Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021).
IEEE Xplore Part Number: CFP21ONG-ART; 978-0-7381-1183-4

2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) | 978-1-6654-1960-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICICV50876.2021.9388428

DOC-BLOCK: A Blockchain Based Authentication System for Digital Documents

Iftekher Toufique Imam*, Yamin Arafat†, Kazi Saeed Alam‡ and Shaikh Akib Shahriyar§
Department of Computer Science and Engineering, Khulna University of Engineering & Technology
Khulna, Bangladesh
*imam1607040@stud.kuet.ac.bd, †arafat1607032@stud.kuet.ac.bd, ‡saeed.alam@cse.kuet.ac.bd, §akib.shahriyar@cse.kuet.ac.bd

***Abstract*—With the rapid growth in the sector of information technology and easy access to cheap and advanced office instruments in the market, the faking of important documents has become a matter of concern nowadays. Therefore, the need for verification and authentication practices of various important documents in the form of banking documents, government documents, transaction documents, educational certificates etc is also increasing. However, various challenging and tedious processes have made document verification very complex and time-consuming which motivated us to conduct this research. In this paper, we present a decentralized web application for digital document verification using Ethereum blockchain-based technology in P2P cloud storage to enhance the verification process by making it more open, transparent, and auditable. The proposed model includes several methods such as public/private key cryptography, online storage security, digital signatures, hash, peer-to-peer networks and proof of work which has made the verification of any uploaded documents for any organization or authority faster and convenient with just a click. Furthermore, respective hash values are also assigned to each individual document. Our proposed model successfully meets up all the criteria for a digital document verification system by alleviating the gaps and difficulties in the traditional methods in document verification.**

***Index Terms*—Blockchain, Hashing, Ethereum, Document Verification, Digital Signature, Cryptography**

I. Introduction

The rapid advancement of information sharing and exchanging is driving more and more companies and individual users towards the use of digitized documents. Moreover, the cumbersome and time-consuming use and validation process of traditional physical documents contribute to motivating people to use modern ways of issuing and validating important documents. Though digital documents are undoubtedly convenient to use, proving the authenticity of these documents is often a matter of concern. Due to the technological revolution and ease of access to cheap and advanced equipment, the forgery of important documents has become quite easy and made document authentication quite a tedious task. The implication arising from the problem of fake documentation is causing serious and alarming impacts and needs to be urgently taken into consideration. Therefore, a system to validate the authenticity of important documents would be greatly beneficial to users for maintaining their digital documents. There is an open-source, immutable, and consensus model available called blockchain to solve this problem [3].

Blockchain technology is a recent invention to enhance the document verification process and entangle the task of reducing document fraud and misuse [4]. Blockchain simply refers to a distributed database that chronologically stores multiple blocks chained together with each data pack or block storing documents in a way that makes it impossible to manipulate these documents [8]. Blockchain is an advanced technology that can play many significant roles in the industry to overcome any failure. Blockchain ensures trust, integrity, consensus, autonomy, and safety [13]. Owing to the purely reliable, transparent, and incorruptible method of storing and validating the transactions, we have also been motivated by this blockchain technology to use it in our work to authenticate important digital documents.

At present, the document verification process includes human interpretations and third-party observations. And as we already know, it is a very lengthy process and also there is always a chance of mistakes and dishonesty. So, the current method of verification doesn't seem reliable and efficient. Several kinds of research stated, there are numerous fake documents and certificates surround the global industry [14]. And how it can affect the economy and development of the society. But blockchain technology can eliminate these difficulties and improve security by maintaining full integrity. In our work, we have built a decentralized web application to avoid the unnecessary loss of time to perform the traditional verification process in a more fast and secure way irrespective of time and place with just a single click using the underlying concept of Ethereum blockchain technology. Our web application serves three purposes mainly:

1. Storing the Main Document
2. Verifying any given Document
3. Download any particular Document

In the first process, users can upload any document by using our system. And all the documents uploaded will be stored directly into the blockchain. Then, we will verify any given documents to find out if it is original or corrupted. To verify digital documents, we need the main document and match it with the given. Here we used the SHA-256 hashing mechanism to encrypt and decrypt every document. Moreover, we have also used smart contracts at the backend linking with the blockchain and stored the encrypted hash value of individual documents which will be cross-checked against the given

document. So, any change in the actual document will change the corresponding hash also. If somebody tries to manipulate a document that document will never pass the verification test. Thus, we can get rid of the problem of fake documents. More importantly, if any organization needs to download any document for their purpose, they can easily download it by using the provided IPFS hash given by our system.

II. Related Works

In recent years, blockchain has become a very popular technology in the industry. Several surveys and research have been conducted to implement blockchain in various sectors. In this portion, we will describe some of these previous works available regarding blockchain.

Leible et al. discussed the possibilities and benefits of blockchain in open science platforms. They described how can we implement blockchain in different sectors, and the contribution of blockchain so far in the industry, etc [1]. As blockchain is gaining popularity worldwide because of its distributed and decentralized nature, Joshi et al. completed a survey focusing on the basic challenges and opportunities in blockchain technology and also its security and privacy concerns are described [2]. The fundamental concept and structure of blockchain are shown very briefly from the beginning. The authors also tried to interpret the use of blockchain in IoT, defense, security, and medical sectors. Chen et al. survey discussed different types of areas where blockchain could bring a better solution. Such as cryptocurrency, healthcare, insurance policy, copyright protection, credit transfer, etc [3]. Gilani et al. did a comprehensive survey on blockchain-based identity management and personal data storage system [4]. They discussed a self-sovereign identity (SSI) concept for the users which is the data ownership control. The survey is all about a user-centric data management system eliminating central authority using blockchain. A survey on using blockchain in intellectual property is summarized by Wang et al. [5]. Rouhani et al. described a brief technical overview of Ethereum blockchain and smart contract [6].

Yue et al. proposed a model for data integrity verification using the blockchain method [7]. They described the flaws in a normal cloud-based verification system that includes third-party owners and made a P2P platform using the blockchain-based Markle tree structure where clients can ensure data integrity. For verification, a random sampling method is used by the authors. And mathematical evaluation of the cost and time propagation for this process is given. Teymourlouei et al. proposed a model for user authentication using blockchain which is more secured than the traditional email and password-based authentication system [8]. The authors described the advantages of using private and public keys for document verification. They also discussed the possibility of the vast use of blockchain in IoT and data tracking, supply chain management, property registration, and protection. Zhu et al. proposed a method for secure credit reporting systems in financial sectors using blockchain technology to build trust among users as blockchain provides the strongest security [9]. Their proposed model covers multidimensional authentication for credit transfer using blockchain, smart contract, and hash function. Arjomandi et al. proposed a document verification method using chipless Radio Frequency Identification (RFID) [10]. Their model scans the documents using chipless RFID and stores the patterns of the individual frequency in the cloud storage. But this model is based on a centralized authority. Musarella et al. proposed digital identity-based encryption using the Ethereum blockchain [11]. Their goal was to develop digital identities for Ethereum transactions using smart contracts. Lakmal et al. proposed a document verification model using a digital signature [12]. They transfer any document which needs to be verified into a machine-readable JSON file and sends it to the validators. Validators verify and add a digital signature to the documents. They also provided another feature that tests them and gives a score based on authenticity. But their concept involves third-party people in the process which makes this system vulnerable to hackers at some point. HamithaNasrin et al. discussed the different sets of ideas that are suitable for using blockchain technology and provided a detailed survey about using blockchain for degree verification in Ethereum smart contract [13]. They also described the mining process of the blockchain and mining algorithms ex. proof of work, proof of stake, and proof of importance. Ghazali et al. described the importance of a graduation certificate and how it can be falsified by anyone easily [14]. As it can lead our society in danger, they proposed a theoretical model to verify academic certificates using blockchain technology. Their model includes encryption with private and public keys and digital signature with timestamping for digital certificate verification. Shah et al. described a system to issue the birth certificate and verify the original birth certificate using blockchain [15]. Their proposed model uses RSA and AES keys both for user registration, login, data retrieval, and birth certificate verification. Thus, their process removes the normal password, pin-based verification system, making their system more transparent and secure.

III. Architecture and Implementation

A. Blockchain

Blockchain is an advanced technology that is more convenient and secure than the current centralized data storage system. It is a transaction-based data storage network. In blockchain technology, all the information is stored in a decentralized manner by creating a distributed network. A blockchain network consists of many personal computers where each computer works as an individual database connecting to other computers in the network. If we assume the connected personal computers as a node, then blockchain is a network of nodes connected by a Peer to Peer (P2P) communication protocol. Any node in the network cannot single-handedly manipulate the data because all the other nodes have access to actual data. Moreover, each node or block is encrypted with an extremely secure hash algorithm. And every block also stores the hash code of the previous block which makes them connected like a chain of blocks in

the network [2]. Any change in one block will change the hash code automatically that will make the whole transaction invalid. Thus, there remains no centralized authority of the network which makes blockchain more transparent and reliable for storing and accessing the actual data. So, in the present decade, any data stored in a blockchain network is immutable. One block can generally contain:

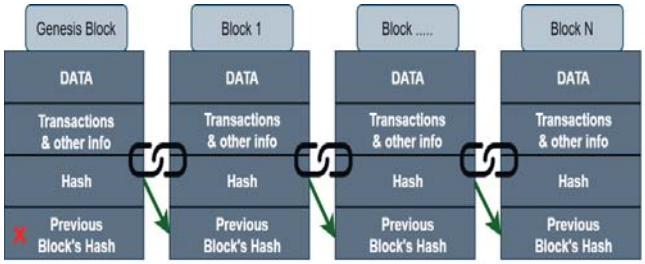

Fig. 1. Blockchain Structure.

1. Its hash value
2. Previous block's hash value
3. Any kind of data or transaction that happened in that block throughout the process.

We can visualize this structure in Figure 1 where the first block is called genesis block, and it has no previous block's hash value [6]. Any transaction on a block is verified by thousands, perhaps millions of computers distributed around the system. Once that block is verified by the other nodes, it can be added to the network.

B. Ethereum

Ethereum is a global, public, and distributed blockchain-based network created for managing the computing system infrastructure of the blockchain network. It is an open-source platform that has multiple functionalities featuring smart contracts, ether, etc. [6]. A smart contract is a programming code deployed in the Ethereum network that executes when a certain event has occurred in any block. Smart contracts are self-executable, distributed, and shared across the blockchain network [11]. Ether is a cryptocurrency for Ethereum based applications. Cryptocurrency means digital money for trading in digital transactions. Ether works as a transaction fee to be paid for any event. So Ethereum provides a basement for developers to create any decentralized application in the blockchain network.

C. Solidity

Solidity is a high-level popular programming language for implementing smart contracts on different blockchain networks like Ethereum. Solidity is an object-oriented programming language. It is inspired by other popular OOP featured programming languages like C++, Python, and JavaScript. Solidity is designed for running the deployed smart contracts in Ethereum Virtual Machine (EVM). Smart contracts are embedded with business logic and computing logic by using solidity language. As solidity supports all the facilities to write smart contracts, it is relatively easy to write smart contracts in solidity.

D. Infura

In our proposed system to run a user's computer as an Ethereum node Infura is used. Generically, to interact with the Ethereum blockchain network users always needed to create an Ethereum wallet first. Without an Ethereum wallet, a user cannot make any transaction or cannot pay the fees for each transaction or in other words cannot use the cryptocurrency "Ether". Infura is a hosted Ethereum node cluster that helps users to interact with any decentralized Ethereum application by avoiding difficulties to set up their own Ethereum wallet.

E. Inter Planetary File System

The InterPlanetary File System (IPFS) stands for the Peer to Peer (P2P) data storage, distribution, and transfer network protocol. IPFS uses a content-based addressing system to independently locate each file linking all computers on a global. Using IPFS in our proposed system will enable a new feature for the users in a similar manner to BitTorrent. So, apparently with this, a user can receive content from any node that has the requested content and also be able to host any content for other users in the network. In the IPFS system, some amount of total data is carried by certain user operators, providing flexible file storage and distribution system. Any network user can host a data file or other information by using its unique content address, then other network users can identify, request or get access to that information from any personal computer that has it.

F. Hashing Function SHA-256

As we can see in Figure 2, the hash value can be understood as a human fingerprint that is unique to each input. Saving the

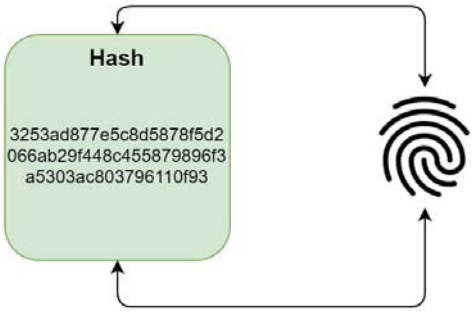

Fig. 2. Hash.

original file in the database would take a huge amount of disk space. So, we need a way to uniquely map files and documents with something smaller than the original size. We will use a hashing function to get this job done A cryptographic hash function can be defined as a complex mathematical function that takes an input data of variable length or size and generates a unique output of fixed length or size for every

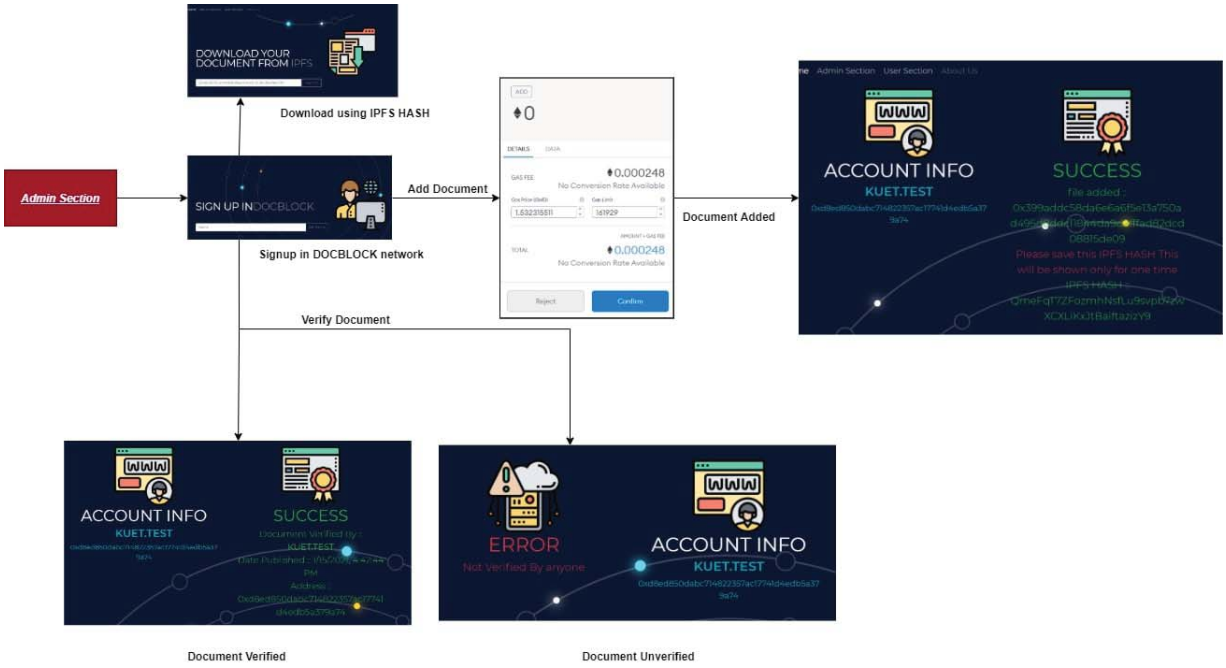

Fig. 3. Admin-Section Workflow.

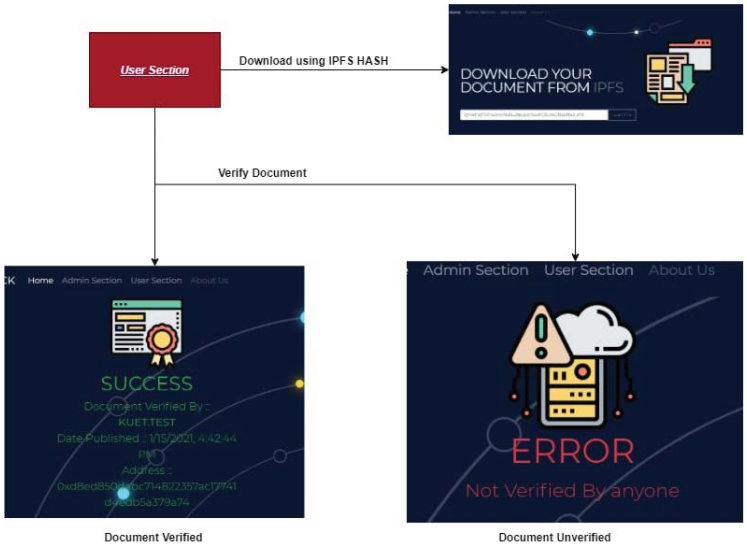

Fig. 4. User-Section Workflow.

given input. Hash always works as a one-way function. So, it is computationally impossible to find the input of a hash function from the hash output. Hash functions are commonly used for digital signatures and cryptography purposes. For example, password security and message verification, key derivation, pseudo-random number generation, and blockchains. Mainly hashing ensures extreme security of any data content. Some of the well-defined hash functions are MD5, SHA-1, SHA-256. In our proposed model. We used the most effective SHA-256 algorithm. SHA-256 can covert large input data to a fixed size 256-bit (32-byte) hash code.

IV. The Proposed Framework

The key concept behind this project is to build a platform that will play a significant role in verifying the authenticity of important files, contracts, certificates, and land/property/asset documents more accurately and quickly. We have used very well-known and secured methods to build this system. We have used Ethereum blockchain to develop this system. SHA-256 is used along with blockchain. It is a one-way and collision-resistant encryption algorithm to encrypt the data.

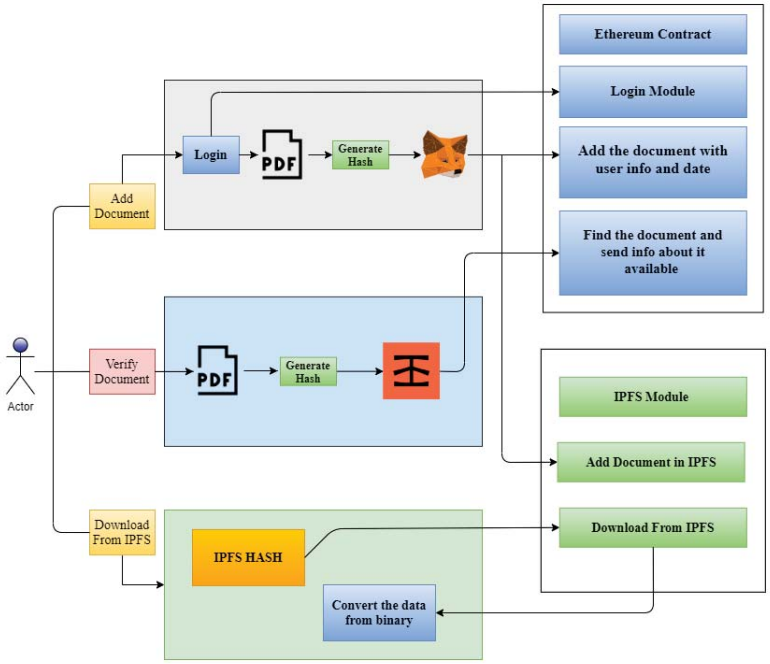

Fig. 5. Proposed DOC-BLOCK System Process.

The entire procedure is very secure that nobody can harm any document. Thus, this project is strongly focused on solid defense against any cyber-attack. Users never need to think about their files being lost or compromised, because they can conveniently access their records anytime they want. Despite having many advanced technologies in the back-end of our system we have designed the front-end part with HTML and JavaScript. Our website's user interface is designed manually keeping the concern in the head of behaving user friendly. For any professional work or something, this website can easily be run by a novice person who does not have any understanding of Ethereum or blockchain. The website does not have a lot of content or have any useless info or unexpected behavior. It is very simple and decent to use for anybody. Our implementation was deployed in Heroku. In our developed system, there are 2 sections (User and Admin) shown in Figure 3 and 4.

The system has 3 important features: upload, verify and download. The whole back-end process is shown in Figure 5. To maintain usability for every scenario of use case, there are mainly two separate sections, one for the organization and the other for general purpose. The admin section refers to any organization or institution. Sometimes an organization or institution needs to check cautiously any documents submitted to them before closing a deal otherwise they will never be sure about the document's authenticity. For any organization's purpose, an organization needs to have Metamask installed in order to upload a verified document in the blockchain. After each upload, the hash of the file is attached with the public key of the organization and the date of adding the document which is used to further verify the authenticity of the document. The organization will also receive an IPFS hash, which only they can see and share with a particular user for further downloading the document. The top portion in Figure 5 of the workflow picture is showing this process of adding files/documents to the Ethereum network. If the same file is added multiple times only the first user who added it, will be shown as result. For example, we added a file and it can be downloaded from Download Document from any section using this IPFS HASH: *QmbfhzU8akbENKJwp3c8eW2vw19fLb8vUYzfjtHKeUsBbN*. This file was added in 1/13/2021, 7:50:29 PM, with the owner name Nuhash with public address: *0xb2863a36f3776c5c7323efe2b3235ce8a3811460*. The cost of adding a document to the contract is 0.04787USD/4.04BDT. So, when an organization uploads any document in our developed system, the system adds the document in blockchain and provides other information about the added document on the screen. For example, the uploader's name and public address, upload time, and an IPFS hash for that particular file are given for further verification. And if the document gets corrupted somehow the authority will know about it because of the change in the document's hash value. On another note, for general-purpose, any user/verifier can verify the given document or download a document with an IPFS hash without having any access to Ethereum/IPFS, which is processed on the server-side for increase usability. Our file SHA256 hash was

3253ad877e5c8d5878f5d2066ab29f448c455879896f3a5303ac803796110f93, this is one way hashing method and we can't get the main file from this hash. Then we made a simple edit in our document and tried to verify it. As expected as even for a small change, the hash (*828f4d885afdecf4ce373aa4ff27ea925a63c07f0a65f8fb6dc6be6b589444aa*) was changed and this file isn't verified. The probability of collision of two hash is, $P \approx \frac{1}{2}(n/2^{128})^2$ where n is the length of the hash which is 256 length, the collision probability is really low that we didn't have to add a double hash method or 512-bit hashing method to make it more secure.

If any genuine copy of file/documents is needed for any query, users of our system will be able to that in a moment of time. After adding a document, the user will be given an IPFS hash, it's given for one time and the user needs to save it somewhere for further usage. We didn't store the IPFS hash in smart contract for making the system more private. When a user gives the IPFS hash into our system, it searches the corresponding file matched with this hash and sends back the original file converted from binary code. The bottom portion in Figure 5 of the workflow picture shows this process. As IPFS automatically clears fewer downloaded files overtime, this file may be inaccessible.

For any individual user or organization, or institution we can see that by following our proposed model we can easily verify any documents and be sure about the document's authenticity and also be able to download the original file always for further checking to find any fake document. And the process to do this entire thing is very simple and easy. So, there is no chance of any unexpected error. Any user with less knowledge about blockchain, Ethereum will never face any problem using this system.

V. Conclusion & Future Works

To avoid document forgery and misuse a better solution was needed for a long time. Therefore, we proposed a model to solve this global problem. The main purpose of our developed system is to create a platform to store and verify any important documents like certificates, land/property/asset records, medical records, etc. We implemented the whole system using Ethereum blockchain network. The collaboration of some well-known features like cryptographic hash, decentralization, and digital signature makes blockchain technology immutable. As a result, there remains no central server to own the data rather all the information regarding any transactions is distributed to the whole network. Our proposed system stands strongly based on security as any manipulation in the documents is quite impossible. The verification result is always accurate and efficient. After comparing our system with the general cloud-based data storage system and verification process we found significant progress in both security enhancement and time optimization. And using our proposed model data corruption and misuse will highly be reduced. Any company, organization, and institution can use this system for better security. In conclusion, our proposed model ensures integrity and security for every use case.

However, as a new and developing technology blockchain has some minor complexities to use in every platform. But still, blockchain technology outperforms any current system application available in the industry by a big margin in security and reliability. Despite all of that our future plan with this model is to create a terminal-based document authentication with the support of multiple file upload and other accessibility features to increase usability for better performance.

References

[1] S. Leible, S. Schlager, M. Schubotz, and B Gipp, "A Review on Blockchain Technology and Blockchain Projects Fostering Open Science," (2019), Front. Blockchain 2:16. doi: 10.3389/fbloc.2019.00016.

[2] A. Prashanth Joshi, M. Han, and Y. Wang, "A Survey on Security and Privacy Issues of Blockchain Technology," (2018), Mathematical Foundations of Computing, Volume 1, Issue 2, pp. 121-147, doi: 10.3934/mfc.2018007.

[3] W. Chen, Z. Xu, S. Shi, Y. Zhao, and J. Zhao, "A Survey of Blockchain Applications in Different Domains," (2018), pp. 17-21, doi: https://doi.org/10.1145/3301403.3301407.

[4] K. Gilani, E. Bertin, J. Hatin and N. Crespi, "A Survey on Blockchain-based Identity Management and Decentralized Privacy for Personal Data," 2020 2nd Conference on Blockchain Research and Applications for Innovative Networks and Services (BRAINS), Paris, France, 2020, pp. 97-101, doi: 10.1109/BRAINS49436.2020.9223312.

[5] J. Wang, S. Wang, G. Junqi, Y. Du, S. Cheng, and X. Li, "A Summary of Research on Blockchain in the Field of Intellectual Property," (2019), Procedia Computer Science, Volume 147, pp. 191-197, doi: https://doi.org/10.1016/j.procs.2019.01.220

[6] S. Rouhani and R. Deters, "Security, Performance, and Applications of Smart Contracts: A Systematic Survey," in IEEE Access, vol. 7, pp. 50759-50779, 2019, doi: 10.1109/ACCESS.2019.2911031.

[7] D. Yue, R. Li, Y. Zhang, W. Tian and C. Peng, "Blockchain Based Data Integrity Verification in P2P Cloud Storage," 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS), Singapore, Singapore, 2018, pp. 561-568, doi: 10.1109/PADSW.2018.8644863.

[8] H. Teymourlouei and L. Jackson, "Blockchain: Enhance the Authentication and Verification of the Identity of a User to Prevent Data Breaches and Security Intrusions," (2019).

[9] X. Zhu, "Blockchain-Based Identity Authentication and Intelligent Credit Reporting," (2020), Journal of Physics: Conference Series, volume 1437, 012086, doi: 10.1088/1742-6596/1437/1/012086.

[10] L. M. Arjomandi, G. Khadka, Z. Xiong and N. C. Karmakar, "Document Verification: A Cloud-Based Computing Pattern Recognition Approach to Chipless RFID," in IEEE Access, vol. 6, pp. 78007-78015, 2018, doi: 10.1109/ACCESS.2018.2884651.

[11] L. Musarella, F. Buccafurri, G. Lax, and A. Russo, "Ethereum Transaction and Smart Contracts among Secure Identities," (2019).

[12] C. Lakmal, S. Dangalla, C. Herath, C. Wickramarathna, G. Dias and S. Fernando, "IDStack — The common protocol for document verification built on digital signatures," 2017 National Information Technology Conference (NITC), Colombo, 2017, pp. 96-99, doi: 10.1109/NITC.2017.8285654.

[13] M. HamithaNasrin, S. Hemalakshmi, and Prof G. Ramsundar, "A Review on Implementation Techniques of Blockchain enabled Smart Contract for Document Verification," International Research Journal of Engineering and Technology (IRJET), Volume 6, Issue 2, 81, February 2019.

[14] O. Ghazali, and O. Saleh, "A Graduation Certificate Verification Model via Utilization of the Blockchain Technology," (2018), Journal of Telecommunication, Electronic and Computer Engineering, 10, pp. 29-34.

[15] M. Shah and Dr. Priyanka Kumar, "Tamper Proof Birth Certificate using Blockchain Technology", International Journal of Recent Technology and Engineering (IJRTE), Volume 7, Issue 5S3, pp. 95-98, February 2019.