Design a Document Verification System Based on Blockchain Technology

Muhammad Dhiyaul Rakin Zainuddin and Kan Yeep Choo(✉)

Faculty of Engineering, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya,
Selangor, Malaysia
kychoo@mmu.edu.my

Abstract. Document forgery is a common method that has been used by many people for their own benefits. Document often has its own unique identity which contains a very sensitive content. The current problem related to document is about the document forgery where advanced technology capable in duplicating and modifying a document. The current document verification system often checks for its availability only and does not check for its integrity which is its content. Moreover, current practices provide a low efficiency in detecting a forged document which resulting in false results. Therefore, blockchain will be introduced where it needs to be used to create a new document verification system while integrating with Interplanetary File System (IPFS) to increase the efficiency in detecting a forged document.

Keywords: Blockchain · IPFS · encryption · document verification

1 Introduction

Blockchain technology is a system that acts as a decentralised database which is a digital ledger that stores data and information throughout the entire network of a computer systems. Blockchain technology were introduced in 2006 with the existents of Bitcoin where it involves cryptocurrency. A cryptocurrency is a digitized currency that can be used to do a transaction and buy stuffs through online. With the help of blockchain technology all the transactions that have been made in the network are being kept safe since the system uses a distributed ledger that will let others know the credibility of a transactions. Blockchain technology had been used by many organizations now such as healthcare, inventory, management, finance and more. Blockchain technology has been helpful in implementing a system that will keep the transactions and storing the data in a safe manner and it is easier to track down the attacker with the implementation of blockchain technology. A document verification system will be able to increase its security feature with the help of blockchain technology.

1.1 Problem Statements

Technology has been evolved throughout the years where people nowadays depend on technology to carry out their daily lives. The current problem that relates to a document

M. Y. bin Alias (Ed.): MECON 2022, AER 214, pp. 229–244, 2023.
https://doi.org/10.2991/978-94-6463-082-4_23

verification system is to prevent the document from being forged. Document forgery involves in copying and imitating the details of the original documents such as identity number and signature.

Other than that, there is a problem with the current system in detecting a forged file where it only checks for its availability of the file in the system. It does not go through content of the file to check for its integrity. There are not many systems that will go through the content of a document and verify each character in the document.

Furthermore, the current document verification system has a low efficiency in detecting the forged file. Current practices of document verification are not efficient and time consuming in getting the results. The verification of the file integrity is not accurate where sometimes it gives false results.

1.2 Objectives

The objective of the system is to investigate the limitation of Ethereum blockchain technology and Interplanetary File System (IPFS) in developing a document verification system and to provide a system that will verify the integrity of a document and a place that will keep a document safe from any forgery to happen.

2 Literature Review

Fraud often happens when people tend to want to get anything they want through illegal method. Document forgery is one of the frauds that people made where it involves in copying and imitating the details of the original documents such as identity number and signature.There are different types of document forgery such as Print, Copy and Paste (PPC), imitation, Reversed Engineered Imitation (REI), Scan, Edit and Print (SEP) forgery [1] and more. Each of the forgery main target was being able to produce an unofficial document that will trick the system or people to gain the things that they want such as money laundering and travel to a country illegally.

A document verification system must be made with a high efficiency to prevent the problems that has been mentioned before. Bhavani Thuraisingham [2] stated that blockchain technologies have become the next technologies that will bring data science to the next level where it secures a lot of things which related to data science such as data processing, data management, data sharing, data collection and data analytics. Blockchain becomes a powerful tool where it also verifies the integrity of the data when there is a transaction within the blockchain.Therefore, blockchain is suitable for storing the documents safely.

Next, the case study of the system uses image capture to store and verify the document using blockchain was based on certificates of a student [3]. The system will convert the captured image of the certificate to a digital certificate uploaded by the admin of the system by using sampling and quantization method. The verification of the certificate was done by comparing the hash value of the original certificate in the blockchain and the hash of the uploaded certificate. A problem arise in the system is the accuracy of converting the analogue image to digital image where after the conversion of the digital image, the hash value may be different from the hash value of the original certificate that have been uploaded into the blockchain since it involves the sampling and quantization method where it will change the nature of the certificate.

Barbara Guidi, Andrea Michienzi, and Laura Ricci [4] had explained one system that can be used to check for the integrity of the file which uses Interplanetary File System (IPFS). IPFS uses the content of the data to locate its address which uses the hash of the of the data instead of using a domain name.IPFS has provided a feature where the data will be always available in the network without automatically removed which is called pinning services where it run long-lived host nodes on a cloud service provider.

Based on the previous research, IPFS can be a useful tool to store a large file and can be used to check the content and integrity of the files. Since the integrity of the files is important in a document verification process, BlockIPFS has been proposed where it uses Interplanetary File System (IPFS) to hash the content of the files uploaded to the IPFS [5] and the blockchain will store only the hash created by IPFS. BlockIPFS uses the concept of IPFS where rather than storing the files into the blockchain, the files will be stored in the IPFS, and the hash pointed to the files will be stored into the blockchain. BlockIPFS only serve the purpose of checking the performance of storing the files into the IPFS and does not serve the purpose of verifying the document but the idea of using IPFS will be useful in verifying the content of the file.Other than that, BlockIPFS uses a public network of IPFS where the files uploaded to the IPFS will be publicly known. The limitation of using IPFS public network is the user will have the access to the files if they have the hash of the files.

Therefore, another literature has been reviewed where it solves the problems of using the IPFS public network. The downside of using a public IPFS network is all the data stored in the IPFS will know and spread publicly to the entire network where all the peers inside the public network will have the access to the data. Based on a conference paper titled as ‘Academic Storage Cluster’, the authors had introduced and used a private IPFS network which uses IPFS Swarm and IPFS Cluster [6]. The reason behind in creating a private IPFS network to limit the access of the data inside the IPFS to a certain number of peers. IPFS Swarm and IPFS Cluster can be used to keep the data safe from the people outside the network to have the access to the data which is suitable for the document verification system.

2.1 Background Theory

2.1.1 Blockchain

Blockchain is decentralized immutable ledger which monitor the transactions and tracking assets which can be a tangible or intangible assets in the network [7]. It is a Distributed Ledger Technology (DLT) which record all the transactions using a cryptographic system which is a hash. Blockchain consists of blocks which are tied up to each other with a chain inside the network. Each block consists of information which are the data, previous and current hash value, and the timestamped of the transactions occurred that are strictly known to the people in the network who made the transaction. The transactions in the blockchain will be completed when all the peers in the blockchain network approve and accept the data that have been stored into the blockchain which means the approval will successfully store the data into the blockchain.

2.1.2 Interplanetary File System (IPFS)

Interplanetary File System (IPFS) is a decentralized peer-to-peer system where it stores and access files, applications, data, and websites [8]. It is a distributed file storage which uses content-based addressing rather than location-based addressing. IPFS uses content-based addressing where the data or the file that need to be access will be based on the content of the files which uses a content identifier. Content identifier uses a cryptographic hash where it will give a unique hash value of the file that have been transfer to the IPFS and the unique hash value will be based the content of the files. The unique hash value is a unique identification for a file where the hash value is only for that file. All the files that have been transferred to IPFS network will be store in a Distributed Hash Tables (DHT). The benefits of DHT in IPFS are based on its scalability and fault tolerance where the network still functioning even though one of the nodes in the network are offline or leave the network. The feature that makes IPFS is a suitable and secure place for a file storage is that whenever there is some changes or modifications to the files, the unique hash value will also be changed.

3 Details of the Design

3.1 Overall System Design

Document verification system using blockchain consists of many software designs which include Ethereum Blockchain, Truffle Suite, IPFS Cluster, AES encryption and web application. IPFS Cluster will be the storage for the file uploaded to the system where it has its own feature. Ethereum blockchain will be the blockchain platform to store the IPFS hash value from the IPFS Cluster which points to the file that has been stored in the IPFS. Truffle Suite is the platform that is used to develop an application uses the test Ethereum network without using any computational power and resources. Finally, web application will be used as the interface for the document verification system where the user will upload and verify a document. Figure 1 shows the diagram of overall system design. Section 3.3 will discuss in detail about the process that occurs in the system.

3.1.1 Ethereum Blockchain

Blockchain system that have been used in designing the document verification system is Ethereum blockchain. Ethereum is an open source blockchain-based platform that

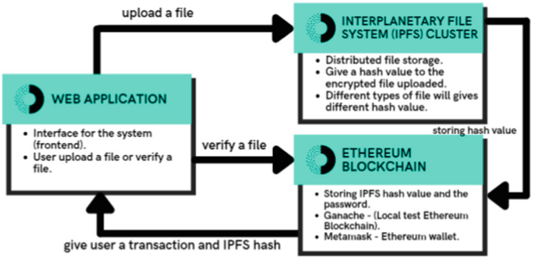

upload a file
INTERPLANETARY FILE SYSTEM (IPFS) CLUSTER
• Distributed file storage.
• Give a hash value to the encrypted file uploaded.
• Different types of file will gives different hash value.
WEB APPLICATION
• Interface for the system (frontend).
• User upload a file or verify a file.
verify a file
storing hash value
ETHEREUM BLOCKCHAIN
• Storing IPFS hash value and the password.
• Ganache - (Local test Ethereum Blockchain).
• Metamask - Ethereum wallet.
give user a transaction and IPFS hash

Fig. 1. Overall System Design

uses ETH as its cryptocurrency to be used for transactions.For the document verification system, Ethereum blockchain will become a place to store the identity of the file which is the IPFS hash value instead of the file because it will cost a less of computational power and resources to store a small credentials. Large files require a higher computational power which cost higher gas price to store the large files into the Ethereum blockchain. Therefore, IPFS will be use as a decentralized storage for the file and the credentials that points to the file which is the IPFS hash value will be stored into the block of the Ethereum blockchain.

3.1.2 Truffle Suite

Truffle suite is a development environment to use Ethereum blockchain to make transactions and smart contracts before deploying the actual Ethereum blockchain into the decentralized application. Truffle suite consists of three components which are Truffle, Ganache and Drizzle. Truffle is a development environment where it will use a test framework to connect between the Ethereum blockchain used by Ganache with the decentralized application. Ganache is a personal Ethereum blockchain that installed locally into a computer for a decentralized application development purpose. It is a useful tool where Ganache is being used as the test Ethereum network for the decentralized application built before deploying the real Ethereum network to the application in a safe environment. The document verification system will be using Ganache as the Ethereum blockchain network as there is a constraint in the computational power.

3.1.3 IPFS Cluster

Interplanetary File System (IPFS) is a storage file system that works in a public network in nature. In a document verification system, security of the file must be considered in a high priority as it may contains sensitive information.A file storage must be known only to the selected peers in the organizations to keep it safe and secure from the unknown people. IPFS Cluster introduced a way where sharing content of the file in the IPFS to be secure where the files will only be shared to the certain number of peers depends on the user. IPFS Cluster provides data coordination across a swarm of IPFS daemons by replicating, allocating, and tracking a global pin set distributed among multiple peers on the network. IPFS Cluster acts as a private network where only selected peers can share and view all the files uploaded into the IPFS. It is a separate system from IPFS where it is a standalone application which uses the IPFS daemon's API. There is no centralized hosting in the cluster where every peer in the network can pin a file into the IPFS Cluster.

3.1.4 AES Encryption Algorithm

There are different types of encryption algorithm in a symmetric encryption. The symmetric algorithm encryption that will be used to encrypt the file is from the modern symmetric encryption which is called as AES encryption. AES encryption is one of the block ciphers that operates on 128-bit blocks. There are three phases in the process of encrypting using AES algorithm which are initial round, rounds, and final round. Each of the phases uses different types of algorithms to encrypt the data. AES algorithm uses only a single key to encrypt and decrypt the data. AES will make the encryption process to be robust from hacking because of its length in key sizes. For document verification system, the file will be encrypted with a password which uses AES algorithm whereby if there is a user who wants to read the content of the file, the user needs to enter the password or the key to decrypt the file.

3.1.5 Web Application

User interface must be created to make the document verification system to be user friendly. A web application will be made to create the user interface for the document verification system using blockchain. Web application is created as the frontend of the system where the user will interact with the system to upload and verify a document or file and the Ethereum blockchain will be the backend of the system. The web application created will be a decentralized application as it is using Ethereum blockchain as its backend. It is developed by using HTML, JavaScript, and CSS which uses a lot of external packages and library to make the application works.

3.2 System Design Implementation

3.2.1 Uploading Process

The purpose of a document verification system is to verify the availability and integrity of the file. Before the verifying process, the file must be uploaded first into the system to indicate that the submission of the file is the original file, and it must be verify based on the existence and content of the original file. The requirements for the user to upload the file to the system are to ensure that it is in a pdf format and being encrypted using AES algorithm which uses password before using the system. The encryption process can be done through Microsoft product and file explorer where the requirement to use it is using Microsoft version 2007 or newer.

The process of uploading the file to the system starts by submitting the password that is used to encrypt the file together with the encrypted file into the system. First, the system will take the submitted encrypted file to be uploaded into the IPFS Cluster which will go through the IPFS daemon. After the file has been uploaded to the IPFS, IPFS will an identity to the file which will be used to access the file called as the IPFS hash value. The IPFS hash value together with the password that has been submitted earlier will be transferred and stored into the Ethereum blockchain where it is stored in a form of a block. Before the storing process of IPFS hash value and the password to the file, the system will ask the user to confirm the transactions made by paying transaction fees using ethers. After the confirmation of the transactions, blockchain will store the IPFS hash value together with the password of the encrypted file. The uploading processes can be design using procedural design which will be described in a flow chart. Figure 2 shows the flow chart design.

3.2.2 Verification Process

Verification process can be done when the file has gone through the uploading process. The requirements that it needs to perform the verification process are the file that wanted

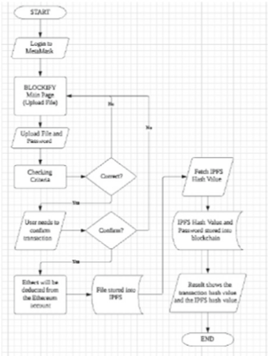

START
Login to MetaMask
BLOCKIFY Main Page (Upload File)
Upload File and Password
Checking Criteria
Correct?
No
No
Yes
User needs to confirm transaction
Confirm?
Yes
Ether will be deducted from the Ethereum account
File stored into IPFS
Fetch IPFS Hash Value
IPFS Hash Value and Password stored into blockchain
Result shows the transaction hash value and the IPFS hash value
END

Fig. 2. Flowchart of Uploading Process

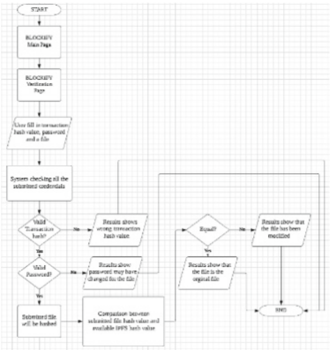

START
BLOCKIFY Main Page
BLOCKIFY Verification Page
User fill in transaction hash value, password and a file
System checking all the submitted credentials
Valid Transaction hash?
No
Results shows wrong transaction hash value
Yes
Valid Password?
No
Results show password may have changed for the file
Yes
Submitted file will be hashed
Comparison between submitted file hash value and available IPFS hash value
Equal?
No
Results show that the file has been modified
Yes
Results show that the file is the original file
END

Fig. 3. Flowchart of Verification Process

to be verified, the password to the encrypted file and the transaction hash of the file when uploading into the system. These three components must be entered into the application to perform the verification process. When these three components have been entered into the system, the verification process starts by using the transaction hash to fetch the transaction block that has the same transaction hash in the blockchain. Then, the system will decode the metadata inside the transaction block to obtain the IPFS hash value and the password of the file. After that, the submitted file will be hashed, and the verification process of the file is by comparing the hash value of the submitted file with the hash value that has been fetched from the transaction block in the blockchain. Then, the submitted password will be compared with the password that has been stored in the blockchain. Since there are three different components that is needed for the system to verify the file, there will be a total of four scenarios where if each of the components entered to the system is incorrect or verified. Figure 3 shows the flow chart design.

3.2.3 History Log Process

This is an extra feature of the system where the user can check the history log of the transactions made in the system. The system will fetch the transaction blocks from the blockchain to be displayed from the application for the user to monitor. It will fetch the latest ten transactions made from the system and it will decode the metadata data from the blockchain to be displayed to the application.

4 Data Presentation and Discussion

4.1 System's Requirements

4.1.1 IPFS

IPFS is a system that can be downloaded and launch from the command line through the command prompt. IPFS need to be available during uploading process where IPFS Cluster need to use the API of IPFS daemon to store the files into the IPFS Cluster. IPFS is an important component to ensure the working mechanism of IPFS Cluster where IPFS Cluster can be used to share files among the selected peers. To initialize the IPFS daemon for the system, the command will be entered through the command prompt where it will start the IPFS daemon.

4.1.2 IPFS Cluster

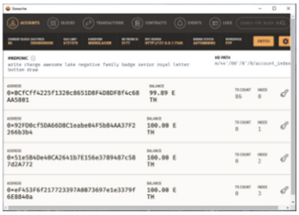
IPFS Cluster is a standalone application that will ensure the files to be uploaded to the IPFS and shared to the selected peers in the network. IPFS Cluster need to use the IPFS daemon API to store the files successfully. After IPFS daemon has been initialized, it is ready to be used by IPFS Cluster to start the clustering process of the peers in the network to store and share the files using the IPFS daemon API. IPFS Cluster can be initialized by using the same method as IPFS daemon where it uses the command line to start the cluster. The command line to initialize the IPFS Cluster is 'ipfs-cluster-service daemon', then it will start to fetch the IPFS daemon's API and start the clustering process.

4.1.3 Ethereum Blockchain

Ethereum blockchain is a must system that is needed to available to verify the existence and integrity of the files stored in the IPFS. Ethereum blockchain is a system that will cost a lot of computational power to use all the feature provided by Ethereum. For this document verification system, Truffle Suite has been used where it provides the Ethereum test network that can be used for development purposes. One of the Truffle Suite which is Ganache will be the one that need to be available as it will automatically connecting the Truffle and Drizzle to the document verification system. Ganache is an application that has its own user interface which can start up on its own and the user will need to choose the network that has the account that is connected to the document verification system. Figure 4 shows the user interface of Ganache after choosing the network.

4.2 File Encryption Process

The first phase of the process of the using the document verification system is by going through the file encryption process where the file will be encrypted using AES algorithm.

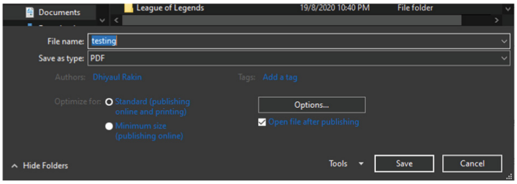


MNEMONIC
write change awesome lake negative family badge senior royal letter bottom draw

HD PATH
m/44'/60'/0'/0/account_index

ADDRESS	BALANCE	TX COUNT	INDEX
0×BCfCff4225f1320c8651D8F4D8DF8f4c68AA5801	99.89 ETH	86	0
0×92FD0cf5DA66D8C1eabe04F5b84AA37F2266b3b4	100.00 ETH	0	1
0×51e584De40CA2641b7E156e3789487c587d2A772	100.00 ETH	0	2
0×eF453F6f217723397A0873697e1e3379f6E8840a	100.00 ETH	0	3

Fig. 4. User Interface of Ganache



File name: testing

Save as type: PDF

Authors: Dhiyaul Rakin

Tags: Add a tag

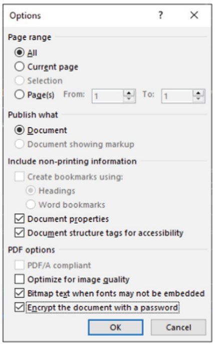
Optimize for: ● Standard (publishing online and printing)

○ Minimum size (publishing online)

Options...

☑ Open file after publishing

Fig. 5. Saving as PDF format



Page range

● All

○ Current page

○ Selection

○ Page(s) From: 1 To: 1

Publish what

● Document

○ Document showing markup

Include non-printing information

□ Create bookmarks using:

● Headings

○ Word bookmarks

☑ Document properties

☑ Document structure tags for accessibility

PDF options

□ PDF/A compliant

□ Optimize for image quality

☑ Bitmap text when fonts may not be embedded

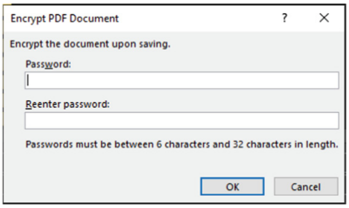
☑ Encrypt the document with a password

Fig. 6. Choose encrypting file option

The encryption process will be done through the file explorer where the user will need to save the file as the PDF format. The first step of the encryption process is to open any file using Microsoft office such as Word, Excel, or PowerPoint. The requirement for the user to encrypt the file is to use Microsoft product from 2007 version or newer. After the content of the file has been fixed, click on the save as option and browse for a place to store the file. Then, select the 'save as type' option to a PDF format and click on the 'Options…' button to see more option about the PDF format (Fig. 5).

After that, tick the box that has the caption of 'Encrypt the document with a password' to start the file encryption process using PDF format (Fig. 6).

The encryption process provide by the file explorer is using the AES algorithm where the file will be encrypted with a single password made by the user itself. The next step would be the user need to create the password for the file to be encrypted successfully.



Encrypt PDF Document

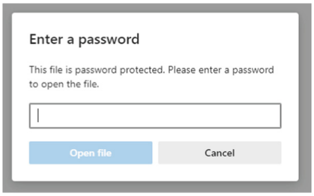
Encrypt the document upon saving.

Password:

Reenter password:

Passwords must be between 6 characters and 32 characters in length.

Fig. 7. Creating password for encrypted file



Enter a password

This file is password protected. Please enter a password to open the file.

Fig. 8. Accessing encrypted file

The password that needs to be entered must have at least 6 characters and at most of 32 characters in length (Fig. 7).

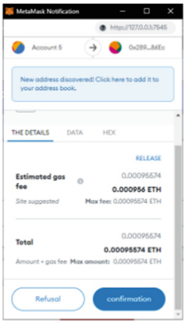
After the password has been entered, the file has been encrypted with the password created by the user and stored in the place that has been chosen by the user. Anyone who wants to access or open the file need to enter the password to that file to decrypt the file and has the access to the content of the file. The purpose of the file encryption process is to restrict the access of the file to the authorised user only and prevents from the file to be forged by an unauthorized user. Figure 8 shows the interface when someone wants to access the open the file.

4.3 Uploading Process

File uploading process can only be done if the user has gone through the file encryption process. In this process, there are two components that need to be prepared which are the encrypted file and the password of the file where it will be submitted and stored into the system.

After all the required credentials has been submitted, MetaMask will prompt a message to confirm the transaction. The confirmation of the transaction will cost a few ethers which will subtracted from the account in the Ganache. The uploading process will start after the confirmation of the transaction has been made. Figure 9 shows the prompt message from MetaMask to confirm the transaction.

The encrypted pdf file will be uploaded to the IPFS Cluster where it will become the storage for the files uploaded to the system. After the file has been uploaded to IPFS Cluster, the file will be given an identity in a form of IPFS hash value which can be used to access the files. Then, the IPFS hash value and the password to the encrypted file will be stored into the blockchain which can be seen in the Ganache application. After that, the result will show the IPFS hash value of the file uploaded and the transaction hash of

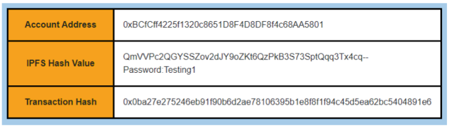


New address discovered! Click here to add it to your address book.

Estimated gas fee 0.00095574
0.000956 ETH
Site suggested Max fee: 0.00095574 ETH

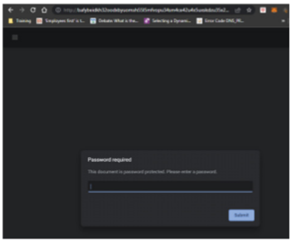
Total 0.00095574
0.00095574 ETH
Amount + gas fee Max amount: 0.00095574 ETH

Fig. 9. Confirmation of transaction



Account Address	0xBCfCff4225f1320c8651D8F4D8DF8f4c68AA5801
IPFS Hash Value	QmVVPc2QGYSSZov2dJY9oZKt6QzPkB3S73SptQqq3Tx4cq-- Password:Testing1
Transaction Hash	0x0ba27e275246eb91f90b6d2ae78106395b1e8f8f1f94c45d5ea62bc5404891e6

Fig. 10. Upload Result



Password required

This document is password protected. Please enter a password.

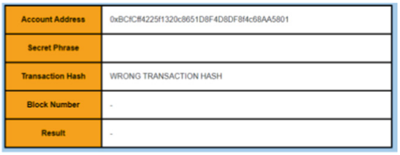
Fig. 11. Accessing file using IPFS link

the transaction block which need to be kept and will be used for the verification process (Fig. 10).

The user can access the files by using the IPFS hash value through a link which is ‘http//localhost:8080/ipfs/{IPFS hash value}’. Even though the user can access to the files by using the link and IPFS hash value, the user still needs to use the password to the encrypted file to have the access to the content of the file. This will increase the security of the system where the password is needed to decrypt the file and see the content of the files. Figure 11 shows the interface when accessing the files using IPFS hash value.

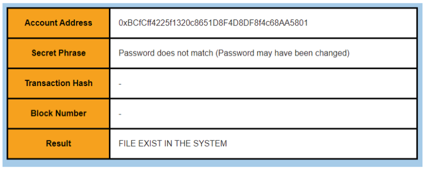
4.4 Verification Process

Verification process can be done after the file has been successfully uploaded into the system. There are three components that are needed to verify a document which are the password of the encrypted file, the transaction hash that is related to the file, and the file that need to be verified.



Account Address	0xBCfCff4225f1320c8651D8F4D8DF8f4c68AA5801
Secret Phrase	
Transaction Hash	WRONG TRANSACTION HASH
Block Number	-
Result	-

Fig. 12. Wrong transaction hash



Account Address	0xBCfCff4225f1320c8651D8F4D8DF8f4c68AA5801
Secret Phrase	Password does not match (Password may have been changed)
Transaction Hash	-
Block Number	-
Result	FILE EXIST IN THE SYSTEM

Fig. 13. Wrong password

4.4.1 Wrong Transaction Hash Value

The first scenario is when the submitted transaction hash does not match to any of the transaction hash inside the blockchain. The function of submitting the transaction is that it is being used to fetch the specific transaction block which stores the desired IPFS hash value and the password top the encrypted file. When the submitted transaction hash does not match to any of the transaction hash in the blockchain, there is no metadata that can be fetch from the transaction block and the verification process will be stop as there is no data to compare with the other submitted components. Figure 12 shows the result when the transaction hash does not match to any of the transaction hash from the blockchain.

4.4.2 Wrong Password of the Encrypted File

After the transaction hash is being verified, the next thing to verify is the password of the encrypted file. The second scenario is when the submitted password of the encrypted file is not match to with the password from the transaction block. The mechanism of verifying the password submitted to the system is by comparing it with the password that has been stored into the transaction block with the submitted transaction hash. If the submitted password does not match with the password inside the blockchain, there are two possibilities which lead to the unsuccessful verification of the password of encrypted file which are either the password of the uploaded file has been modified or the submitted password is entirely wrong. Figure 13 shows the result when the submitted password does not match with the password stored inside the blockchain.

4.4.3 Wrong IPFS Hash Value

When both the submitted transaction hash and password has been verified, the next step will be to verify the file using the IPFS hash value. The system will fetch the IPFS hash value that has been stored in the blockchain and the submitted file will be hashed and both IPFS hash value will be compared to verify the file. If the IPFS hash value does not match with each other, then the file is either has been modified or the user has submitted

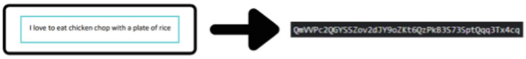

I love to eat chicken chop with a plate of rice
QmVVPc2QGYSSZov2dJY9oZKt6QzPkB3S73SptQqq3Tx4cq

Fig. 14. Original file with its IPFS hash value

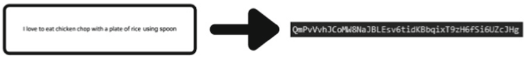

I love to eat chicken chop with a plate of rice using spoon
QmPvVvhJCoMW8NaJBLEsv6tidKBbqixT9zH6fSi6UZcJHg

Fig. 15. Modified file with its IPFS hash value

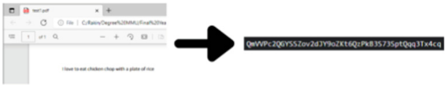

test1.pdf
File | C:/Rakin/Degree%20MMU/Final%20Yea
1
of 1
I love to eat chicken chop with a plate of rice
QmVVPc2QGYSSZov2dJY9oZKt6QzPkB3S73SptQqq3Tx4cq

Fig. 16. PDF file format with IPFS hash value

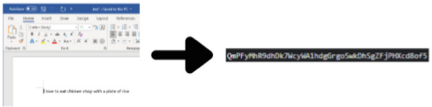

AutoSave
File
Home
Insert
Draw
Design
Layout
References
Paste
Clipboard
Font
I love to eat chicken chop with a plate of rice
QmPFyMhR9dhDk7WcyWA1hdgGrgoSwkDhSgZFjPHXcd8of5

Fig. 17. Word file format with IPFS hash value

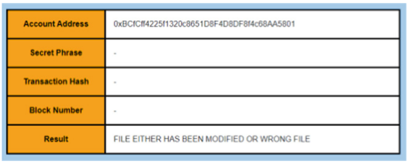
a wrong file. There are many possibilities which lead to a file that will give a different IPFS hash value.

4.4.3.1. Factors Affecting IPFS Hash Value

There are three factors that are affecting the IPFS hash value of a file which are when the content of the file has been modified, the file has been converted to a different file format, and when the file is being compressed to a zip file. The first factor indicates that whenever there is a slight change in the content of the file, it will give a different IPFS hash value. For example, Fig. 14 shows the content of the original file and Fig. 15 shows the content of a file which has a slight change and both files will give a different IPFS hash value.

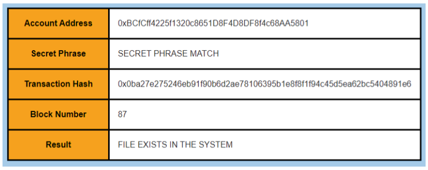
The next factor is when the file format has been changed to a different file format. For example, if a pdf file is being converted to a word file, the word file will have a different IPFS hash value from its original format file. Then, if the same word file converted back to pdf format, the new pdf format file will also have a different IPFS hash value from the word and original pdf file format. This shows that any changes to file format will give a different IPFS hash value (Figs. 16 and 17).

IPFS hash value will remained the same when the name of file is being modified or the file is being extracted from the compressed zip file. These two changes will not affect the IPFS has value. Therefore, the system has concluded that whenever there are any changes made to the file, it will be considered as a forged or modified document, and it is not a verified file. Figure 18 shows the result when the IPFS hash value of the file does not match with the IPFS hash value stored in the blockchain.



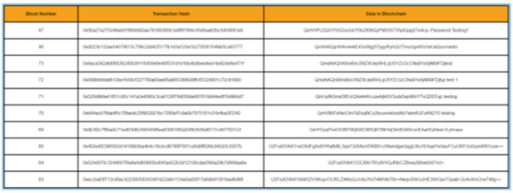
Account Address	0xBCfCff4225f1320c8651D8F4D8DF8f4c68AA5801
Secret Phrase	-
Transaction Hash	-
Block Number	-
Result	FILE EITHER HAS BEEN MODIFIED OR WRONG FILE

Fig. 18. Wrong file



Account Address	0xBCfCff4225f1320c8651D8F4D8DF8f4c68AA5801
Secret Phrase	SECRET PHRASE MATCH
Transaction Hash	0x0ba27e275246eb91f90b6d2ae78106395b1e8f8f1f94c45d5ea62bc5404891e6
Block Number	87
Result	FILE EXISTS IN THE SYSTEM

Fig. 19. File exists in the system



Block Number	Transaction Hash	Data in Blockchain
87	[illegible]	[illegible]
86	[illegible]	[illegible]
73	[illegible]	[illegible]
72	[illegible]	[illegible] test 1
71	[illegible]	[illegible] testing
70	[illegible]	[illegible] testing
69	[illegible]	[illegible]
85	[illegible]	[illegible]
84	[illegible]	[illegible]
83	[illegible]	[illegible]

Fig. 20. History Transactions Log

4.4.4 File Has Been Verified

The last scenario is when all three components match with all the data stored in the system. This occurs when the submitted transaction hash match with the transaction hash in blockchain, the submitted password match with the password stored in the blockchain and the IPFS hash value of the submitted file match the IPFS hash value of inside the blockchain. Figure 19 shows the result when the file has been verified as it does be forged or modified.

4.5 History Log

Extra feature has been provided by the application where the user can monitor the history transactions log from the application itself. The history of transaction log consists of ten rows of data which each column shows the transaction block number, transaction hash and the metadata inside the specific transaction block. Figure 20 shows the interface of history transaction log.

5 Conclusions

Blockchain is a technology that provides a lot of features and benefits in creating a document verification system. The integration between Ethereum blockchain and IPFS helps in distinguishing between the original file and the modified file. Blockchain provides a better storage system where it can keep all the related credentials safe and secure from any attack as it is permanently stored in the blockchain. There are still some limitations to the system that need to be considered for future work. The limitation of the system is related to the encryption method where the system does not provide any file encryption method since it uses other software. Other than that, the system checks for any changes made to the file and not going through the content of the file. Optical Recognition Character (OCR) can be implemented into the system to overcome the limitation of checking the content of the file. For future work, these limitations need to be considered to create a better document verification system based on blockchain technology.

Acknowledgements. The authors would like to thank Multimedia University for the supports given to this work.

References

1. A. Husain, "Printed Document Integrity Verification Using Barcode," JurnalTeknologi (Sciences & Engineering), vol. 70:1, p. 99–106, 2014.
2. B. Thuraisingham, "Blockchain Technologies and Their Applications in Data Science and Cyber Security," in 2020 3rd International Conference on Smart BlockChain (SmartBlock), Texas, 2020.
3. J. Jayachitra, Dr. S. Matilda, A. Gayathiri, "Certificate validation using blockchain," in 2020 7th International Conference on Smart Structures and Systems (ICSSS), Villupuram, 2020.
4. Barbara Guidi, Andrea Michienzi, Laura Ricci, "Data Persistence in Decentralized Social Applications: The IPFS approach," in 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), Italy, 2021.
5. Emmanuel Nyaletey, Reza M. Parizi, Qi Zhang, Kim-Kwang Raymond Choo, "BlockIPFS - Blockchain-Enabled Interplanetary File System for Forensic and Trusted Data Traceability," in 2019 IEEE International Conference on Blockchain (Blockchain), USA, 2019.
6. Alexander Von Tottleben, Cornelius Ihle, Moritz Schubotz, Bela Gipp, "Academic Storage Cluster," in 2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL), Germany, 2021.
7. IBM, "What is blockchain technology?" IBM, [Online]. Available: https://www.ibm.com/my-en/topics/what-is-blockchain.
8. Ivan, "InterPlanetary File System Explained – What is IPFS?" Moralis Academy, 8 April 2021. [Online]. Available: https://academy.moralis.io/blog/interplanetary-file-system-explained-what-is-ipfs.