.wikipedia.org/wiki/Checksum. [Accessed: April, 23, 2022]