DocsInBlocks - A Blockchain-Based Document Handler for Candidate Verification

Vaibhav Vesmaker[1], Prof. Sagar Korde[2]

[1]Student, Dept of Information Technology, K.J. Somaiya College of Engineering, Mumbai, India
[2]Faculty,Dept of Information Technology, K.J. Somaiya College of Engineering, Mumbai, India

Abstract – *Verifying a candidate's CV is crucial when recruiting for a company or accepting students for education. In all circumstances, we must ensure that the applicant supplies the system with accurate and legible data. The forgery of certificates and papers is a big roadblock in this procedure. Also, handling applicants' constant requests for copies of their academic certificates or other documents, and exchanging documents like transcripts between organizations can be a tedious task for the organizations. We are attempting to overcome these obstacles and simplify the procedure through this initiative. This paper's objective is to provide a full explanation of the application. It will describe the system's goal and characteristics, its interfaces, the limitations it must work under, and how the system will respond to external inputs. Along with that, we strive to make the task easier for the students or applying employees and for the educational institution and organization. The current system takes a lot of time and energy to process these documents. The new blockchain system based on Hyperledger Fabric will make the process much more streamlined.*

Key Words: **Hyperledger fabric, Blockchain, Education, Credential Verification, Verified Documents.**

INTRODUCTION

One of the greatest challenges in today's world when it comes to appointing someone in any position of responsibility is ensuring that person has the right skill set for the job. Usually, this is done by looking at an individual's past records like his education, experience, and any other credential that might be relevant to the position.

However, in recent times with the advent of technology it has become very easy to produce documents like degree certificates, score results, experience letters, or any other type of achievement or certificate which are not authentic. This poses a huge problem for organizations that pay large sums of money for an individual to occupy a position of responsibility and for educational institutions that accept students for advanced education based on their previous accolades when that person might not have the aptitude for the position at all. In both cases, we need to verify that the candidate provides correct and legible information to the system. Certificate and documents forgery is a major hurdle, in this process. Under the existing system, organizations are not able to efficiently differentiate between true and forged documents which may be submitted by some applicants.

In order to tackle this, some organizations directly contact the institution which has issued the credential for verification, but this process is long and cumbersome and often involves monetary compensation. Another drawback also includes the amount of paperwork and time required for any process like admissions, or recruitment, which involves the exchange of documents. Through this project, we are trying to leverage the advantages of blockchain technology like data integrity and immutability to overcome these hurdles and streamline the process. We have made an online platform based on Hyperledger Fabric which offers an easy way to issue, check and verify academic certificates by various organizations like schools, colleges, universities, online certifications platforms, companies, etc. It also helps organizations to reduce the immense amount of paperwork, save money, as well as provide the applicants with all their documents in a single place around the clock.

LITERATURE SURVEY

Bitcoin's decentralized architecture, which uses Blockchain technology and the proof-of-work consensus method, increased transaction transparency, trust, and hence verifiability. (refer to bitcoin paper). Yet it lacked a way to implement business logic into the system. Vitalik Buterin published a white paper in late 2014 that addressed Bitcoin's flaws, and as a result, he devised and developed the then-new concept of smart contracts. The smart contract was a novel approach to applying business logic, but the Ethereum platform was based on cryptocurrencies and mining and it also was a public blockchain on top of that. (refer to Ethereum paper). Organizations needed something where the adding of entities in the blockchain is permissioned. This was achieved by Hyperledger Fabric which gave many perks like faster transaction speed, and multiple language support.

(Privacy-preserving transparent supply chain management through Hyperledger Fabric Deebthik Ravi, Sashank Ramachandran, Raahul Vignesh, Vinod Ramesh Falmari, M.Brindh). The goal of this study is to look into how Blockchain might help improve supply chain management standards. This article also explains how to use Hyperledger

Fabric, a permissioned blockchain platform, to execute the proposed concept. In this paper's scope, current supply chain concerns such as data integrity, provenance transparency, privacy, and security are addressed more specifically in the context of the coffee supply chain business, while also seeking to generalize the solution to effectively handle other supply chain activities.

(Security analysis of a blockchain-based protocol for the certification of academic credentials Marco Baldi, Franco Chiaraluce, Migelan Kodra and Luca Spalazzi Dipartimento di Ingegneria dell'Informazione Universita Politecnica delle Marche ` Ancona, Italy, 60131). The study looks at Blockcerts, a blockchain-based protocol for certifying academic credentials that is currently being utilized around the world to validate digital certificates of competence that comply with the Open Badges standard. It investigates the certification stages used by the Blockcerts protocol to certify a certificate and discovers that they are vulnerable to impersonation attacks of a certain sort. In more detail, authentication of the issuing institution is accomplished by retrieving an unauthenticated issuer profile from the internet and comparing some of the data reported there with the data contained in the issued certificate. It demonstrates that an attacker can impersonate a legitimate issuer and produce certificates that the Blockcerts validation mechanism cannot distinguish from originals by constructing a false issuer profile and generating an appropriately changed certificate. We present some possible defenses to such an assault, which entail the usage of either a traditional public key infrastructure or a decentralized identity system that is connected with the Blockcerts protocol.

(Eductx: A Blockchain-Based Higher Education Credit Platform By (Muhamed Turkanović, Marko Hölbl , Kristjan Košič, Marjan Heričko, Aida Kamišalić.))It is a globally trusted, decentralized higher education credit and grading system that can provide a globally united viewpoint to students, higher education institutes (heis), and other potential stakeholders such as corporations, institutions, and organizations. One of the issues with the approach described above is that it excludes corporations and educational institutions, making the solution untrustworthy because the decision authority is a third party.

(Blockchain, academic verification use case by (Federico Bond, Franco Amati, Gonzalo Blousson)). Since blockchain technology provides better transparency, lower maintenance, and lower cost than traditional options, we propose using it to verify the legitimacy of academic certificates using techniques such as digital signatures and timestamps. Based on discussions made on the stage of the inaugural Bitcoin forum hosted by the administration of Ciudad de Buenos Aires on July 31, 2015.

A permissioned blockchain could be used to close these loopholes. Permissioned blockchains, in addition to the benefits of a public blockchain, provide various capabilities such as private data sharing, complete anonymity, and more, potentially making it the ideal choice for businesses. Hyperledger Fabric and other permissioned blockchain systems provide a pluggable consensus mechanism as well as cutting-edge blockchain security. In this situation, the major research questions would be:

(i) What would be the paradigm transition from a centralized to a decentralized education and hiring system?

(ii) How can we overcome the flaws of classic non-blockchain systems?

(iii) Furthermore, how can one go beyond a public blockchain to address issues like privacy, scalability, and modularity?

(iv) What would be the best method for bridging the important gap of complete anonymity?

(v) How can a network built on permissioned blockchain technology produce a stable and balanced system that meets all of the above criteria?

EXISTING SYSTEMS

Candidates' CV verification is a major task for hiring an organization and accepting students for higher education. This task is one of the most time and resource consumptive under the existing system. Many of the methods used here are outdated and involve unnecessary manual work, which can be easily replaced by faster and more secure methods.

This manual method involves contacting the issuing authority for verification which is a tedious process and may take varying amounts of time and effort depending upon how cooperative the issuing authority is, as there is no standardized procedure for this. Also, the process of organizations requiring to handle the applicants' requests for copies of their documents, and the exchange of documents when changing the organization requires a lot of time and tremendous paperwork and also involves monetary compensation.

Apart from this, the traditional non-blockchain system also has many other fundamental flaws in its system. These include but are not limited to

a. Non-transparency.

b. Data tampering and falsification which can be caused by human intervention.

c. Limited scalability.

These issues mainly arise from the use of a centralized system and thus can be easily solved if a private permissioned decentralized blockchain-like system is put in place instead. Some of the features of a private blockchain network that make it a perfect candidate for replacing the existing system are –

- Transparency.
- Data Integrity and Immutability.
- No single point of failure.
- Scalable for very large networks.
- No manual/human intervention.
- Only trusted organizations will be added to a private permissioned network.

Why Hyperledger fabric?

For Educational and Employment, we have to consider the following requirements.

- **Network needed to be Permissioned**

  Hyperledger fabric is a distributed ledger technology that is permissioned in nature. It means that all the organizations which are added to the network are known to each other. In blockchains like ethereum, bitcoin the participants are not known to each other whereas in the case of the Education and Employment sector organizations should know each other and should be enrolled by some trusted authority. In Hyperledger Fabric the Membership Services serves as the authority, which enrolls the trusted and verified organizations and gives them a cryptographic certificate.

- **No involvement of cryptocurrencies**

  Unlike Ethereum and Bitcoin blockchain, Hyperledger fabric does not use any cryptocurrency. For the Education blockchain there is no requirement of cryptocurrency or fuel for running the smart contracts and for any transaction processing. Education sector should not involve any currency which makes Hyperledger best suit for our project.

- **Pluggable Consensus**

  Hyperledger Fabric provides pluggable consensus which means that you can use consensus mechanisms as per your need. In blockchains like bitcoin and ethereum only Proof of work and Proof of Stake algorithms can be used for consensus. The small organizations don't have sufficient computational power for mining and other stuff so Ethereum and other blockchains can be used for this purpose whereas in Hyperledger Fabric we can implement our own consensus algorithm which makes it suitable for our purpose.

- **Privacy and Confidentiality**

  Privacy and Confidentiality of the applicant's data are of utmost importance in the Education and Employment sector. The data of all the students and employees need to be secured. To store the applicant's data Hyperledger Fabric provides a large variety of cryptography algorithms and protocols.

- **Large Programming Language support** Chaincode in Hyperledger Fabric can be written in a variety of languages such as go, node js, java, python, etc. These languages are very general, the developer won't have to learn a new set of languages for development.

DESIGN

A. Overview

- **How are organizations added to our network?**

  Any representative of the organization who wants to join the consortium can contact the Membership Service Provider (MSP) or the concerned authority sending the details of the organization. The MSP can verify the details, genuineness and other necessary security checks of the organization in order to maintain and increase the credibility of the consortium by not allowing fake or doubtful organizations to be a part of the network. The MSP can later send the required configuration details like the required scripts and keys in order to make the organization set up its peer and join the network. The admin who will be representing the organization has to later run the configurations on a server or computer which will be later recognized as the representative peer of the organization. The organization then uses the configurations, identifies itself in the network and also performs operations in the blockchain.

- **How are applicants added to our network?**

The applicant who wants to join the network can contact the applicant organization of his/her region. He/she can send the required details to the organization. The applicant organization can then add the applicant to the network. The applicant then becomes a part of the network and is also eligible to perform his/her permissioned operations on the network. A temporary password will be shared with him via mail. He can login and change the password immediately as soon as he is registered to the network.

- **Users and Use Cases**

There are mainly 3 characters in the system namely Admin, registrar and Applicant.

- **Admin:**

An admin is created for each organization that joins the network. The details of the admin must be present during the addition of the organization to the network. The admin details (username and password) are present in the fabric-ca-server-config.yaml of the respective organization CA configuration. The admin can add a registrar for their organization and the admin can view all the registrars of the organization.

- **Vice Admin (Registrar):**

The registrar can view the Organization dashboard where all the documents issued to the applicants by that organization are displayed. Registrar can also view all the current and previous applicants belonging to their organization. Registrar can view the personal details and academic records of the applicant if they have granted the permission. Registrar can verify the self uploaded documents of an applicant and after verification the status of the document gets changed to verified. The registrar can also update the current organization of an applicant.

- **Applicant:**

The applicant gets his credentials from the applicant organization. The applicant can view his dashboard where all the academic and other certificates are visible to him. Applicants can view their profile, update their personal details. The applicant can create self uploaded documents, this feature allows them to add their existing non digital certificates to the network and get them verified by the concerned organization (Only the organization who has previously provided the hardcopy of the document is allowed to verify the document).The registrar can than view the document, confirm the genuineness of the document and then verify(change the status of the document to verify). The applicant has the right to grant or revoke the access to/from the organization. When an applicant grants access to the organization only then the organization can view applicants personal details and all his documents.

B. Network Architecture

Figure 1 shows the proposed architecture of the network. Below are the components of the hyperledger network:

- Membership Service Provider:

A MSP manages the hierarchy of the network and also the identities. It provides the identity for organizations, i.e for the peers, orders, etc. It includes the TLS crypto materials for secure communication. There can be one or more MSPs in a network.

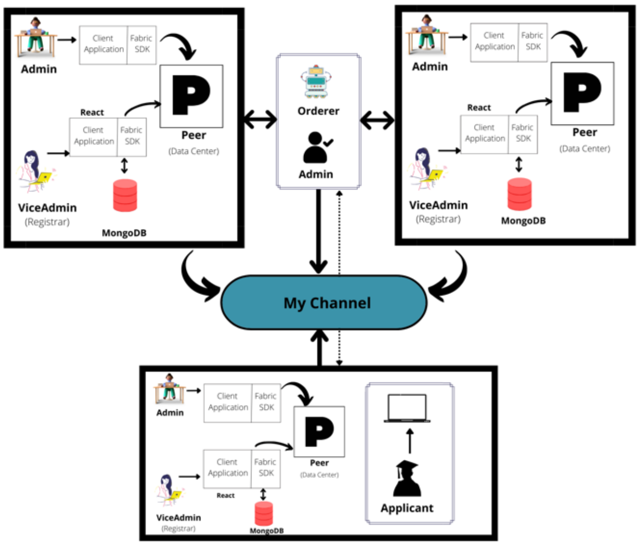


Figure:1 Network Architecture

- **Peer:**

A peer is the necessary element of the network which holds all the assets like smart contracts(chaincode), the ledger which includes the world state and blockchain, and all other configuration details. It is the element which connects to the channel, performs operations in the blockchain, and interacts with the client application.

- **Channel:**

A channel is the part that brings the consortium together, maintains privacy between organizations. A public blockchain consists of a single channel, but hyperledger enables organizations to maintain different private channels if they want to exchange shared data privately. The ledgers and the chaincode are bound to channels, i.e. there is a common ledger and a common chaincode per channel and peers can connect to multiple channels thereby holding more than one ledgers and smart contracts.

- **Chaincode:**

Chaincode is the same concept as that of smart contracts, which holds the business logic, the functions with predefined output for an input. The organizations wanting to be the part of the channel, agrees upon the common chaincode for the channel through consensus.

- **FabricSDK:**

Fabric SDK can be configured to a server (built using NodeJS and ExpressJS for our application) to connect to peers, interact with the chaincode and also hold the backend and certificates of the users of the application.

- **Client application:**

A client application is a front end application(built using ReactJS in our case). It uses the Fabric SDK Node Server to connect to channels/channels, interact with the peer node, and receive events from the peer node.

- **Ordering service:**

An ordering service is a collection of one or more orderers, where the orderer connects to channels, takes the transactions from the peers, orders and packages them and sends it again to the peers. There are many configurations for ordering service like solo, kafka depending on the number of orders and the fault tolerance of the system.

In the network representing our system, every single organization must contain components like peer, Fabric SDK server, a client application, a personalized MongoDB database in order to store the registration details as well as for authentication and authorization. There must be an administrator which represents the organization, a registrar. The admin representing the organization receives the configuration details and keys from the membership service provider. He can then manually set up the peer node. The peer can be considered as a server or data center which holds all the databases, chaincode, certificate and every other necessary configuration. The process of configuring the peer node is a one time process. He can also register the registrars using the client application provided. The registrars can be considered as one or more persons who manage the documents on behalf of the organization. The registrars are also provided with a client application which allows them to perform various operations like login into the system, sending and verifying the documents etc as discussed above.

The applicant organization helps to register the applicants into the system, authorizing them to perform operations on the network. It can be considered as a standalone organization or it can even be further divided into departments that are allowed by fabric. It can be considered a government organization whose sole purpose is to register applicants in order to maintain the genuineness of the applicants. After being registered to the network, the applicants can use the client application to interact with the network and perform operations like login, sending documents for verification, etc.

All transaction data goes through the orderer organization. All the organizations are connected to a single channel named "mychannel" and share the same ledgers. The data stored in the blockchain is encrypted hence restricting any organization to view the data which is not authorized for them and the process of encryption is discussed further.

C. Implementation

- **How are requests processed?**

When a request is posted from frontend, it goes to the node api (Fabric SDK). Fabric SDK is used to connect to the network. To call a smart contract, the network configurations are loaded, A new file system based wallet is created if it is not present to manage the identities. Then the role based authentication of the user happens. After a network is created the new gateway is setup for connecting to the network.The network details are fetched and channel details are fetched. Then the contract details are fetched and the specific transaction is submitted and the response is sent back to fabric sdk and through sdk the json data is fetched to the frontend.

- **Vice Admin (Registrar) Registration**

The admin is created for all the organizations that join the network. Admins can create Vice Admins (registrars) for their organization. Admin tries to login, its request goes to Fabric SDK, then the login credentials are verified through Mongo Database. If the credentials are valid the admin logins successfully. After the admin is created and admin log in to the network, admin can create a registrar. He enters the details of the registrar, through fabric sdk tries to connect to a network via gateway. After connecting to the network, the admin connects to a channel and registers the registrar. The wallet and x509 certificate gets created for the registrar and the registrar credentials is inserted into the mongo database.

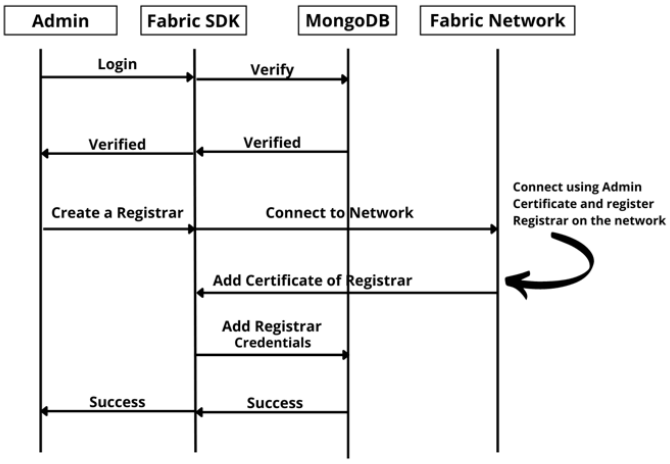

Admin
Fabric SDK
MongoDB
Fabric Network
Login
Verify
Verified
Verified
Create a Registrar
Connect to Network
Connect using Admin Certificate and register Registrar on the network
Add Certificate of Registrar
Add Registrar Credentials
Success
Success

Figure:2 Vice Applicant Registration

- **Applicant Authentication and Registration**

For registration of the applicant, the applicant needs to send the required details and documents to the registrar of the applicant organization of his/her region. The registrar can then login into the system. He can now create an applicant by filling the details in a provided form hence registering the applicant to the network. As soon as the applicant is registered, he/she is provided with a temporary password with which they can login. Later they can change the password. If the login credentials are correct, the applicant is provided with the certificate which is stored in his local storage and is later utilized for performing the authorized operations on the network.

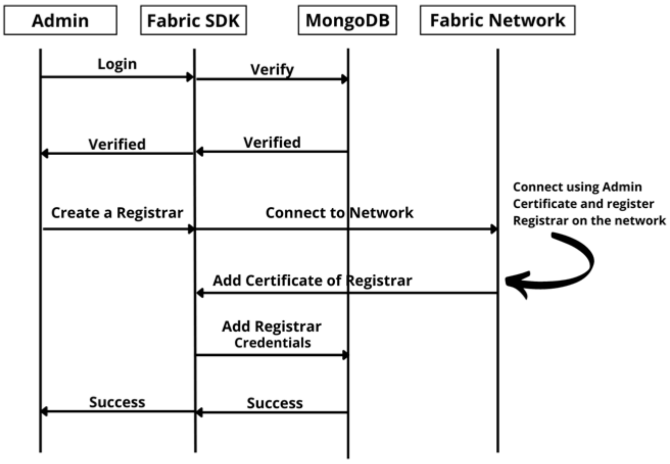
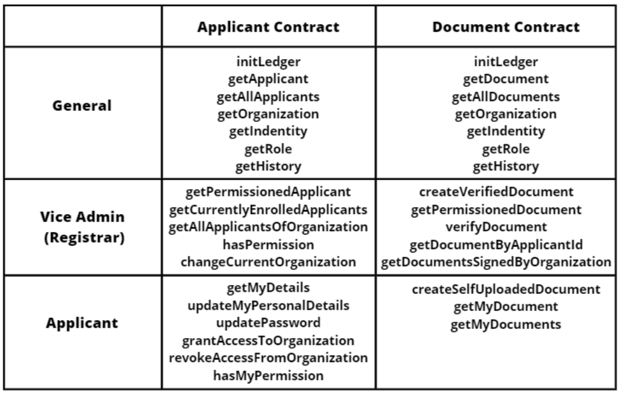

Admin
Fabric SDK
MongoDB
Fabric Network
Login
Verify
Verified
Verified
Create a Registrar
Connect to Network
Connect using Admin Certificate and register Registrar on the network
Add Certificate of Registrar
Add Registrar Credentials
Success
Success

Figure:3 Applicant Registration

- **Chain Code**

There are two contracts implemented for our system considering two assets, i.e the applicant and the document which will be transferred or to be more specific whose data will be changed. Below are the chaincode functions of the two chaincodes.



	Applicant Contract	Document Contract
General	initLedger getApplicant getAllApplicants getOrganization getIndentity getRole getHistory	initLedger getDocument getAllDocuments getOrganization getIndentity getRole getHistory
Vice Admin (Registrar)	getPermissionedApplicant getCurrentlyEnrolledApplicants getAllApplicantsOfOrganization hasPermission changeCurrentOrganization	createVerifiedDocument getPermissionedDocument verifyDocument getDocumentByApplicantId getDocumentsSignedByOrganization
Applicant	getMyDetails updateMyPersonalDetails updatePassword grantAccessToOrganization revokeAccessFromOrganization hasMyPermission	createSelfUploadedDocument getMyDocument getMyDocuments

Figure:4 Contract Functions

- **Data Encryption**

Despite the fact that blockchain is a highly effective tool for data persistence and security, there is still an issue. Everyone in the consortium has a copy of the data, as well as the blockchain, which has all of the transaction details with the information of the applicant filled in. We are striving to remedy this issue, so the data is only available to those who have permission to see it. This is accomplished through the use of public and private key cryptography. The data is encrypted using the applicant's public key, ensuring that only the applicant with the private key may decrypt and view the information.

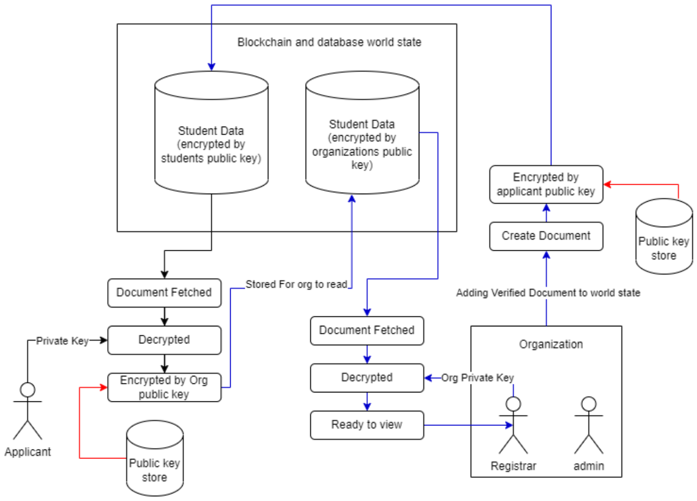

Blockchain and database world state
Student Data (encrypted by students public key)
Student Data (encrypted by organizations public key)
Encrypted by applicant public key
Public key store
Create Document
Document Fetched
Stored For org to read
Adding Verified Document to world state
Private Key
Decrypted
Document Fetched
Organization
Encrypted by Org public key
Decrypted
Org Private Key
Applicant
Ready to view
Public key store
Registrar
admin

Figure:5 Data Encryption flow

This raises a new issue: the applicant must provide access to the organization in order to provide them with the applicant's information; how will he or she do so and make his or her data available to the organization? The current option is to make a duplicate of the data that will be encrypted using the organization's public key. As a result, the organization can use that copy for viewing, confirming, and other functions. The backend server will decrypt the data using the applicant's private key. The server will also retrieve the organization's public key and encrypt the freshly copied data, making it only available through the organization's private key.

How does the organization include certificates, transcripts, grade sheets, and other documents into the applicant's file?

The answer to this challenge will be to encrypt each document using the applicant's public key and add it to the world state while also adding transactions to the blockchain. As a result, everyone possesses the data, but only those with access to it can see it.

- **Document softcopy Storage and Retrieval**

Document verification and secure storage is at our core. We strive to make the process effortless for the applicant as well as the organization. We achieve this by using document hash generated while uploading from a trusted source as well as storing it in immutable blockchain. No-one has the authority or the rights to change the document hash uploaded in the blockchain as the chaincode agreed and defined during the inception doesn't allow it.

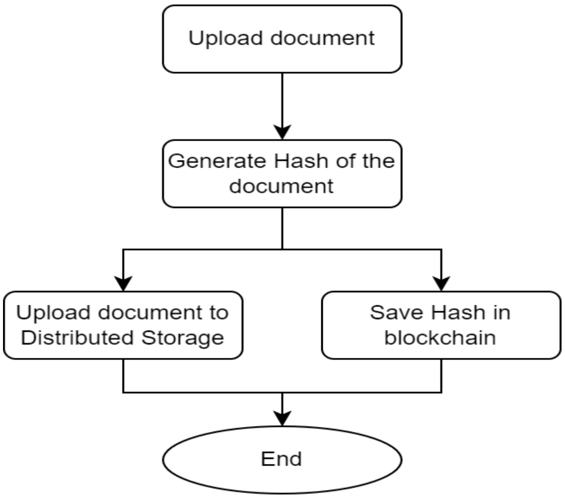

Upload document
Generate Hash of the document
Upload document to Distributed Storage
Save Hash in blockchain
End

Figure 7: Flow of Document Storage and Retrieval

Once the document is uploaded in the distributed secure storage, everytime the document is accessed the document hash will be generated and compared with the hash retrieved from the blockchain, thus, ensuring that even a single pixel isn't changed of the document. This provides a secure and effortless way of managing, storing and retrieving the documents.

The diagrammatical representation of the flow of how the documents would be stored and retrieved using two diagrams. The documents are uploaded and requested as per the requirements of the user.

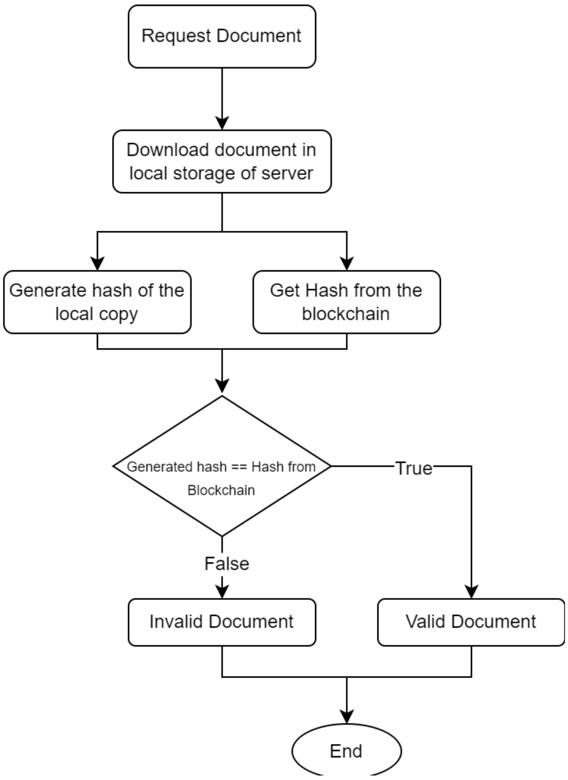

Request Document
Download document in local storage of server
Generate hash of the local copy
Get Hash from the blockchain
Generated hash == Hash from Blockchain
True
False
Invalid Document
Valid Document
End

Figure 7: Flow of Document Storage and Retrieval

FUTURE WORK

- We can decentralize the cloud storage where all the documents are stored. The current implementation involves Microsoft Azure storage because of its established security. Later, it can be decentralized and the documents can be stored on personalized more secure servers throughout the demographics.
- The applicant organization can be later separated into countrywide single organization with various departments being the states which are supported by fabric.
- The applicants can be later divided into year-wise batches for the ease of organizations to view the applicants according to the year in which they were enrolled.
- Later, the list of documents of applicants can be used to create a personalized resume with a common template making it much more easier for recruiters to find important achievements as well as hiring platforms which takes the resume and autofill the fields.

CONCLUSION

The issue with the current procedures was that it was exceedingly challenging for businesses or recruiters to distinguish between authentic and falsified documents. A significant quantity of paperwork and money were also involved with the current systems during the admissions process as well as other procedures like the issue of transcripts.

It involves a number of issues, including confirming the applicant's background, the given documents' originality, their arrangement, storage, consistency of the records, etc. Additionally, for any official task, the applicants had to carry and submit each document separately.

This sparked a requirement to digitise the entire system in order to address these issues. However, digitization raises security concerns, like the possibility of document content changes, all documents being accessible to hackers in one location, etc., if they are all held by a single centralised institution.

Blockchain was the ideal platform to address these security and decentralisation challenges. However, the main worry was that anyone might use the resources and operate on the blockchain.

Thus, hyperledger fabric emerged as the structure that met all the requirements. Due to the application's use of the hyperledger platform, which only permits permissioned companies and users to operate, it has far higher security than other blockchain platforms. Given that every document's state is visible to the user, this platform aids organisations in distinguishing between authentic and fake documents. Organizations may access the applicant's whole academic and professional history in one location, giving them a sense of a CV that follows a standard format. Additionally, our programme eliminates the need for candidates to submit the same documents repeatedly and allows them to share their profiles with any organisation by giving that organisation permission with just one click. The benefits of this application also include the efficient migration of the current system to our application, the ability of smaller organisations to afford the setup due to the elimination of the need for local storage of the images of documents using cloud services, and improved security against unauthorised access to or tampering with data. Smaller groups can even join an umbrella organisation if they can't afford the establishment.

REFERENCES

[1] Satoshi Nakamoto, ``Bitcoin: A Peer-to-Peer Electronic Cash System'', March 2009.

[2] Deebthik Ravi, Sashank Ramachandran, Raahul Vignesh, Vinod Ramesh Falmari,M. Brindh, "Privacy preserving transparent supply chain management through Hyperledger Fabric", Elsevier 100072, 16 March, 2022.

[3] Marco Baldi, Franco Chiaraluce, Migelan Kodra and Luca Spalazzi, "Security analysis of a blockchain-based protocol forthe certification of academic credentials", October 2019.

[4] Muhamed Turkanovic , Marko Holbl, Kristjan Kosic, Marjan Hericko, Aida Kamisalic, "EduCTX: A Blockchain-Based Higher Education Credit Platform", IEEE Access PP(99), October 2017.

[5] Federico Bond, Franco Amati, Gonzalo Blousson, "Blockchain, academic verification use case", 31 August, 2015.

[6] Sandner, P., Nägele, T. and J. Gross, "Liechtenstein Blockchain Act: How can nearly any right and therefore any asset be tokenized based on the Token Container Model?", Medium, October 7, 2019.

[7] Office of Ed Tech, "Education have a problem? Put a blockchain on it!", Medium, 26 February, 2020.

[8] Mara-Florina Steiu "Blockchain in Education: Opportunities, application and challenges", First Monday, Volume 25, Number 9, 7 September 2020.

[9] Maryville University, "How Blockchain Is Used in Education".

[10] Rajeev Sakhuja, "Hyperledger Fabric 2.x Network Design & Setup", Udemy, July 2020.

[11] Aditya Joshi, "The Complete Guide on Hyperledger Fabric v2.x on Kubernetes", Udemy, May 2021.

[12] Hyperledger White Paper.

[13] Hyperledger Documentation.

[14] Inter Planetary File System Documentation.