Digital Media Copyright and Content Protection Using IPFS and Blockchain

Kwame Opuni-Boachie Obour Agyekum[1], Qi Xia[1], Yansong Liu[1], Hong Pu[1], Christian Nii Aflah Cobblah[1], Goodlet Akwasi Kusi[1], Hanlin Yang[1], and Jianbin Gao[2(✉)]

[1] Center for Cyber Security, University of Electronic Science and Technology of China, Chengdu 611731, China
obour539@yahoo.com, xiaqi@uestc.edu.cn, johnsonlys@outlook.com, puhong1997@outlook.com, kriscobblah@gmail.com, gkusikat@gmail.com, p1usj4de@gmail.com

[2] School of Resources and Environment, Center for Digital Health, University of Electronic Science and Technology of China, Chengdu 611731, China
gaojb@uestc.edu.cn

Abstract. With the development of the Internet, the total amount of digital media transmission and storage has been very large. However, traditional digital copyright solidification and rights protection are offline processes, which are time-consuming and labor-intensive. This paper focuses on digital fingerprint technology, Inter-planetary File System (IPFS) and Fabric Alliance blockchain technology to create a digital system to optimize traditional processes and improve the efficiency of digital media copyright solidification.

In this work, we build an IPFS system where all digital media files exist in the form of IPFS objects. Digital fingerprints, which reflect the characteristics of the digital media files are also designed. Finally, all files are recorded onto the Fabric blockchain, ensuring immutability and provenance. Our system proves to be efficient and accurate in ensuring complete ownership of digital media.

Keywords: Blockchain · Digital fingerprinting · Fabric · IPFS

1 Introduction

The massive Internet adoption has resulted in rapidly growing cases involving the protection of digital media, content copyright and access to patented multimedia files. These are now major concerns for authors and holders of copyrights [1]. The Internet is widely used to trade and distribute digital content such as music, software, games, footage, video, and text. At the same time, the traditional protection mechanism for copyright is difficult to adapt to the protection of digital content copyright requirements with the arrival of the information age [2].

Incidents regarding the infringement of digital content online are increasing as well as the infringement and piracy challenge of digital content on the Internet [2]. Users deliberately or accidentally disclose valuable copyrighted resources

Y. Zhao et al. (Eds.): ICIG 2019, LNCS 11903, pp. 266–277, 2019.
https://doi.org/10.1007/978-3-030-34113-8_23

when using and disseminating digital content, consume potential user resources and harm creators and related copyright holder's economic interests. Digital protection of copyright has become a key issue, which has received great attention.

It is essential for the creator of media files to have complete access, authorization and a secure medium for sharing their information to stop such incidents and to expose and punish actual guilty parties. The owner should be able to show copyright for its contents if information is leaked [3]. Online media like the DropBox or peer-to-peer (P2P) media are the best available tools for rapid and effective data sharing and privacy protection [2].

Existing solutions such as digital rights management, Micro-payments, paywalls, etc. have their merits and setbacks: no regard is given to the author. Consequently, our studies revealed that Blockchain technology in conjunction with Inter-planetary File Systems (IPFS), which is a P2P distributed file system that seeks to replace HTTP and build a much better web for all, can help alleviate some of the issues regarding digital content copyright. IPFS is a P2P file system that incorporates the knowledge gained from many previous successful systems [4]. It merges a distributed hash table, an incentivized block exchange and a name-space that certifies itself. IPFS is a P2P hypermedia protocol that renders the web faster, safer, and more open [4].

The concept behind decentralized copyright management is premised on the storage in an extensive blockchain of creative works and their metadata [3]. This would enable the distribution of digital content to be monitored in real time, licenses checked, certain applications permitted, and precisely calculated. This can also verify the authenticity of a work. This is already in progress, for example, pay-per-use offerings are available that use blockchain technology. Examples of such systems are PeerTracks, Ujo Music, Mediachain Attribution Engine, Blo [2].

In this paper, we introduce a novel idea of combining the advantages of IPFS and Blockchain technology to help protect the copyright of authors and copyright holders. To the best of our knowledge, this is the first of its kind.

The organization of the paper is as follows: Sect. 2 presents the works related to this study, with Sect. 3 introducing the technologies utilized in this work. Section 4 presents the model and implementation of the system, while Sect. 5 provides the operation and results obtained. Finally, Sect. 6 concludes the paper.

2 Related Works

The idea of utilizing the blockchain technology for licenses verification has been illustrated in some research works of which Herbert and Litchfield [3] stand to have a solid preposition based on their research works. They characterized the blockchain into two main forms for licenses permit verification. Within the Bitcoin transaction model, clients demonstrate possession by demonstrating that they have the bitcoin from an active software vendor they engage in the transaction. Another model of Bitcoin is the Bespoke model which is more advanced with the extra data fields that permit computer programmers, and software manufacturers alike to also subscribe and store their license information on the

system such as permit expiration, permit limits, and others. Custom models are utilized for clients with a single permit. However, the limitation is to have these stakeholders to hold numerous licenses.

McConaghy and Holtzman [5] utilized Bitcoin blockchain to also record media data by focusing on image ownership by Legal Registry capacity terms, and timestamps are recorded as the image attributes. The registry is put in a blockchain network with proprietorship data; web crawlers utilize machine learning techniques to distinguish pictures put on websites without the owner's consent, which in so doing violates the data owner's rights and discredits the work since individuals can claim the sole ownership of such data as well as enjoying any benefits that come with it thereof.

Kishigami et al. [6] depict a distributed blockchain based system as a management hub for media files and documents. The key potential of their system is the use of blockchain to oversee rights for over 4k video in a distributed environment which permits copyright holders to manage their content licenses and simultaneously counting the connected client licenses for its content. Their design framework does not allow offline access of the content on the system and the clients also expect the online miners to exchange ID to facilitate decoding of the content of the media file.

Gao and Nobuhara [7] also proposed a strategy for putting away up to N times, 20 bytes in the blockchain and for creators and owners, computerized records technically don't need timestamp anonymity. The strategy can embed the hashes of electronic information and related data such as author name, file name, and comments. Mostly these comments and other related data can be accounted for by writing it in a plain text basically. The owner of the digital media file can ensure data protection and integrity by this approach. The limitation was the improvement in the application of subsequent intellectual property rights other than patent protection policies and obvious security.

"Mediachain", also made use of IPFS (Inter-planetary File System) record framework [8], enhances copyrights of computerized works. As of now, it's basically used for computerized picture copyright security applications. Mediachain, however, could be a collaborative solution that combines media metadata convention that gives copyright recognition for imaginative works. Copyright proprietors are able to sign metadata statements with cryptographic marks on their finished media file and with the use of timestamps on the blockchain, the records are stored in IPFS. These claims can be accessed by inputting a query that calls the data.

IPFS ideology as it stands can be acknowledged from many past effective systems and frameworks, and it has overcome numerous challenges and constraints. Nevertheless, we think a few perspectives of IPFS can be improved with the coupling of blockchain, and until presently, this thought has not yet been realized.

3 Background

3.1 Inter Planetary File System (IPFS)

IPFS [4] is a file sharing platform, which identifies files through its content. A distributed hash table (DHT) is used to recover file locations and node connectivity data. When a file is fed into the IPFS system, it is divided into chunks, with each chunk containing a maximum of 256 kbs of data and/or connections to other chunks. Each chunk is recognized by a cryptographic hash, also referred to as content identifier, which is calculated from its content [11].

The connections also comprise contents' identifiers, forming a Merkle directed acyclic graph (DAG) that explains the entire file and can be used to recreate any file from its chunks. Due to the Merkle DAG, the root hash can be used to identify a complete file. The node is registered as a provider through DHT once the node is divided into chunks, and the Merkle DAG has been formed. The DHT [12] is basically a distributed store of key-value pairs. It utilizes node identifiers and keys, with a distance metric to store and collect data readily, both having to be the same length. As a result of the use of IPFS's content identifiers, it is particularly appropriate for blockchain use in order to locate, verify and transfer chunks and files [11].

Indeed, the root hash of a Merkle DAG file can be sent via a transaction to the blockchain and thus, does not expand the chain. Simultaneously, no information other than the hash is needed to retrieve the file from IPFS. This is distinct to file sharing schemes that do not depend on cryptographic hashes for content identification, as they need a file whose hash is discovered in the blockchain with extra name resolution mechanisms.

Moreover, it is difficult to create a second, separate file with the same hash so IPFS cannot be flooded with files with a specific target file identifier [11]. In conclusion, an IPFS-based file can be checked readily and it is hard to thwart a user with the same name/identifier in several files. The issue of decentralized storage of large files is resolved with IPFS. When you upload a new file to IPFS, it is divided into several chunks of data. Each chunk has a hash value for its content to be clearly identified. Finally, for the whole file, a hash value is created which can be downloaded by any node.

3.2 Blockchain

Blockchain [10] is a decentralized infrastructure and a distributed computing paradigm which uses cryptographic chained block structures to validate and store data, consensus algorithms to generate and update data, and smart contracts, programmable scripts, to program and manipulate data.

Blockchain management is achieved by multiple participants, with each participant providing nodes and storing data on the chain. This enables an easy sharing of data. The use of a time-stamped chain of data makes for easy verification and traceability. Each modification to the data is recorded and can be tracked on the blockchain. The utilization of Merkle trees and blockchain hash

tables enables an irreversible modification of the data. Information recorded in the blockchain cannot be tampered with, after reaching a consensus.

Blockchain transactions only hold very small amounts of data, wherefore it is crucial to choose what data should be placed On-chain and what should be kept Off-chain [11]. There are many off-chain storage solutions customized for blockchain such as IPFS [4]. These solutions share the concept of a distributed peer-to-peer file system where the information is split into different chunks, encrypted and disseminated across various network nodes to guarantee security and accessibility.

4 System Design and Implementation

4.1 System Model

In this section, we outline the various components, and also illustrate the flow, of the system.

Users upload files and metadata information through the web-client. The web client is written in JavaScript, and node.npm and yarn, which are tools that ensure the efficient functioning of the system. The server-side, which is designed by the Flask framework and based on python programming language, finds the file transfer port through the HTTP response and saves the file at the localhost. The server-side transfers the file to IPFS by invoking the HTTP protocol, and IPFS returns the hash for the corresponding file. Meanwhile, the digital fingerprint extraction module is used to extract the digital fingerprint of the corresponding file (digital fingerprint is a string of binary characters). After the hash characters and digital fingerprint are received at the server-side, fabric network is called remotely through the Remote Procedure Call (gRPC) service to take the digital fingerprint as the key, and combines the hash characters and the metadata information after re-decoding as the value. The key-value pairs are then uploaded to the fabric network. The data in fabric network are synchronized and updated by kafka consensus algorithm.

When other users want to upload the same file to the system (which can be interpreted as a copyright infringement), the system extracts the digital fingerprint after receiving the file, and sends the digital fingerprint as the key to the fabric network for a thorough search. If the digital fingerprint already exists, the system will report an error, roll back the upload operation of the current user and remind the user that the same file already exists.

4.2 System Implementation

IPFS provides a new platform for writing and deploying applications, as well as a new distribution system for versioning big data. So the sum is greater than the parts. Like all decentralized systems, IPFS is point-to-point, with no privileged nodes. The node stores the object IPFS in local storage. These files or other data structures are then linked and transmitted through the nodes.

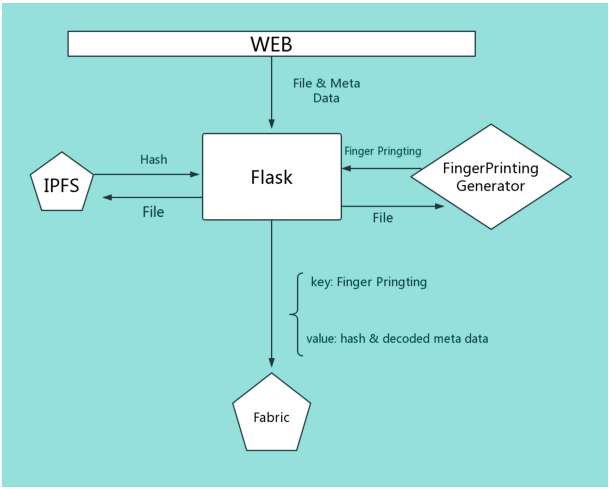

Fig. 1. System model

By transferring files with private information off-line, some weaknesses of blockchain can be overcome. This guarantees that files can be removed when necessary and that the chain does not expand too quickly. However, the guarantee that files have not been edited or modified must be maintained for information sharing between participants on the network. In addition, the Fabric blockchain monitors which participants have access to which data.

IPFS stores data in a dispersed manner by dividing them into pieces that can be asked for, and transmitted between nodes. With its cryptographic hash, each file is identified. This makes it simple to verify that the right information is obtained by creating the hash of the file chunks and ensuring that it matches the required hash. IPFS hashes identifying files that contain personal data can be stored on the blockchain instead of the data itself.

The blockchain containing the hash guarantees that the file has not been altered. The file itself being stored in IPFS means it can be removed as required, by nodes deleting it from their local storage. File hashes can be used to connect owner's files with access permissions. Thus, the growth of the chain is significantly decreased as hashes are normally smaller than the data they represent. The combination of blockchain security and a private-network IPFS guarantees data protection and decrease in storage costs. The structure of the block is given in Fig. 2.

4.3 Digital Fingerprint Generation

The digital fingerprint generation, which has been presented by several authors in literature [14]–[18], is the key component of this system, and it marks the unique key of identifying copyright information from the system. The requirement for the algorithm is to be reliable, fast and accurate. For digital media

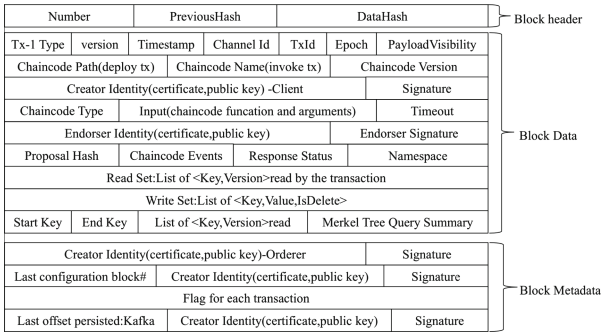

Fig. 2. Block structure

files of different carriers, different digital fingerprint algorithms should be used to reflect the uniqueness of the files.

1. **Perceptual hash algorithm used by an image:** An image is composed of pixels. Each pixel has its own color, and we can use the rgb channel to describe this color. But if the color of the pixel of each image is described as the digital fingerprint of the picture, the length of this fingerprint will be very large. It's difficult to use this in specific projects, and not conducive for communication. Since each picture has different image features, if we can digitally describe these features, it can also be used as a digital fingerprint of this picture.
   To make a distinction between two similar photos, we can use the encryption hash algorithm. The calculated hash will output a fixed length hash-key according to different inputs (different binary arrangements). If the two images are the same, then their hash-key will be the same. This paper uses the perceptual hash algorithm. The perceptual hash algorithm mainly includes ahash, phash, and dhash, and is not a strict hash calculation method, but a hash value is calculated in a relative manner. This similarity can be used to digitize image features.
   – **ahash:** average hash, faster, but generally accurate.
   – **phash:** perceives hash, with higher accuracy, but at a fair speed.
   – **dhash:** difference value hash, higher precision, faster.
2. **Video Segmentation Key-frame dhash algorithm:** The video processing method in this paper adopts the dhash algorithm mainly for the consideration of processing speed and accuracy. But a video has a few more steps than dealing with a single image. The video is formed by a continuous playback of each frame of the picture, but like the picture, we can't say that each frame of the picture is included in the calculation range. We can extract the picture number of the specific frame by equally dividing the key frame. The picture dhash fingerprint of a specific frame is calculated, and then all the fingerprints are spliced. The result of this stitching is the dhash fingerprint of the video.

3. **Audio fingerprinting:** After the whole audio is divided into blocks, Fourier transform is performed on each block respectively, and then the molecular band extracts the subscript of the highest energy point.

4.4 Fabric Consensus Algorithm

The essence of each node of Fabric is a state machine that is replicated, and the same ledger needs to be stored between the nodes. In Fabric, the consensus process is used to achieve the consistency of distributed nodes and the consistency of ledger state. Fabric's consensus process has three phases: endorsement, sorting, and verification. The fabric network is started and the chaincode is successfully invoked. The fabric network guarantees that nobody can tamper with the data, and that ensures traceability.

In order not to limit links or data flow between nodes, the IPFS is intended to communicate information as broadly as possible. The situation in question concerns personal information. Thus, IPFS cannot exchange files with any entity that asks for it. In accordance with permissions registered on the Fabric blockchain by the data owners, only participants with access permissions must be allowed access. As such, the provided solution is an IPFS that makes use of a chaincode (smart contract) which has the ability of adding, deleting and updating file ownership and accesses. Participants of the system are identified by their public keys and files are identified by their unique cryptographic hash.

5 Operation Manual and Results

In this section, we discuss the operable nature of the setup and show some results obtained. For the setup of the system, the following are the steps we adhered to.

We first initiate the IPFS setup. Once a *Daemon is ready* message appears, it indicates a successful setup. Next, we start the Fabric blockchain network. We pre-install Docker and docker-compose, and also pull the images beforehand. All startup steps needed for the complete functionality of the Fabric network are already written in the script file, so the file is run once the path that contains the script file is located. Once the network has been started, the chaincode is invoked. We then start the gRPC service, by entering the path where the gRPC files are located. The gRPC service connects the server-side and the Fabric blockchain network. Next, we start the Flask server and run the main function. Figures 3 to 5 illustrate the setup processes and the results obtained.

For an owner who wants to publish his copyrighted work on the blockchain network, s/he has to send the name and possibly, remarks on the work. Once the file exists on the network, an error message will be generated once anyone tries to claim ownership of that same work and publishes it on the blockchain network.

We further compare our system to other related systems by taking into consideration some performance metrics such as data security, bandwidth efficiency,

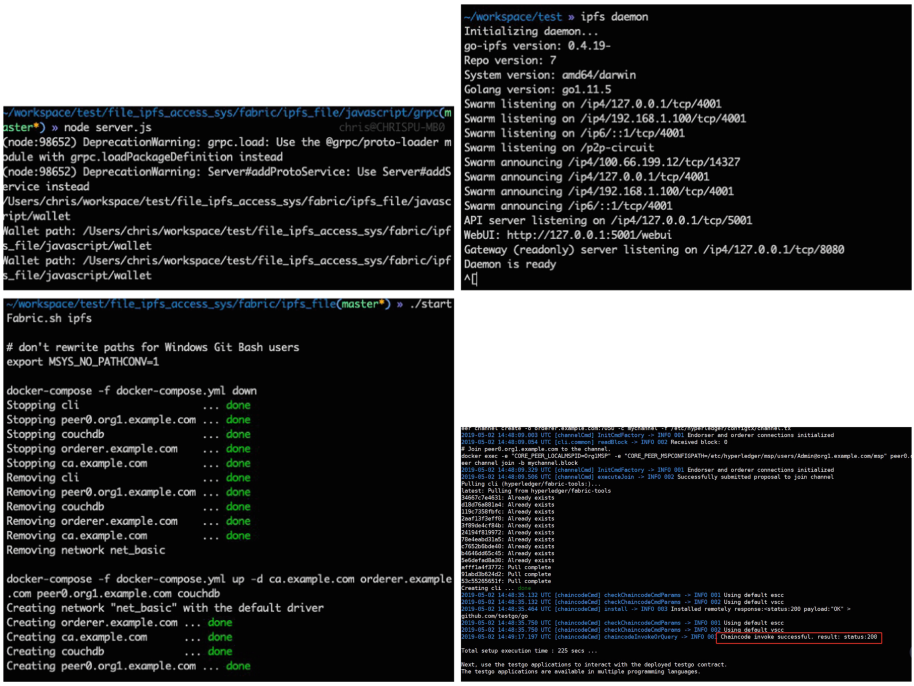

Fig. 3. Setup of the system: 1. gRPC 2. IPFS 3. Fabric 4. Network

access control, provenance and auditing, privacy and storage costs. Table 1 provides a summary of the comparison, with the following notations: Y - Yes, N - No, H - High, M - Moderate, and L - Low. The figure depicts that our system performs very well as compared to other related existing systems.

Figure 3 illustrates the interfaces in starting the system, while Fig. 4 represents the interface for uploading the files onto the system. The first picture of

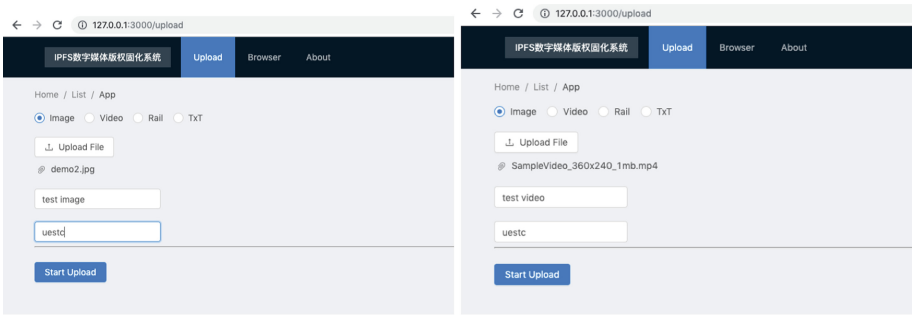

Fig. 4. Data upload onto server: 1. image 2. video

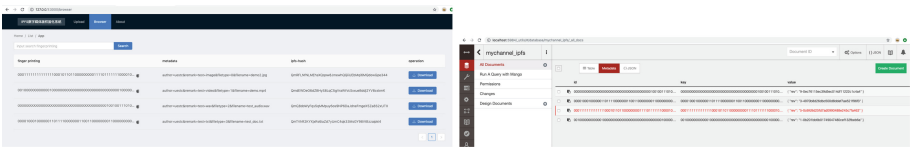

Fig. 5. 1. Interface showing finger printing, metadata and hash of uploaded files 2. Records on the fabric blockchain

Table 1. Comparison between proposed and other related systems

Metric	[3]	[5]	[6]	[7]	[8]	Our system
Blockchian copyright protection	Y	Y	Y	Y	Y	Y
Data security	Y	Y	Y	Y	Y	Y
Bandwidth efficiency	N	Y	N	N	N	Y
Access control	Y	N	N	N	Y	Y
Provenance and auditing	N	Y	N	Y	Y	Y
Privacy	N	Y	Y	N	N	Y
Storage costs	H	H	M	H	M	L

Fig. 4 illustrates an image upload while the second one illustrates a video upload. Once the file has been uploaded onto the server, there is an interface that shows the fingerprint of the uploaded file, the name of the file and the generated hash value of the file. This can be seen in the first image of Fig. 5. The second image shows the details of the uploaded files after querying the IPFS server.

6 Conclusion

Digital media is a very sensitive data and can easily be stolen or copied by third parties. The copyright holders should be able to own and monitor their data without compromising safety instead of trusting a third party to storing or distributing the data. Similarly, traditional digital copyright protection mechanisms are inefficient simply because notarization is done offline and it takes a very long time before it's completed.

This paper therefore proposes a system that incorporates the use of a digital fingerprint technology, IPFS and Blockchain technology (Fabric) as its core constituents, to optimize the traditional processes and help protect digital media copyright and ensure complete ownership. This ensures that copyright holders or authors know when their data is being accessed and how it is used, so that they don't have to trust any third party. Moreover, investigations into the using IPFS and Blockchain in the protection of digital media copyright has a high potential value, which is important in improving the current situation of digital media copyright protection. Copyright organizations should also cooperate actively in support of Blockchain's digital copyright efforts and development policies.

Acknowledgment. This work was supported in part by the programs of International Science and Technology Cooperation and Exchange of Sichuan Province under Grant 2017HH0028, Grant 2018HH0102 and Grant 2019YFH0014.

References

1. Vishwa, A., Hussain, F.K.: A blockchain based approach for multimedia privacy protection and provenance. In: IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, vol. 2018, pp. 1941–1945 (2018). https://doi.org/10.1109/SSCI.2018.8628636
2. Qi, Y., Liu, X.: Digital copyright protection based on blockchain technology. In: 2018 International Journal of Knowledge and Language Processing, vol. 9, no. 2, pp. 61–70 (2018)
3. Bhowmik, D., Feng, T.: The multimedia blockchain: a distributed and tamper-proof media transaction framework. In: 2017 22nd International Conference on Digital Signal Processing (DSP), London, pp. 1–5 (2017). https://doi.org/10.1109/ICDSP.2017.8096051
4. Benet, J.: IPFS - Content Addressed, Versioned, P2P FileSystem, July 2014
5. McConaghy, M., McMullen, G., Parry, G., McConaghy, T., Holtzman, D.: Visibility and digital art: blockchain as an ownership layer on the Internet. Strateg. Change **26**(5), 461–470 (2017)
6. Kishigami, J., Fujimura, S., Watanabe, H., Nakadaira, A., Akutsu, A.: The blockchain-based digital content distribution system. In: 2015 IEEE Fifth International Conference on Big Data and Cloud Computing, Dalian, pp. 187–190 (2015). https://doi.org/10.1109/BDCloud.2015.60
7. Gao, Y., Nobuhara, H.: A decentralized trusted timestamping based on blockchains. IEEJ J. Ind. Appl. **2017**, 252–257 (2017). https://doi.org/10.1541/ieejjia.6.252
8. How Mediachain Works. http://www.mediachain.io/
9. Baumgart, I., Mies, S.: S/Kademlia: a practicable approach towards secure key-based routing. In: 2007 International Conference on Parallel and Distributed Systems, Hsinchu, pp. 1–8 (2007). https://doi.org/10.1109/ICPADS.2007.4447808
10. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system, October 2008. (cited on pp. 15 and 87) (2017). http://www.bitcoin.org/bitcoin.pdf
11. Steichen, M., Pontiveros, B.F., Norvill, R., Shbair, W., et al.: Blockchain-based, decentralized access control for IPFS. In: The 2018 IEEE International Conference on Blockchain (Blockchain-2018), pp. 1499–1506. IEEE (2018)
12. Maymounkov, P., Mazières, D.: Kademlia: a peer-to-peer information system based on the XOR metric. In: Druschel, P., Kaashoek, F., Rowstron, A. (eds.) IPTPS 2002. LNCS, vol. 2429, pp. 53–65. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45748-8_5
13. J. Herbert and A. Litchfield, A Novel Method for Decentralised Peer-to-peer Software License Validation Using Cryptocurrency Blockchain Technology, ACSC 2015, 2015(27)
14. Cappelli, R., Erol, A., Maio, D., Maltoni, D.: Synthetic fingerprint -image generation. In: Proceedings 15th International Conference on Pattern Recognition, Barcelona, vol. 3, pp. 475–478, September 2000
15. Cappelli, R., Maio, D., Maltoni, D.: Synthetic fingerprint -database generation. In: Proceedings 16th International Conference on Pattern Recognition, Quebec City, vol. 3, pp. 744–747, August 2002

16. Bontrager, P., Togelius, J., Memon, N.: Deepmasterprint: generating fingerprints for presentation attacks. https://arxiv.org/abs/1705.07386 (2017)
17. Abdelnur, H.J., State, R., Festor, O.: Advanced network fingerprinting. In: Lippmann, R., Kirda, E., Trachtenberg, A. (eds.) RAID 2008. LNCS, vol. 5230, pp. 372–389. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87403-4_20
18. Minaee, S., Abdolrashidi, A.: Finger-GAN: generating realistic fingerprint images using connectivity imposed GAN. arXiv preprint arXiv:1812.10482 (2018)