Data Privacy and Sharing

How certificates will be stored on the blockchain network.

Certificates are placed on a distributed ledger so that users can access information anytime, anywhere and at any place. Using a private key, the documents are kept secure allowing only authorized persons to access it. It takes about 5 min to 2 hours for the submission to be verified. If it is done manually due to some reasons, it may take up to 5 days. If a university or an institution decides to store its certificates or carry out its certification on blockchain technology, we would need to convert these documents to a hash code using cryptography. The code is further stored on the blockchain network. The hash code generated resembles a copy of the document but represented as a string of codes. These strings of codes act as a key for the document. When the user presents this document elsewhere, the code must remain the same as stored in the blockchain for it to be valid. In order to verify the authenticity of the certificates, digital signatures and timestamps are important. We could use an interface base contract that provides a standard for signed digital asset verification.As blockchain is a decentralized system, digital signatures can be accessed by anyone without relying on trusted intermediaries, this therefore saves time to the company. We store the proof that a digital asset has been certified by an institution on blockchain.

This will includes these steps that :

- **Creation** of the digital asset and storing the digital signature to the blockchain network.
- **Transmission** of the digital asset. For e.g Email, file sharing, etc…
- **Validation** of the digital assets utilizing the digital signature and the institution mentioned in the blockchain.

Details stored on the certificate to prevent fake online certifications are :

- A signed digital certificate containing the owner's distinguished name
- Owner's public key
- Certificate authority's distinguished name
- Signature of the certificate authority over the fields

List and discuss the type of permissions to be used in the solution and reasons for selecting it.

1. **Public blockchain -** Decentralized, permissionless distributed ledger system operates on a public level.
2. **Private blockchain -** Restrictive, permissioned blockchain operates in one organization only.
3. **Consortium blockchain -** Semi-centralized. Same like private blockchain but allow for multiple organizations to manage.
4. **Hybrid blockchain -** Combination of private and public blockchain. Allow only a selected section of data to go public, keeping the rest private.

So the question arises: public vs private or other types of permissions for authentication of certificates by the institution?

In my humble opinion, this depends on the type of institution. As every institution across the world operates differently. If we take the university of Mauritius as an example, the more suitable to it, would be the consortium blockchain. Because it would allow the collaboration of different organizations compared to private blockchain which allow only one organization to manage it. For example let's take this scenario, me as a student is participating in a competition. So I would obtain a certification stating whether I won the competition or if I have only participated. If we allow only private blockchain, it will not be wise to do so because if ever I did not win but obtain only a participation certification. Considering the fact that unethical behavior is exempted from me. I can lie to the job interviewers saying that I did win a competition at my university. There is no proof to whether what I stated is true or not. Because my graduation degree will count as only one organization being allowed, which in this case is the University of Mauritius, no proof for the competition that collaborated with the university.

But in terms of authentication, there are many educational institutions online where their main aim is to only issue their own certification and there are no collaborations between universities or companies. Private blockchain will be considered most suitable since it is established to validate whether participants are trustworthy or not. It is easy to control and maintain. As all participants are known, if ever there is counterfeit law sanctions can be implemented.

Data ownership, Privacy and sharing

Data ownership :

By default blockchain data is equally owned at each place it is distributed but for personal data, regulations require one owner to be accountable for all data privacy. Instead of storing the data in a blockchain we could store index numbers that will be tied to personal data in a separate database. This will thus, make one organization own and secure that database while still being able to share the blockchain. Storing data in another database can allow data deletion in an organization. Blockchain data is permanently stored, if there are data changes there is still a record of what it used to be many users would want to delete their data entirely from the system for, of course, data privacy. This would allow for an editable database where the users would be able to change and delete data.

Data Privacy :

Data privacy is secured by the use of public and private keys. The issue of possessing transactions visible to everyone who has access to the blockchain should also be recognized, since this data may be obtained and made public elsewhere. Still, to overcome these arguments, data may be kept on the blockchain in encrypted form. . In order to make the digital's asset unique. We would need to use cryptographic hash functions. For e.g MD5 is one of the hashing functions available. For each block in the blockchain, it contains a cryptographic hash of the previous block,a timestamp and transaction data. Ensuring data privacy is crucial especially when a student's career may be at risk. Educational institutions may need to implement stronger privacy measures by the use of private or permissioned blockchain.

Data Sharing :

As data can be shared publicly or privately. Users can publicly share their certification with different organizations which will therefore be difficult for malicious hackers to attack and manipulate the authenticity of the certification. Data sharing in blockchain would allow users to create profiles and store data and make it accessible to others as per agreement. Users are able to specify a range of purposes of data sharing, kinds of data that can be shared and classes of applications/companies that can access the data.

How can organizations validate certificates for individuals that they are hiring?

Let's assume a recruiter is seeking to hire someone for his service and requests his certificates for verification. With blockchain, the recruiter can accomplish this by slotting the key generated from hashing the content of the document into the institution's portal or a verification app which includes the database of all their legit graduated students. If the hash matches the details stored on the blockchain by the institution he claimed to have graduated or certified from, the certificates produced are hereby genuine.

Chaincode

REST API's that are needed to call validation from the solution

As there are two type of chaincode :

System Chaincode : it runs part of peer process

Normal chaincode : it runs a separate container managed by peer

So the chaincode preferred is system chaincode because it is used to perform low-level ledger features like the validation system, query system, and endorser system.

The REST API's for the oracle blockchain platform that would get called are :

- **Get chaincode information**: Method GET
  Path `/console/admin/api/v1.1/chaincodes/{chaincodeName}`
- **Get installed chaincode List:** Method GET
  path `/console/admin/api/v1.1/chaincodes`
- **Install chaincode:** Method POST
  path `/console/admin/api/v1.1/chaincodes`
- **Instantiate Chaincode:** Method POST
  path `/console/admin/api/v1.1/chaincodes/{chaincodeName}/instantiate`

Identify and outlines the Channels that would be created from the case study

The channels are used when we want to keep confidentiality of data within a set of organization so

1. private data is suitable : Because the institutions can keep data about the users private from other organizations on the same channel.
2. We also have the option to a new channel where only organizations that need the data would be allowed to. But creation of separate channels creates additional administrative overhead. It does not allow it.
3. Private data purging where government regulations require data to "purge", a hash would be left so as to maintain privacy of data.

Private Data sharing using chain codes can be carried out :

- Use a corresponding public key for tracking public state
  A matching public key can be provided for public state tracking such as current ownership. In each educational institution's private data collection, a private key is generated for each organization that should have access to the corresponding private data of the asset.

- Chaincode access control
  Specify which user can query private data in a collection.For ex a user requests an educational institution for his certificate then the user's credentials in the chaincode

are obtained and verify if they can have access before returning private data. A passphrase is required into the chaincode.

- Sharing private data out of band

As sharing of private data is permissible to other organizations we are able to write private data to other organizations then. Hashing of the value by the use of GetPrivateDataHash() is done to verify if it matches the on-chain hash. This way companies or organizations are able to check if you are the legitimate owner of the certification.It is (off-chain)

- Sharing private data with other collections

Private data could be shared by creating a matching key in a private data collection of another organization. A transient field is passed through the chaincode confirming a hash using GetPrivateDataHash(). It is (on-chain)

- Transferring private data to other collections
  Transfer of private data is done with chaincode where it deletes the private key in the collection and generates it in the collection of another organization. The use of transient fields is again used to allow verification of data existing in the collection before deletion of the key from one collection and added to another collection. The use of endorsements is required

- Using private data for transaction approval.
  Pre-approve transaction is required if a counterparty needs approval before completion of a transaction. Private key is written to their private data collection then the chaincode will check using GetPrivateDataHash(). More powerful with collection-level endorsement policies.

- Keeping transactors private

  Proof of transaction with hash and references are recorded on-chain. For a given transaction, variations of the previous pattern can also eliminate the leaking of transactors. Only the educational institutions and users are aware that they are transactos. But can reveal pre-images if necessary.

Algorithm for Chaincode

We would have to package and zip the node.js chaincode. We would have 2 section in the package.json file with two sections:

- The scripts section declares how to launch the chaincode.
- The dependencies section specifies the dependencies.

Sample package of node.js chaincode:

```
{
	"name": "chaincode_example02",
	"version": "1.0.0",
	"description": "chaincode_example02 chaincode implemented in Node.js",
	"engines": {
		"node": ">=8.4.0",
		"npm": ">=5.3.0"
	},
	"scripts": { "start" : "node chaincode_example02.js" },
	"engine-strict": true,
	"license": "Apache-2.0",
	"dependencies": {
		"fabric-shim": "~1.3.0"
	}
}
```

We must abide by the rules for a node.js chaincode such as:

- package.json must be in the root directory.
- The entry JavaScript file can be located anywhere in the package.
- If "start" : "node <*start*>.js" isn't specified in the package.json, server.js must be in the root directory.

Then, we place the chaincode and package file in a zip file to install it on Oracle Blockchain Platform.

Challenges Identified

1. Not easy to use for novice people but developers are working on it to make it accessible to everyone, considering the huge public interest which has grown regarding blockchain.

2. Requires huge initial investment that only well-stable universities can invest in, putting smaller universities with a lack of financial resources as an outcast.

3. Courses with a huge number of users will have to authenticate one by one then proceed to add them in the blockchain which is time consuming.

4. Aso in terms of data Privacy and security

Issue : Blockchain is built in such a way that all transactions are transparent while its actors can be identified. This increases the lack of privacy and security

Solution : The use of Permissioned or private blockchain platforms like Quorum, Hyperledger Fabric and Corda, which provide the capability of executing private transactions between two or more participating nodes. This ensures that the transaction details pertaining to the sender and recipient are part of a private ledger and will not be revealed to unauthorized participants.

Also, according to google it states that nobody has single-handedly hacked a blockchain. A group of hackers or people within the organization collaborate to breach the blockchain' security.

Summary

This report aims to respond to how blockchain can help when it comes to prevention of fake online certification. The issue has been addressed as how blockchain is useful, how smart contracts would work, how data privacy and sharing is maintained, how chaincode operates, what are the challenges faced and how we can overcome them. The need for proper security is needed so as to prevent hackers from getting access and not disrupting the normal function of an educational institution. As most places around the earth are affected by covi-19, online-learning and e-certificate being more implemented in our society. There is eradication from traditional-based systems and turning to digital systems. Blockchain will definitely be a thing in the near future as it provides a better guarantee of data privacy and helps users to not get their data stolen and sold to other third-parties. It noticeably helps against the use of fake online certifications.