**PROJECT REPORT**

**Decentralized Application for**

**Digital Certification**

**Personnel:**

Aravind. G [14BT1IS009]

Varun. K [14BT1IS040]

**1. Project Overview**

One of the great promises of blockchain technology is that it can serve as a decentralized permanent unalterable store of all types of information or assets, not just as a currency or payment system. The project comprises of an application which would provide information about the certification of student’s educational qualification which is digitally signed by the university/Education Board using blockchain technology. This is also applied to POA or POI documents as well. Similar to the idea of an E-Aadhaar but with a different use of technology.

**2. Purpose of the Project**

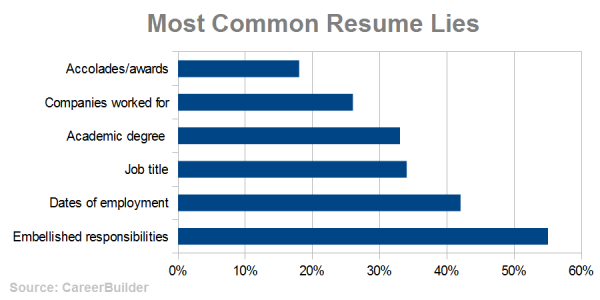
* **Problem Statement**

Many aspirants like to pursue higher education at countries which specialize in a particular domain. Application to such universities requires document verification which is done by contacting the respective schools and colleges to provide confirmation about the applicant’s qualification.

Similarly, in business companies perform background and educational verification of their employees. The reason behind such verification is that across the globe there are numerous fraudulent cases. Employee and students are found to have duped or lied in their resume about the certifications. This confirmation is done only done in the later stages of the verification process and would have given enough time for the fraudulent to have taken advantage over the company or university.

Time is wasted upon performing such tasks. On paper nothing seems to be believable unless confirmed by the board or institution. There are many such cases even in India. For instance the Dr. BR Ambedkar University in Agra is alleged to have handed out thousands of fake degrees. Over 100s of fake degrees have been to relatives of the employees of the university. This wasn’t confirmed until mid 2015.

 A survey by one of the largest online job finder sites, CareerBuilder, shows that a staggering 58 percent of employers have caught a lie on a resume. The site has more than 23 million unique visitors and over 1.6 million jobs. Just over half of employers, 51 percent, said that they would automatically dismiss a candidate once caught. Only seven percent said they would be willing to overlook a lie, if they liked the candidate. The HireRight report revealed that 50% of employers check education verification



We can see that over 30% of the resume lie are academic degree.

To prevent such fraudulent cases and also provide an ease of presentation of information to the respective organization is one of the reasons behind such a project.

* **Motivation**

It was the Hacker Earth Blockchain workshop which initiated the motivation to build such a product. They demonstrated how Blockchain is used for linking and securing blocks of information about crypto currency such as Bitcoins. Any transaction performed through the Blockchain a permanent and verified, which means that information cannot be modified and is digitally signed by an organization.

Blockchain is designed to be secure as it operates on a peer-to-peer decentralised network, that is the information is linked through ledgers which share the information and are geographically separate from one another. This is what they call Decentralised storage.

By utilising such a feature data cannot be accessed easily by hackers and there are various ledgers which are connected by pointers. Once transaction is completed, then the information cannot be modified. This enables the Education Board or University to permanently certify a student and his/her certificate can be presented to any university.

The process of verify this certificate is much quicker than the previous methods and is more convincing. Such technology is already being implemented in many countries. For instance MIT Lab has taken the initiative to even provide open source their project called BlockCerts. For even participation of workshops such certification is awarded. But Indian educational board hasn’t taken that step forward yet. Thus to enable it to move forward is what motivated us to come up with such a project.

**3. Goals of the project**

* To deliver hassle free data verification for organization bodies.
* To provide digitally accurate records of official records.
* Use of a smart phone to provide records/documents instead of numerous files of paper.
* Secure and prevent the records from any alteration/fraud.
* Permanent and indestructible records unlike paper documents.

**4**. **Blockchain’s Significance**

Despite the blockchain being in its initial stages, lot of applications and services are being rapidly developed. Services such as transfer of money, proof of consistency, proof of ownership, smart contracts and various other concepts. Blockchain is even will be used for voting in the distant future. But what concerns us the most is its robustness and immutability. Upon looking up these concepts enabled us to envision the project which can be used in any institution and alleviate many difficult processes.

Blockchain is the technology used behind the digital crypto currency known as Bitcoins. This technology was developed by a group known as Satoshi Nakamoto to solve the double spending problem which was problem of duplication/falsification. This paved way for a transaction without a trusted authority or a central server.

Web 3.0, a term coined for the change in the protocol of the Internet. At the moment, majority of the internet works on a centralized network where there is always a central server look after functions of the network. Since the introduction of bit coins, the technology behind it is being applied to the web as well. Decentralization is the key word here, which tells how Web 3.0 is totally a different path from the old Web 2.0. The protocols behind the new Web are in contrast with the old version, which means many applications are to be built from a scratch.

Without the evolving internet many web application had to upgrade to the emergence of the new technology. Tech giants such as IBM, Microsoft, Amazon, Infosys and many more have dwelled deep into this technology. Blockchain as a service (BaaS) has been initiated by cloud service providers, for instance this service is available on Bluemix by IBM, Amazon web services and even Microsoft Azure. This enables developers to build, test and deploy decentralised applications.

With such services already being provided, there is one such platform which is has been widely at the current point in time. This platform is called Ethereum. Ethereum is an important component in this report as we develop the application on this platform which shall be discussed further in the report.

**5. Literature Survey:**

Upon looking up the subject digital certification, one of most frequent occurring paper would be the MIT digital certificate paper. This provided the foundation to the application which we wanted to develop. This paper redirected to the new initiative which was undertaken by MIT known as Blockcerts.

This initiative was led by MIT Media Labs and Learning Machine. Learning Machine is a company which is solely dedicated to decentralising records which along with the partnership with MIT Media Labs co-created Blockcerts.

The above system architecture is based on the process by which Blockcerts function their digital certificates.

Blockcerts is an open source standard for digital certification. It is aligned with the following Decentralization and Data Signature standards.

* [IMS Open Badges](https://www.imsglobal.org/sites/default/files/Badges/OBv2p0/index.html)
* [W3C Verifiable Claims](https://w3c.github.io/vc-data-model/)
* [W3C Linked Data Signatures](https://w3c-dvcg.github.io/ld-signatures/)
* [W3C / Rebooting Web of Trust Decentralized Identifiers](https://github.com/WebOfTrustInfo/rebooting-the-web-of-trust-fall2016/blob/master/draft-documents/DID-Spec-Implementers-Draft-01.pdf)

With these standards Blockcerts assert that a viable application can be created on a decentralised system.

Maltese government have partnered with Learning Machine to provide digital academic certification allowing its people to store their academic qualifications and other records for free.

Holberton School, which was the first institution to issue digital certificates to students. Holberton School has partnered up with Bitproof, a company which is focused on producing digital certificates. This reported enabled employers to find at least 86% of the employee who lied in their resume. Bitproof also now provides developer tools to create certificates and also enable to develop blockchain applications as well.

Raman Technologies as well developed an academic digital certification as well. They have used the following stack for the development.

1.Ethereum Blockchain (Ethereum ropsten network)

2.PHP/AngularJS for web app development

3.Solidity smart contracts

4.IPFS distributed file storage

5.RabbitMq /whisper – messaging framework

6.PHP MVC for development of model view controller

7.PHP Laravel for RESTful web service framework

8.PKI and digital certificates (X.509 digital certificates)

Their application was based on Web applications and had expressed the desire enable this for Android and iOS.

The Humanized Internet, which includes a so-called identity-as-a-service, relying on the blockchain system. Blockchain offers an immutable, transparent, and distributed ledger that can provide a secure means of identifying every person on Earth. Think of blockchain as a universal, secure digital lockbox that could store information with your legal ID, such as property title, education certificates, and medical records, all in one place. The owner of the documents could access the system via mobile phone, and the identity could be confirmed using the owner’s biometrics.

Sapien’s Project was a digital certification project which focused on something little different. It focused on scalability of blockchain technology. They state that Lisk Sidechain would be able to provide SDK’s for local computation which is less cost effective and scalable with high computation processes.

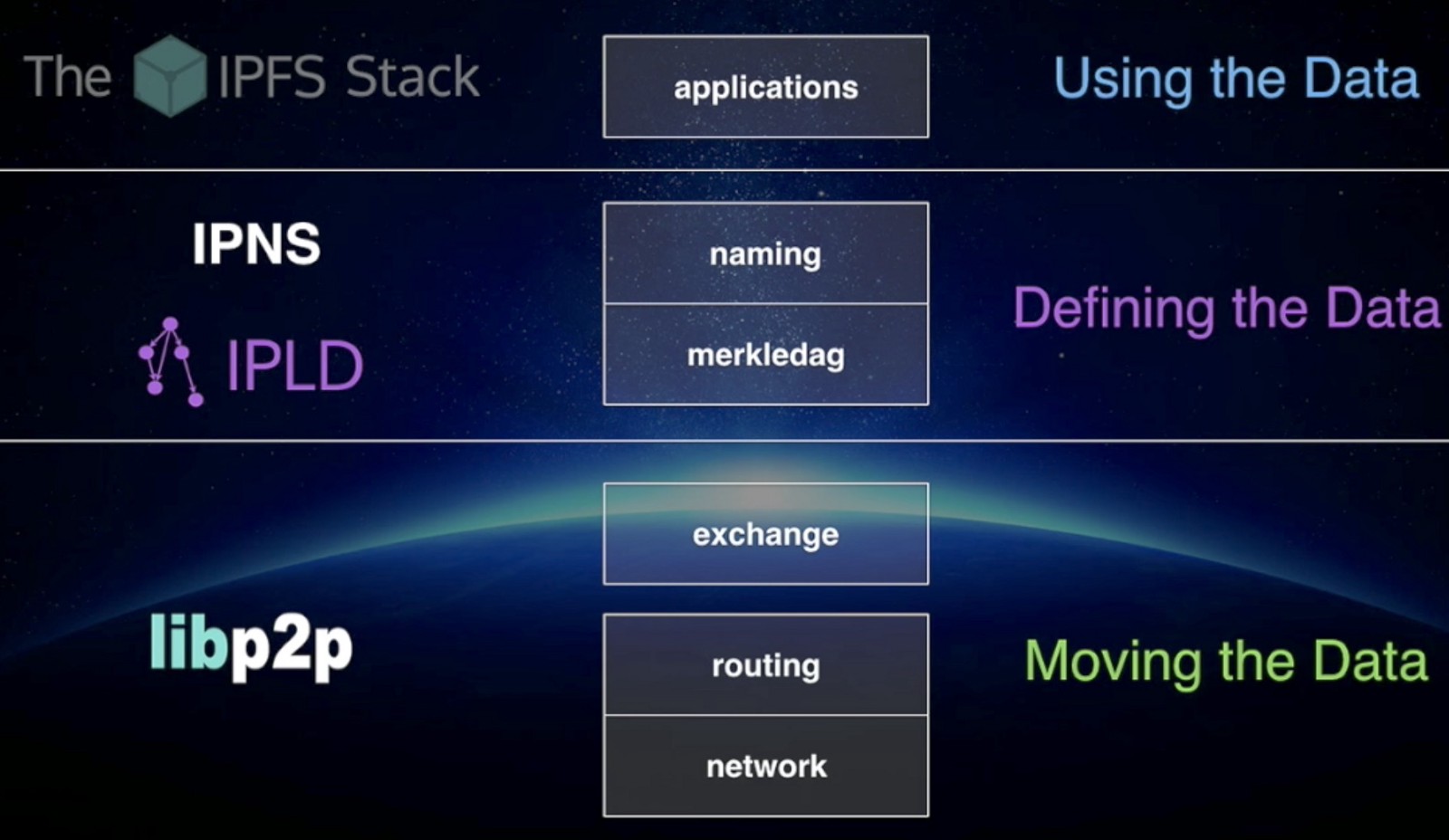
Proof of Existence, a blockchain based website which is used to prove the existence of a document. It was noted that some amount of money had to be paid as miner fee. Another observation was that word processing documents were not advisable. This is because word documents possessed metadata, the cryptographic digest generated is solely based on the document’s content. So the metadata of the document does not enable the blockchain to verify the existence of the document at the timestamp. This proves how exact the document must be for the blockchain to verify it. PDF and other unalterable format would be better for this usage.

**Protocols behind blockchain based applications**

In any application, there are set of protocols that are followed. For instance, in DBMS can have a RDBMS protocol, which works on relationship between two or more entities. Similarly, in a web application, it would be based on SMTP, HTTP, FTP and so on.

Blockchain applications are based on a newly developed protocol which is known as IPFS. IPFS stands for InterPlanetary File Systems.

IPFS (the InterPlanetary File System) is a new hypermedia distribution protocol, addressed by content and identities. IPFS enables the creation of completely distributed applications. It aims to make the web faster, safer, and more open.



IPFS is combination of few properties, which is stated below:

* **DHT (Distributed Hash Table)**

DHT is a class of a decentralized distributed system that provides a lookup service similar to a hash table: (*key*, *value*) pairs are stored in a DHT, and any participating node can efficiently retrieve the value associated with a given key. Responsibility for maintaining the mapping from keys to values is distributed among the nodes, in such a way that a change in the set of participants causes a minimal amount of disruption. This allows a DHT to scale to extremely large numbers of nodes and to handle continual node arrivals, departures, and failures.

* **Block Exchanges - BitTorrent**

BitTorrent is a widely successful peer-to-peer file sharing system, which succeeds in coordinating networks of un-trusting peers (swarms) to cooperate in distributing pieces of files to each other. Key features from BitTorrent and its ecosystem that inform IPFS design include:

1. BitTorrent's data exchange protocol uses a quasi-tit-for-tat strategy that rewards nodes who contribute to each other, and punishes nodes who only leech others' resources.

2. BitTorrent peers track the availability of file pieces, prioritizing sending rarest pieces first. This takes load off seeds, making non-seed peers capable of trading with each other.

3. BitTorrent's standard tit-for-tat is vulnerable to some exploitative bandwidth sharing strategies. PropShare is a different peer bandwidth allocation strategy that better resists exploitative strategies, and improves the performance of swarms.

* **Version Control Systems - Git**

Version Control Systems provide facilities to model files changing over time and distribute different versions efficiently.

The popular version control system Git provides a powerful Merkle DAG 2 object model that captures changes to a file system tree in a distributed-friendly way.

1. Immutable objects represent Files (blob), Directories (tree), and Changes (commit).

2. Objects are content-addressed, by the cryptographic hash of their contents.

3. Links to other objects are embedded, forming a Merkle DAG. This provides many useful integrity and workflow properties.

4. Most versioning metadata (branches, tags, etc.) are simply pointer references, and thus inexpensive to create and update.

5. Version changes only update references or add objects.

6. Distributing version changes to other users is simply transferring objects and updating remote references.

* **Self-Certified File systems - SFS**

SFS proposed compelling implementations of both (a) distributed trust chains, and (b) egalitarian shared global namespaces. SFS introduced a technique for building Self-Certified File systems: addressing remote file systems.

Thus, the name of an SFS file system certifies its server.

The user can verify the public key ordered by the server, negotiate a shared secret, and secure all traffic. All SFS instances share a global namespace where name allocation is cryptographic, not gated by any centralized body.

Another important concept behind this application would be the use of Merkle trees. This provides the most important aspect of the application which is discussed below.

**Merkle tree** is a tree in which every leaf node is labelled with a data block and every non-leaf node is labelled with the cryptographic hash of the labels of its child nodes. Hash trees allow efficient and secure verification of the contents of large data structures. Hash trees are a generalization of hash lists and hash chains.

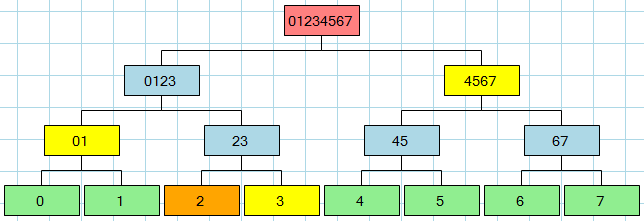
Demonstrating that a leaf node is a part of a given binary hash tree requires computing a number of hashes proportional to the logarithm of the number of leaf nodes of the tree: this contrasts with hash lists, where the number is proportional to the number of leaf nodes itself.

The reason behind the use of Merkle trees:

1. Merkle trees provide a means of proving that integrity / validity of your data.
2. Merkle trees require little memory / disk space and proofs are computationally easy and fast.
3. Merkle tree proofs and management requires only a very small and terse amount of information to be transmitted across a network.

* **Data Existence Verification with Merkle trees:**

Let's say you are the owner of the record "2" in the below diagram.  You also have, from a trusted authority, the root hash, which in our simulation is "01234567".  You ask the server to prove to you that your record "2" is in the tree.  What the server returns to you are the hashes "3", "01", "4567" as illustrated here:



Using this information (including the right-left flags that are sent back along with the hashes), the proof is that:

* 2 + 3 from which you compute 23
* 01 + 23 from which you compute 0123
* 0123 + 4567 from which you compute 01234567

Since you know the root hash from your trusted authority, the proof validates that "2" exists in the tree.  Furthermore, the system from which you have obtained the proof is proving to you that it is an "authority" because it is able to provide valid hashes so that you can get from "2" to your known root hash "01234567."  Any system pretending to validate your request would not be able to provide you with the intermediate hashes since you're not giving the system the root hash, you're just telling it to give you the proof - it can't invent the proof because it doesn't know your root hash -- only you know that.

In order to verify the proof, very little information about the tree is revealed to you.  Furthermore, the data packet that is needed for this proof is very small, making it efficient to send over a network and to make the proof computation.

* **Ethereum**

Ethereum is an open blockchain platform that lets anyone build and use decentralized applications that run on blockchain technology. In late 2013, Ethereum’s inventor Vitalik Buterin proposed that a single blockchain with the capability to be reprogrammed to perform any arbitrarily complex computation could subsume these many other projects. Ethereum is a programmable blockchain. Rather than give users a set of pre-defined operations (e.g. bit coin transactions), Ethereum allows users to create their own operations of any complexity they wish. In this way, it serves as a platform for many different types of decentralized blockchain applications, including but not limited to cryptocurrencies.

Ethereum in the narrow sense refers to a suite of protocols that define a platform for decentralised applications. At the heart of it is the [Ethereum Virtual Machine (“EVM”)](http://ethdocs.org/en/latest/contracts-and-transactions/developer-tools.html#the-evm), which can execute code of arbitrary algorithmic complexity. In computer science terms, Ethereum is “Turing complete”. Developers can create applications that run on the EVM using friendly programming languages modelled on existing languages like JavaScript and Python.

Like any blockchain, Ethereum also includes a peer-to-peer network protocol. The Ethereum blockchain database is maintained and updated by many nodes connected to the network. Each and every node of the network runs the EVM and executes the same instructions. For this reason, Ethereum is sometimes described evocatively as a “world computer”.

This massive parallelisation of computing across the entire Ethereum network is not done to make computation more efficient. In fact, this process makes computation on Ethereum far slower and more expensive than on a traditional “computer”. Rather, every Ethereum node runs the EVM in order to maintain consensus across the blockchain. Decentralized consensus gives Ethereum extreme levels of fault tolerance, ensures zero downtime, and makes data stored on the blockchain forever unchangeable and censorship-resistant.

Ethereum platform satisfies the needs of the project providing robustness and unadulterated verification to ensure the integrity and security of the data.

From a practical standpoint, the EVM can be thought of as a large decentralized computer containing millions of objects, called "accounts", which have the ability to maintain an internal database, execute code and talk to each other.

There are two types of accounts:

**1. Externally owned account (EOAs)**: an account controlled by a private key, and if you own the private key associated with the EOA you have the ability to send ether and messages from it.

**2. Contract**: an account that has its own code, and is controlled by code.

The code has the ability to read/write to its own internal storage (a database mapping 32-byte keys to 32-byte values), read the storage of the received message, and send messages to other contracts, triggering their execution in turn. Once execution stops, and all sub-executions triggered by a message sent by a contract stop (this all happens in a deterministic and synchronous order, i.e. a sub-call completes fully before the parent call goes any further), the execution environment halts once again, until woken by the next transaction.

Contracts generally serve four purposes:

1. Maintain a data store representing something which is useful to either other contracts or to the outside world; one example of this is a contract that simulates a currency, and another is a contract that records membership in a particular organization.

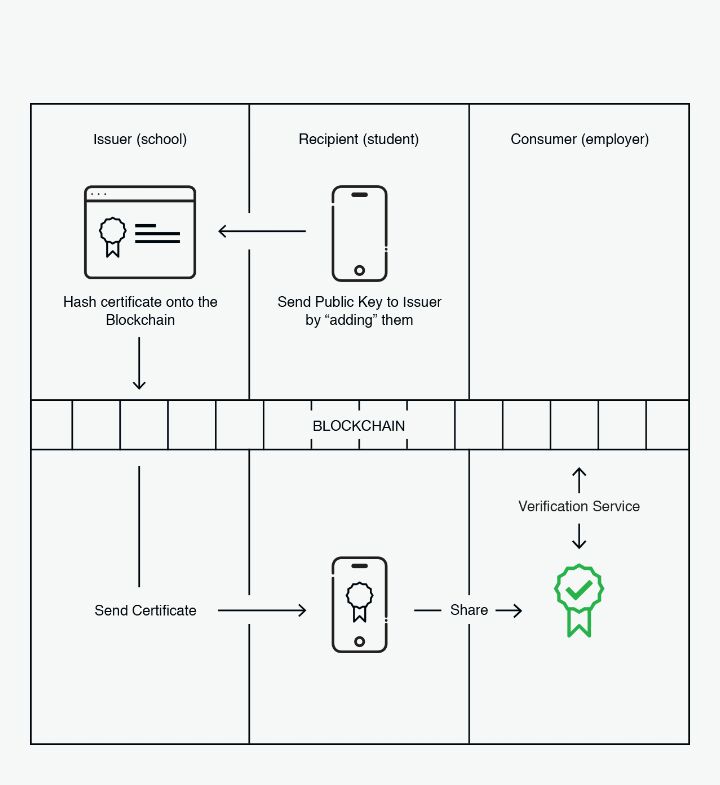
2. Serve as a sort of externally owned account with a more complicated access policy; this is called a "forwarding contract" and typically involves simply resending incoming messages to some desired destination only if certain conditions are met; for example, one can have a forwarding contract that waits until two out of a given three private keys have confirmed a particular message before resending it (i.e. multisig). More complex forwarding contracts have different conditions based on the nature of the message sent; the simplest use case for this functionality is a withdrawal limit that is overridable via some more complicated access procedure.

3. Manage an ongoing contract or relationship between multiple users. Examples of this include a financial contract, an escrow with some particular set of mediators, or some kind of insurance. One can also have an open contract that one party leaves open for any other party to engage with at any time; one example of this is a contract that automatically pays a bounty to whoever submits a valid solution to some mathematical problem, or proves that it is providing some computational resource.

4. Provide functions to other contracts; essentially serving as a software library.

Contracts interact with each other through an activity that is alternately called either "calling" or "sending messages". A "message" is an object containing some quantity of ether (a special internal currency used in Ethereum with the primary purpose of paying transaction fees), a byte-array of data of any size, the addresses of a sender and a recipient. When a contract receives a message, it has the option of returning some data, which the original sender of the message can then immediately use. In this way, sending a message is exactly like calling a function.

The first use satisfies the requirement of the project which states the records of the membership of a person in an organization which is similar to the verifying an existence of a person in an institution.

**6. Application Architecture**

As the above diagram mentions, initially the user whose certificate is to be verified sends a public key to the issuer. The issuer in our case is the University who provides validation of a student’s educational qualification in that university. Basically, the student requests the School/University to issue his/her certificate.

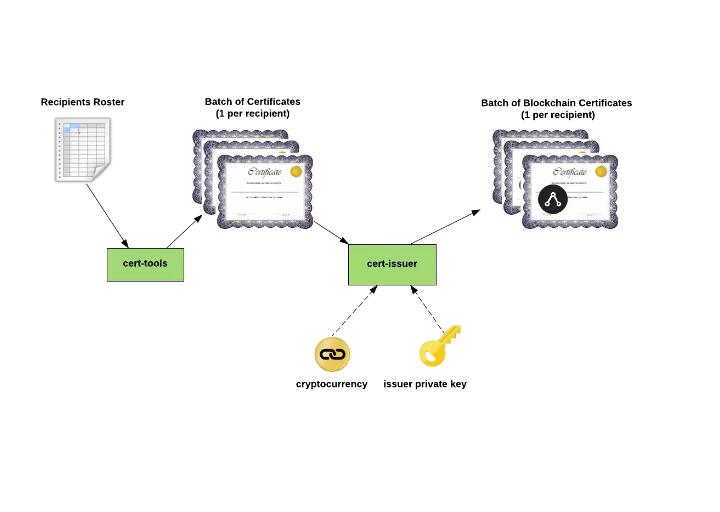
This sharing of certificate is performed by using a certificate wallet which is in simple terms an interface.

The school/university then issues the certificate through the blockchain. This is done by performing a transaction on the bitcoin blockchain. Along with the certificate the issues adds the hash or the digest of the certificate.

Upon receiving the certificate the recipient shares it with the Employer using a cert-wallet or even cert-viewer. cert-viewer is similar to the cert-wallet which acts an interface to view/display the certificate shared by the recipient.

To ensure the credibility of the certificate, the employer may use a third party application to verify the certificate. This is done using a cert-verifier which uses the digest or hash given by the issuer.

The below image describes how the cert-tools such as cert-waller, cert-display are used during a transaction. The symbol on the right most certificate would stand for the certificate with the digest or the hash. The below image demonstrates how the recipient requests for the issue of a certificate and the issuer encrypts with the blockchain hash. One of more thing to note here is that the digest is computed based on the contents of the certificate which makes it more authentic and verifiable.

**7. Known issues and Challenges**

* Investment for decentralised ledgers.
* Requires maintenance and synchronization between servers.
* Requires more computation and increases network size.
* More complicated than usual server technologies.
* Immutability isn’t always a boon.
* Transaction costs and network speed.

**8. Conclusion**

Blockchain is a technology that clearly has applications in the world of learning at the individual, institutional, group, national and international levels. It is relevant in all sorts of contexts: schools, colleges, universities, MOOCs, CPD, corporate, apprenticeships, and knowledge bases.

Rather than the old hierarchical structures, the technology becomes the focus, with trust migrating towards the technology, not the institutions. It is really is a disintermediation technology.

Traditionally institutions have been a source of trust: universities, for example, are trusted “brands”. In finance, where blockchain is nowadays a ubiquitous hot topic, banks exist to enact transactions, creating an environment in which blockchain’s advantages are readily obvious.

In education, however, there needs to be trust beyond the technology. We are looking, I think, at a hybrid model rather than a wholesale blockchain takeover. Reputation will still matter, and this will continue to be derived from the quality of the instruction, teachers, research, and so on. However, blockchain can play a role here, too, as one could imagine a sort of web of teachers and learners that deploys blockchain to cut out institutions. This, in my view, is not impossible, but it is unlikely.

It must also be recognized and conceded that blockchain is not without its problems. There are data-regulation issues, and a cloud has been created over the technology by the fact that one of the exchanges in the Bitcoin system – which is based on blockchain – saw $500 million disappear. And last but certainly not least, after considerable difficulty, US authorities were able to close down the infamous “Silk Road” drug-dealing exchange, which was also blockchain based.

Yet the biggest obstacle to blockchain’s more widespread use is cultural. Education is a slow learner and a very slow adopter. Despite its obvious advantages, the learning world is likely to be slow in implementing this technology, as most of the funding and culture is centred around the individual institution. Bologna was dead the day it was signed as nobody really wanted to lose their students and suffer financially, but it nonetheless became the framework for European higher education. This indicates clearly that the stimulus for change will have to come from elsewhere.

Despite the known issues and compromises from using Blockchain technology for certification, the technology is still in its development. As more researches advance, the technology can be optimized and be more widespread than it already is. Until it tackles sensitive issues, it can be used to solve some general and domestic problems.

**9. References:**

* Using Blockchain Technology to Prove Existence of a Document https://bravenewcoin.com/news/using-blockchain-technology-to-prove-existence-of-a-document/
* Malta Pilots Blockchain-Based Academic Certificate Recording System https://cointelegraph.com/news/malta-pilots-blockchain-based-academic-certificate-recording-system
* Sapiens http://sapiensproject.io/
* Blockcerts - The Open Initiative for Blockchain Certificates https://www.blockcerts.org/
* Using the blockchain as a digital signature scheme https://blog.signatura.co/using-the-blockchain-as-a-digital-signature-scheme-f584278ae826
* IPFS P2P File System, Juan Benet https://securityintelligence.com/why-blockchain-as-a-service-should-be-on-your-radar/