VERIFYING THE INTEGRITY OF DIGITAL FILES USING DECENTRALIZED TIMESTAMPING ON THE BLOCKCHAIN

Akash Dhande
Dept. of Computer Enginieering
Smt. Kashibai Navale College of Engineering, Pune, India
akashdhande.1997@gmail.com

Anuj Jain
Dept. of Computer Enginieering
Smt. Kashibai Navale College of Engineering, Pune, India
anujjain0703@gmail.com

Tejas Jain
Dept. of Computer Enginieering
Smt. Kashibai Navale College of Engineering, Pune, India
tejasjain52@gmail.com

Tushar Mhaslekar
Dept. of Computer Enginieering
Smt. Kashibai Navale College of Engineering, Pune, India
tusharmhaslekar@gmail.com

Prof. P. N. Railkar
Dept. of Computer Enginieering
Smt. Kashibai Navale College of Engineering, Pune, India
poonamrailkar@gmail.com

***Abstract*—In today's day and age, the integrity and the authenticity of the digital files and document is a critical issue. Especially if those digital files are to be submitted as the evidence in court. For example, a video file of an accident. Fakers can exploit such files by editing the video and leading the court into the wrong judgement. Therefore, this paper proposes a system to prove the integrity of a digital file such as video proof of accident in the above example. The complete system consists of three functions, one for calculating and storing the hash value of the digital file and its details, second for proving the integrity of the given file by comparing it with stored hash and its timestamp and the third function is for storing and retrieving the original file stored on the InterPlanetary File System (IPFS) network. In this approach, one can store the integrity of a file and can use it to prove the authenticity of that file by comparing the hash of another file with the stored hash. This paper proposes a system that uses the new and emerging technology of Blockchain to secure the integrity of the digital files.**

Keywords—Decentralized, Blockchain, Timestamping, IPFS

1. INTRODUCTION

The proposed system uses the new and emerging technologies like Blockchain to store the hash of the files uploaded by user which will be the integrity of that file and IPFS (InterPlanetary File System) Network which will be used to store the original files uploaded along with the trusted timestamping and the location of the file from where it was uploaded.

Blockchain is a decentralized ledger or data structure. It can be referred as blocks in a chain where the corresponding blocks refer to the blocks, prior to them [5]. A blockchain is a decentralized, distributed and public digital ledger that is used to record transactions across many computers so that the record cannot be altered retroactively without the alteration of all subsequent blocks and the consensus of the network. This allows the participants to verify and audit transactions inexpensively. A blockchain database is managed autonomously using a peer-to-peer network and a distributed timestamping [7]. They are authenticated by mass collaboration powered by collective self-interests. The result is a robust workflow where participants' uncertainty regarding data security is marginal. So essentially a blockchain is a distributed ledger which cannot be tampered with ensuring the security of the data stored in it.

IPFS (InterPlanetary File System) is a peer-to-peer distributed file system that seeks to connect all computing devices with the same system of files [1]. In some ways, IPFS is similar to the World Wide Web, but IPFS could be seen as a single BitTorrent swarm, exchanging objects within one Git repository. In other words, IPFS provides a high- throughput, content- addressed block storage model, with content-addressed hyperlinks. This forms a generalized Merkle directed acyclic graph (DAG). IPFS combines a distributed hash table, an incentivized block exchange, and a self-certifying namespace. IPFS has no single point of failure, and nodes do not need to trust each other not to tamper with data in transit. Distributed Content Delivery saves bandwidth and prevents DDoS attacks, which HTTP struggles with.

Section 2 contains State of Art, Section 3 contains Gap Analysis, Section 4 User Classes and Characteristics, Section 5 contains Proposed Work, Section 6 contains Conclusion and Future Work and the Section 7 contains References.

2. STATE OF ART

A. Timestamping

The paper talks about timestamping which is used in cryptocurrencies like bitcoin. This system was proposed by Norman Neuschke & Andre Gernnandt in 2015. System uses hash of digital data and can be used to record the transactions into the blocks [2][3][4].

B. Blockchain

This paper was proposed by Rishav and Rajdeep Chaterjee in 2017. It consists of detailed implementation of blockchain technology and its use-cases includes transactions of multiple parties based on Hyperledger [5][7].

C. Digital watermarking

Copyright management system based on digital watermarking includes blockchain perceptual hash function, quick response code (QR), InterPlanetary File System (IPFS) related work to compare copyrights of digital files. The concept of digital watermarking was introduced by Meng Zhaoxiong Morizumi Tetsuya in 2018 [1].

D. IPFS

It is a peer-to-peer distributed file system that seeks to connect all computing devices with the same system of files. IPFS combines a distributed hashtable, an incentivized block exchange, and a self-certifying namespace. IPFS has no single point of failure, and nodes do not need to trust each other. This concept was introduced by Juan Benet [8].

E. Blockchain: Challenges and Applications

Blockchain is a form of database storage that is noncentralized, reliable, and difficult to use for fraudulent purposes. Transactions are made with no middle men in blockchain. This work proposed by Pinyaphat Tasatanattakool and Chian Techapanupreeda in 2018 [9].

F. Secure and Trustworthy Application

This paper was proposed in 2018 by Huayi Duan. The design system gives secure and trustworthy blockchain applications and systems in their own workplaces [10].

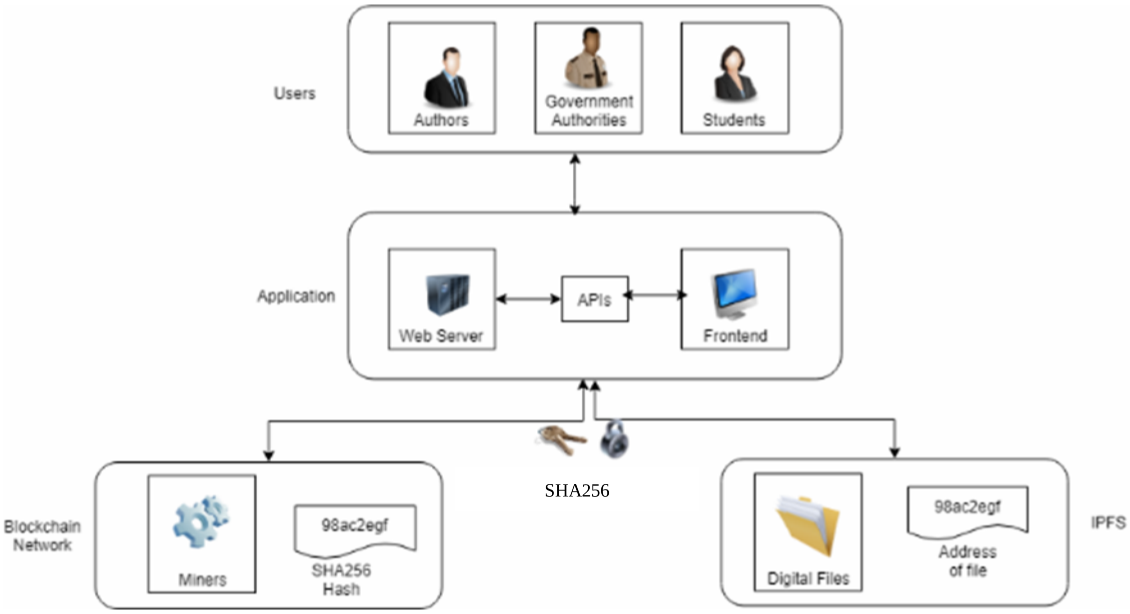

Fig. 1 System Architecture

3. GAP ANALYSIS

	Manual Verification	Govt. DVS	Proposed services
Validity	Unlimited	Unlimited	Unlimited
Confidentiality	Moderate	Medium	High
Cost of verification	Medium	Medium	Low
Security	Low	Medium	High
Energy Consumption	Nil	Medium	High

Table 1: Gap Analysis

Table 1 compares various methods available today for verifying the integrity of any document.

Manual Verification methods like attestation by Govt. officials is the most common method used today. The security is low since documents can be easily tampered. Also, a moderate cost is associated with each attestation.

The Document Verification Service (DVS) is an online system that allows organizations to compare a customer's identifying information with a government record. The DVS is a secure system that operates 24/7 and matches key details contained on Australian-issued identifying credentials, providing a 'yes' or 'no' answer within seconds. This helps protect governments, businesses and civilians from identity crime. Drawback of the method is that the entire database is centralized. One can only verify government documents using this service.

The proposed system securely stores the hash code of any type of digital file on a decentralized, tamper-proof ledger. Once a hash fingerprint is embedded in the blockchain, it is immutable and will exist "forever" as a trusted timestamp. Confidential files can also be used since only the hashes are stored and not the actual file. This service is free of charge.

4. USER CLASSES AND CHARACTERISTICS

PRIMARY USER: The main User Class that is going to use this product is Insurance companies, Government authorities, Private sectors like IT companies. The platform frequency of use could be on a daily basis as every sector, every operations need raw data in the form of files, images, videos i.e. digital documents.

VERIFIERS: Another User Class is of verifiers who need to verify their digital documents by uploading them on the web portal. By generating different hash values, user can distinguish between the files. As this platform is innovative, any user can have free and easy access of the platform. The platform has a web-based platform and users will require an account to secure their files.

4.1 SYSTEM FEATURES

Decentralized Timestamping: Trusted timestamping is the process of securely keeping track of the creation and modification time of a document. Security here means that no one—not even the owner of the document—should be able to change it once it has been recorded provided that the time stamper's integrity is never compromised [9][2].

Blockchain Network: The Blockchain Network consists of two parts:

1. Calculation of Hash using SHA-256 function.
2. Creation of Blocks with has values.

IPFS System: InterPlanetary File System (IPFS) is a protocol and network designed to create a content-addressable, peer-to-peer method of storing and sharing hypermedia in a distributed file system [1][6].

5. PROPOSED WORK

This project aims to achieve and maintain the digital integrity of the files. As the court procedures and insurance procedures take a long time to identify, analyze and verify the digital files our portal will shorten the process largely and help the users to get the overall process go much faster.

5.1 ALGORITHM

```
I.     register(userDetails)
II.    upload_file(file)
III.   fileHash = calculate_sha256_hash(file)
IV.    createNewTransaction(owner, fileHash, remarks)
V.     upload_file_IPFS(file)
VI.    register(verifierDetails)
VII.   search_blockchain(hash)
VIII.  verify_results()
```

5.2 PROCESS OF THE SCHEME

1. Users like government authorities, artists, students who want to secure the integrity of a digital file or claim the ownership of a file register themselves on the web portal.

2. User uploads the file on the web portal.

3. SHA256 hash of the uploaded file is calculated.

4. The newly created block contains various information like owner name, file hash, timestamp, block of previous hash and custom remarks by the user (if any). This transaction is added in the list of pending transactions. Once the block is mined by a network node, it is added in the blockchain network along with the calculated nonce.

5. With user's consent, add the file on IPFS for future retrieval. In IPFS, each file and all of the blocks within it are given a unique fingerprint called a cryptographic hash. Each network node stores only content it is interested in, and some indexing information that helps figure out who is storing what. When looking up files, you're asking the network to find nodes storing the content behind a unique hash [6].

6. Verifying authorities like courts, government bodies (e-Seva Kendra, UIDAI, Universities) register themselves on the web portal.

7. Verifiers can search the blockchain network by uploading the file or searching by name, hash, date or time.

8. The results are fetched from the blockchain and if hash matches, then the file is valid.

6. CONCLUSION AND FUTURE WORK

The proposed web-based application converts the inputs to hash values using Secure Hashing Algorithms. The converted hash codes are stored in decentralized and tamperproof transaction ledger, i.e. the blockchain along with its timestamp. Currently, courts do not routinely accept video/image footage as evidence, because it is impossible to prove that the files were not manipulated after the incident. By using the blockchain of a to store a hash of the digital file it can be proven that the footage was not manipulated. Any tampering with the file in retrospect would result in a file hash that no longer matches the hash that was embedded in the blockchain.

The scope of our application in future is by extending it to a mobile application which can be used to record live and onsite data and secure its integrity. Also, an application may be extended to add second factor of authentication by tracking and storing live locations on to the blockchain network.

7. REFERENCES

[1] MENG Zhaoxiong, MORIZUMI Tetsuya, MIYATA Sumiko and KINOSHITA Hirotsugu, “Design Scheme of Copyright Management System Based on Digital Watermarking and Blockchain” 42nd IEEE International Conference on Computer Software & Applications 2018.

[2] Bela Gipp, Corinna Breitinger, Norman Meuschke and Joeran Beel, “CryptSubmit: Introducing Securely Timestamped Manuscript Submission and Peer Review Feedback using the Blockchain”, ACM/IEEE Joint Conference on Digital Libraries (JCDL) 2017.

[3] Norman Neuschke and Andre Gernandt, “Decentralized Trusted Timestamping using cryptocurrency Bitcoin” iConference 2015, Newport Beach, CA, USA, March 24-27, 2015.

[4] Haber, S. and Stornetta, W.S. 1991. How to Time-Stamp a Digital Document. Advances in Cryptology-Crypto ‘90 Proceedings. 3,2 (1991),99-111.

[5] Rishav Chatterjee and Rajdeep Yadav, “An Overview of the Emerging Technology: Blockchain” 2017 3rd International Conference on Computational Intelligence and Networks (CINE) 2017

[6] IPFS is the Distributed Web: https://ipfs.io/

[7] Decentralized timestamping on the blockchain, available: https://en.wikipedia.org/wiki/Trusted_timestamping

[8] IPFS - Content Addressed, Versioned, P2P File System (White Paper) by Juan Benet.

[9] Blockchain: Challenges and Applications, Pinyaphat Tasatanattakool and Chian Techapanupreeda in 2018 at ICOIN.

[10] Chengjun Cai, Huayi Duan, and Cong Wang, “Tutorial: Building Secure and Trustworthy Blockchain Applications”, IEEE Conference 2018