DocsChain: Blockchain based IoT Solution for Verficiation of Degree Documents

Saqib Rasool, Afshan Saleem, Muddesar Iqbal, Tasos Dagiuklas, Shahid Mumtaz and Zia ul Qayyum

Abstract—Degree verification is the process of verifying the academic credentials of successfully graduated students and universities annually spent millions of dollars for handling this process of degree verification. Hence, there is a dire need to minimize the degree verification cost and Massachusetts Institute of Technology has introduced the blockcerts, a blockchain based solution for freely handling the degree verification requests. Although Blockcerts eliminates the cost of degree verification but it also alters the existing workflow of degree issuance and verification. Blockerts is primarily focused on facilitating the students and there is a room for improvement from the prospective of educational institutes.
Contribution of this paper is to introduce the solution of DocsChain which is focused on facilitating both students and education institutes by maintaining the existing social workflow of degree issuance and verification. In contrast to Blockcerts, DocsChain allows the educational institutes for bulk submission of multiple degree documents. For the verification, it operates over the photocopies of the degree documents and collect digital information from these copies using the OCR (Optical Character Recognition) enabled IoT/WoT cameras. Extracted digital information is then passed to a one-way hashing algorithm to collect the equivalent hash which is then used for searching from the DocsChain using the concept of PoE (Proof of Existence).

Index Terms—DocsChain, Blockchain, PoE, OCR, Degree Verification, Document Verification

1 Introduction

Degree documents are issued after the completion of degree programs and are considered as the proof of completion of the corresponding degree programs. Hence, the employers use the degree documents for confirming the educational history of the applicants. Similarly, applying to other educational institutes, for higher studies, also requires the verification of degree documents. This pivotal role of degree documents attracts the scammers and motivates them to try securing jobs or admissions based on the fake degree documents. Hence, the employers or educational institutes need to verify the degree documents before accepting these and millions of dollars are annually spent, by the universities, for handling the verification requests of degree documents [1]. This results in an important challenge of simplifying the degree verification process along with reduction in cost.

Massachusetts Institute of Technology has introduced an open standard for academic credentials on the blockchain, known as the blockcerts [2]. Although blockcerts can be used for freely verifying the academic documents but it changes the existing workflow of degree issuance and verification. Students need to install the mobile application of blockcerts and also need to use it for maintaining a digital wallet of their degree documents and other certificates. Only then the blockerts will be able to verify their degree documents. Hence, the degree awarding institutes cannot place the data of a degree document on blockcerts, without involving the students.

This paper presents the DocsChain[1], a semi-private blockchain based solution that incorporates within the existing workflow of degree issuance and verification.

by focusing on the verification of the hard copies of degree documents. The process of degree issuance is executed by the degree awarding institutes and it results in the issuance of hard copies of degree documents to the students along with submitting its details to the DocsChain. When a student submits the copy of her degree documents to some organization, it can accomplish the degree verification in two different ways. Either it can request the degree awarding institute to verify the degree or it can deploy the CoT (Cloud of Things) client of DocsChain to independently verify the degree documents, without involving the degree-holding student.

According to the US federal law of FERPA (Family Educational Rights and Privacy Act) [3], it is not allowed to share the students' data without their permission. However, blockchain is based on a distributed ledger that is shared among all the participants and in order to make the FERPA compliant blockchain solution, it must restricts the availability of data stored on the distributed ledger of blockchain. PoE (Proof of Existence) is a popular solution to ensure the privacy of the data stored in the blockchain.

PoE ensures privacy by replacing the original data with an equivalent but irreversible hash. A one-way hashing algorithm is used for the said purpose and the output of that algorithm is used for representing the original data. There are many projects [4], [5], [6] that are using the blockchain-based PoE for various use cases. Both blockcerts [2] and

- *S. Rasool and A. Saleem work with University of Gujrat. E-mail: Saqib@ieee.io*
- *Muddesar Iqbal and Tasos Dagiuklas work with London South Bank University, UK*
- *Shahid Mumtaz works with Instituto de Telecomunicaes, Portugal*
- *Zia ul Qayyum works with Allama Iqbal Open University, Pakistan*

Manuscript received December 19, 2018; revised September 23, 2019.

1. www.DocsChain.org

our proposed solution of DocsChains are also based on the PoE for ensuring the data integrity without sacrificing the privacy of the actual data of the stored degree documents.

The contribution of this paper is threefold. First, it introduces a blockchain-based solution for degree verification which can be easily adopted within the existing fabric of the academic social system. Second, the participating institutes of DocsChain can add the data of a degree document without involving the input from the corresponding student. This is not possible in the existing solution of blockcerts where each student needs to create its blockchain address and then share it with the document issuing authority. Third, the academic institutes can simultaneously shift the data of both recently graduating and the already graduated students to the DocsChain.

Proof of Existence (PoE) requires a one-way hashing algorithm. It is an algorithm that takes different types of inputs and generates the output according to a single standard, which is also irreversible. For example, in the case of DocsChain, we are using the SHA-256 hashing algorithm [7]. As its name states, it always generates the output of 256 characters and if we are using the hexadecimal then the same output can be represented in 64 characters.

IoT camera automatically detects the edges of degree document using OpenCV [8] and then collects the image from the live stream. It then collects the text from the collected image using OCR. Optical Character Recognition (OCR) is a technique that can be used for extracting the content from an image. As every degree awarding institute may have its unique format for degree documents, therefore, each participating university has to contribute with one OCR program specialized for each degree format to ensure the proper data extraction from degree documents. Data extraction is the first step that degree awarding institutes need to perform after issuing the hard copies of degree documents.

IoT camera automatically detects the edges of degree document using OpenCV [8] and then collects the image from the live stream. It then collects the text from the collected image using OCR. Optical Character Recognition (OCR) is a technique that can be used for extracting the content from an image. As every degree awarding institute may have its unique format for degree documents, therefore, each participating university has to contribute with one OCR program specialized for each degree format to ensure the proper data extraction from degree documents. Data extraction is the first step that degree awarding institutes need to perform after issuing the hard copies of degree documents. This extracted data needs to pass to the one way hashing algorithm of SHA-256 to get the hashed value for that data. Each participating degree awarding institute is having its unique private key as a signature and it then adds its signature to the generated hashed value which is then stored in the blocks of DocsChain, as shown in Fig. 3.This whole process of collecting text from the degree document is executed by the IoT camera associated with the miner.

Section two elaborates the limitations of blockcerts and how the proposed solution of DocsChain solves that limitations. Section three covers the technical details of DocsChain. Section four presents the effectiveness of our proposed architecture against three different scenarios of configuration. Section five covers the related work and the last section concludes this paper along with some possible future extensions.

2 Related Work

To the best of our knowledge, no research work has integrated the OCR within the blockchain to secure the hard copy of documents through the blockchain. There are efforts available for integrating the educational certificates within the blockchain [9] [2]. There is also a project which secures the digital form of degree documents through blockchain. This project is known as blockcerts and is currently being experimented at MIT [2]. However, as it is not including the hard copies of degree documents therefore, a change is required in the workflow of degree assurance and verification.

Through DocsChain, we have incorporated the hard copies of degree documents within the blockchain and therefore, there is no need to change the existing workflow of degree issuance and verification. Even the already awarded degrees can also be verified with the DocsChain which is not possible in blockcerts. This is because degree awarding institutes are already maintaining a record of previously issued degrees and they just need to pass the degrees through OCR and these will become part of the DocsChain.

To the best of our knowledge, there is no existing semi-public blockchain available and DocsChain is the first semi-public blockchain platform. The concept of a semi-private blockchain platform has been already introduced for the addition of more consortium members by the centralized authority. Few efforts of security and privacy are also available for controlling the access to the data stored in blockchain [10], [11], [12], [13]. We have also included two access levels for achieving the security and have used the PoE for achieving privacy in a very simple way by replacing original data with its equivalent hash value.

Many studies have been conducted for using the camera-based devices to perform the degree verification [14], [15], [16] and document verification [17], [18], [19], [20], [21]. However, none of the existing research work has combined the blockchain with the camera to achieve the aim of degree verification. This is the first research study that combined the IoT camera and blockchain for ensuring the authenticity of the degree documents.

3 Incorporation of Blockchain within the Social Workflow of Degree Processing

A social workflow originates from the interactions of a group of people that are working on personal tasks or data [22]. Social workflow of degree issuance and verification involves the interaction of three types of entities, viz. issuer, recipient and verifier. Fig. 1 and Fig. 2 shows an institute (working as issuer) that issues the degree documents to the students (acting as the recipient). Students submit copies of the received documents to the employer (labelled as verifier) that are supposed to verify the received documents. This section elaborates the differences in the interactions of three discussed entities for both blockerts and DocsChain.

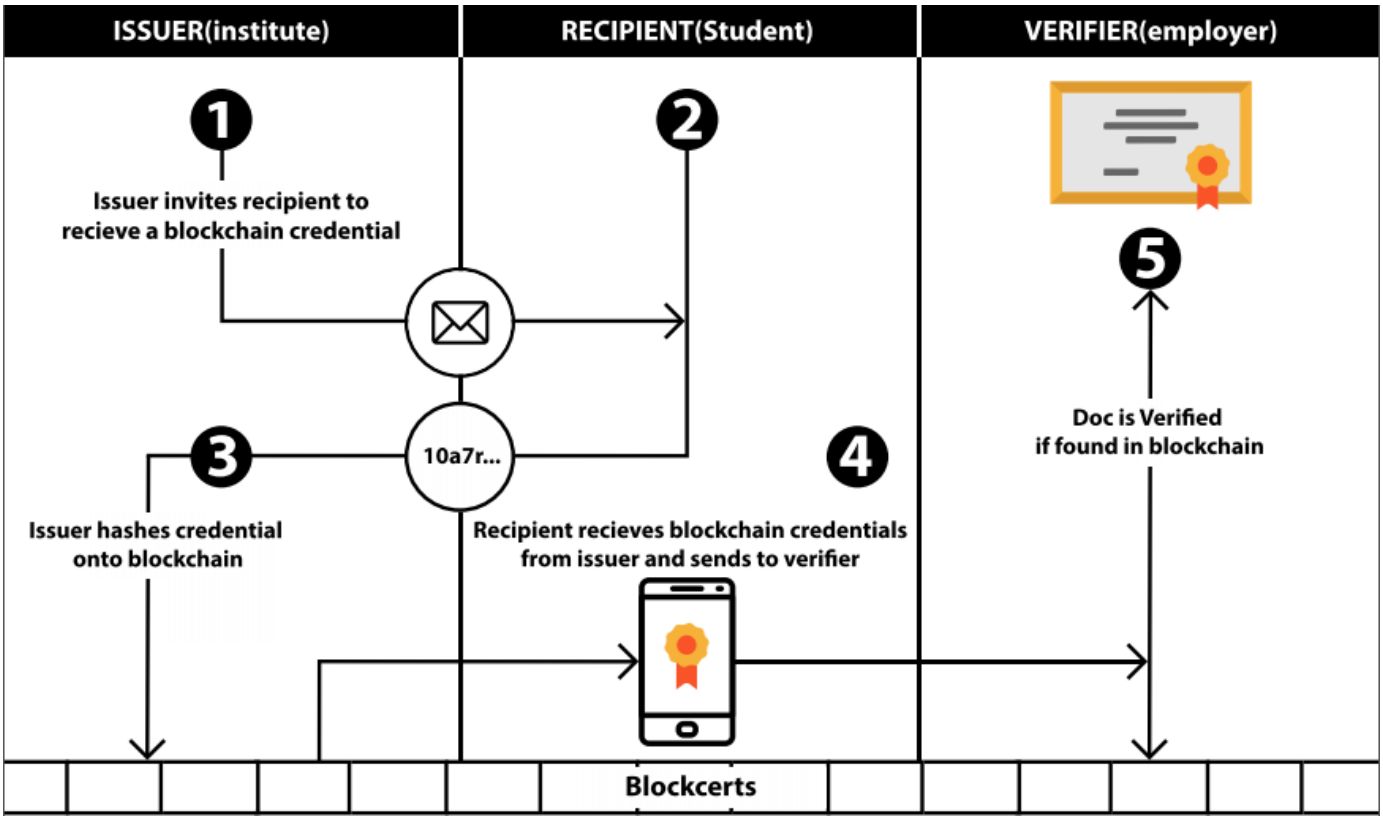

Fig. 1. Workflow of degree issuance and verification through the existing solution of blockcerts

3.1 Workflow of Degree Issuance and Verification through Blockcerts

Fig. 1 shows the workflow of degree issuance and verification through the blockchain of blockcerts. Students are the main focus of the blockcerts as it requires the students to maintain a digital wallet through the mobile application of blockcerts. Following is the explanation of each of the steps mentioned in Fig 1:

1) Institute invites the students to create an account on blockcerts and share their credentials with the institute. However, if institute is not already using the application of blockcerts then students can also initiate the process by inviting the institute to join the blockcerts.
2) Students create their accounts on blockcerts and share their credentials with their institute.
3) An institute uses the received credentials of the student for mapping the data of her degree documents against these credentials and then stores this information on the blockchain of the blockcerts.
4) Student can use the mobile application of blockcerts to retrieve the information stored on blockchain and can share details of her degree/degrees with the employer through their credentials.
5) Employer can use the provided details of degree certificate/certificates for free of cost verification through the blockcerts.

Blockcerts is an effective way of providing free of cost degree issuance and verification solution. However, it has the following limitations for the three main participants of the social workflow of degree issuance and verification:

- **Institutes** face the limitation of being tightly bound with students because the data of degree documents cannot be placed on blockcerts, without involving the students. This dependency on students becomes more drastic when an institute decides to place the already issued degrees of alumni on the blockcerts. This is because it is difficult to trace and convince the already passed-out students for taking part in the process of placing the data of their degree documents on the blockchain. Another limitation of blockcerts is that the institutes can place only a single document per transaction and there is no option available for the data submission in bulk. DocsChain provides solution to all of the discussed problems of institutes by eliminating the need of involving students during the submission of degree documents to the blockchain.
- **Students** have no option of using the hard copies of their degree documents and are bound to use the mobile application of blockcerts. DocsChains solves this problem by adding the support of hard copies of degree documents. Hence, students no longer need to install any mobile application. They can simply take the hard copy of degree document from the institute and can email its scanned image to the employees or can even give the photocopy of the degree documents to the employees.
- **Employees** do not have the option of verifying the scanned copies or photocopies of the degree documents through the blockcerts. Our proposed solution of DocsChain provides this feature by applying the OCR over the scanned copies or photocopies of degree documents to collect their digital representation.

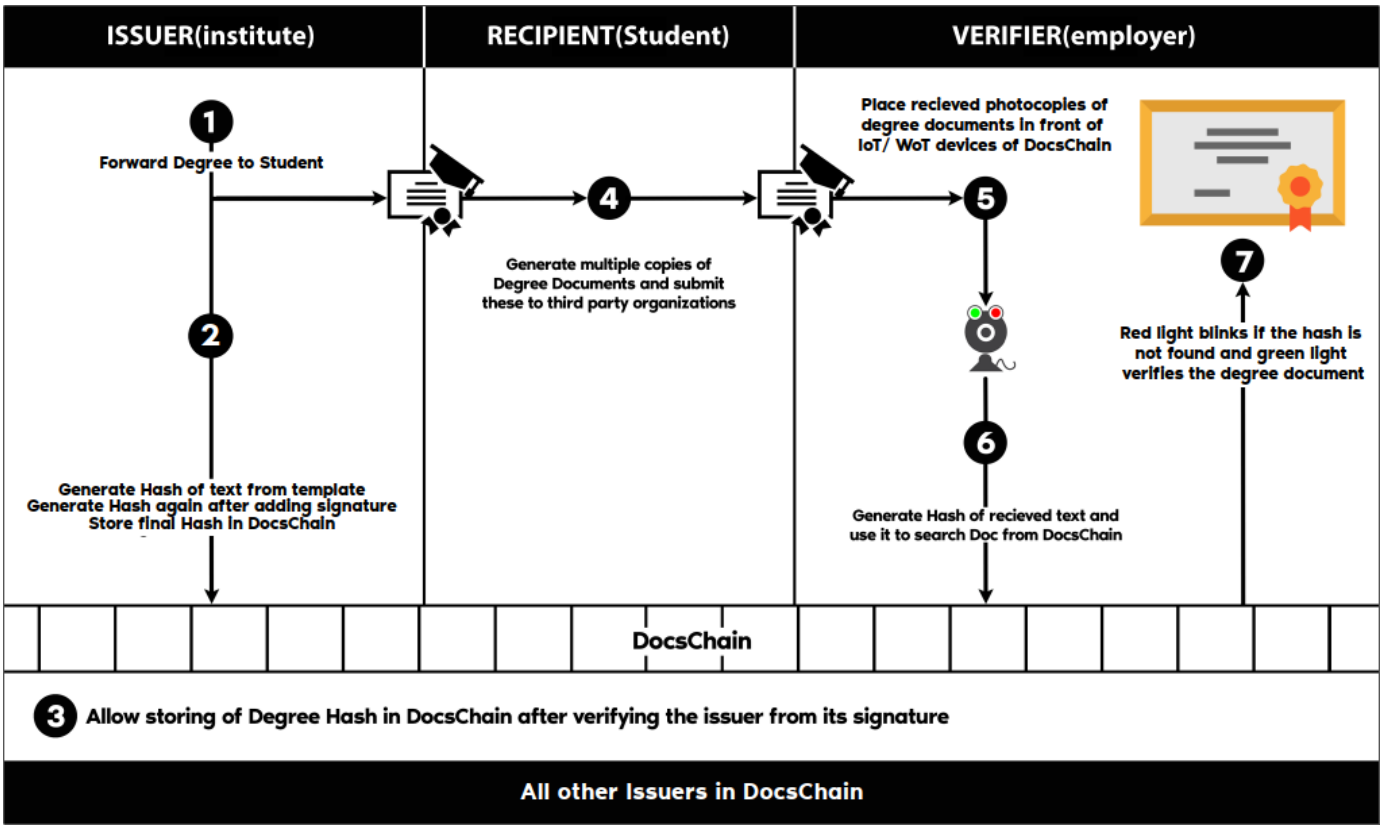

Fig. 2. Workflow of degree issuance and verification through the proposed solution of DocsChain

3.2 Workflow of Degree Issuance and Verification through Proposed Solution of DocsChain

DocsChain allows the institutes to issue the degree documents without involving the students and it solves the discussed limitations of blockcerts for institutions. DocsChain also operates over the hard copies of degree documents and it results in the overcoming the mentioned limitations of blockcerts for students. Fig. 2 shows the workflow of degree issuance and verification through DocsChain and each of its steps are discussed below:

1) Degree documents are forwarded to the students by the degree awarding institute and the scanned image of the same degree document is forwarded to the OCR library for extracting the details of that degree.
2) Data extracted from OCR is added with four other values of 1) issuer number, 2) private key of the issuer 3) OCR library number and 4) the hash of the previous block of DocsChain. All of these values are mentioned in Fig. 3 and have been already explained in section 2. The hashed value of the complete block generated from step 2 is then placed in the block and is forwarded to all other participating degree awarding institutes.
3) All other degree-awarding institutes identify the public key of the block creator (or degree issuers) from the issuer number and use this to validate if the block is submitted by the authorized degree awarding institute. In case of consensus among all the participating degree awarding institutes, a new block is added to the DocsChain.
4) Students can take multiple copies of the received degree documents and can submit these to multiple organizations for higher studies or jobs etc.
5) Verifier receives the hard copy of the degree document from the student and identifies the issuer number from the institute name. It then has to identify the OCR library number from the program of the degree as all degrees of the same program shares the same OCR library. After identifying the OCR library, the verifier takes the scanned copy of the student's degree and passes it to the related OCR library.
6) OCR library extracts the data from the degree document and generates a hash code from the extracted data using the SHA-256 hashing algorithm.
7) Hash code generated from step 6 is then searched from the DocsChain. The degree is verified if the passed hashed code is found on DocsChain and vice versa.

In comparison to blockcerts, DocsChain has three main improvements in the workflow of degree verification. The first improvement is that DocsChain supports the conventional hard copy of degree documents while blockcerts only supports the digital copy of the certificates. The second improvement is that the workflow of degree verification through DocsChain is independent of the student while blockcerts cannot operate without students. The third improvement is that DocsChain can also verify the degree documents that have been already issued while blockcerts cannot verify the documents that are not issued through it.

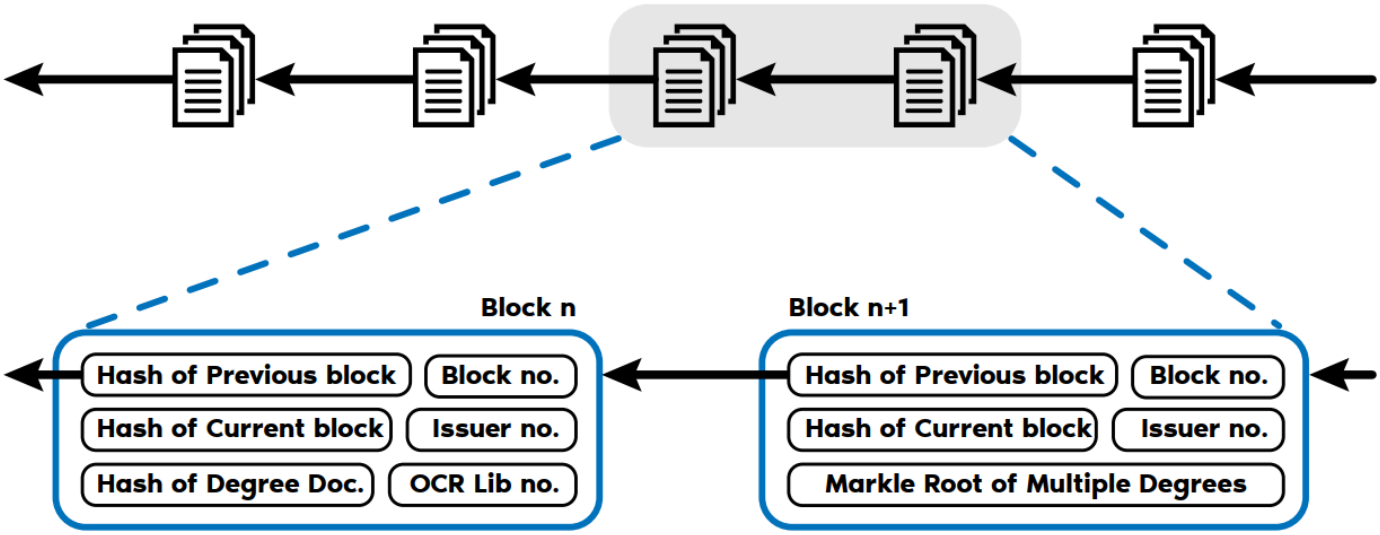

Fig. 3. DocsChain with two blocks having different number of degrees

4 Blockchain of DocsChain

This section focuses on the blockchain nature of the DocsChain and explores the features of DocsChain that are derived from the blockchain. Blockchain [23] has already been proven as an effective solution for removing the dependency on a single entity and distributing the authority among multiple independent entities. Bitcoin is the first and most popular application based on the blockchain [24]. There are many important features in the blockchain that make it unique and effective in comparison to other application development techniques and following are some of the salient blockchain relevant strengths of the DocsChain:

- **REST Services** are used for exposing all the features of the DocsChain. It allows the integration of multiple types of client applications with DocsChain. We have integrated the IoT/WoT cameras as clients but same REST services can be integrated within the mobile applications to access DocsChain features from mobile devices.
- **PoE** uses the one-way hashing algorithm that produces the Irreversible output therefore, it ensures the privacy of the data passed as input. It also ensures the integrity of input data as even a minor change in input results in the change of the output value. Hence, PoE not only ensures the privacy of the data but also maintains its integrity.
- **Autonomy** is the flexibility of decision making wherever its possible [25] [26]. DocsChain supports autonomy by allowing the institutes to add their specific OCR templates and degree documents. DocsChain also allows the organizations to operate the camera with different configurations according to their needs (more details in section 6.4).
- **Openness** is an important feature for supporting transparent coordination among multiple institutes participating as the consortium members of the DocsChain. This feature is supported by sharing a distributed ledger among all the institutes operating under the DocsChain.
- **Trust** needs to be established among the mutually cooperating institutes of DocsChain. Step 3 of Fig. 2 shows a feature named as consensus algorithm [27] which establishes the trust among the particpants of the DocsChain.

4.1 Structure of a Block in Distributed Ledger of DocsChain

Hashed value of extracted data of degree document along with the signature becomes a single block of DocsChain. Each block of DocsChain may contain 1 to N number of documents to accommodate the variable number of degree documents against each graduating batch. Moreover, each block must belong to only a single degree awarding institute. The signature of private key exists in pair with a public key [28] which is used by the other degree-awarding institutes for confirming the access to create new blocks in DocsChain. Fig. 3 shows the basic structure of two blocks in DocsChain with the different number of degree documents. Each block contains a minimum of six values that are explained below:

- **Hash of the previous block** is used to link multiple blocks of DocsChain in an immutable manner.
- **Hash of current block** is used to verify the integrity of the block as it is collected after passign all the data of a block through a one-way hashing algorithm.
- **Hash of Degree Document** is used for verifying the existence of a degree document by any third-party verifier using the concept of PoE.
- **Block number** is used for identifying multiple blocks in the DocsChain.
- **Institute** contains the name of degree awarding institute and it is unique for every institute. These name from each block are collected during the first step mentioned in Fig. 5.
- **OCR template** is based on multiple segments which refers to the information sections on a hard copy of a degree document. Atleast one OCR template is required against a degree awarding institute and any number of templates can be created by the institute where each template refers to a different format of degree documents. Each segment of the OCR template also contain the constraints for data collected for that segment (like the roll number segment shows the format of the roll number).

Fig. 4. Blockchain platforms with respect to Access and Ownership

4.2 Semi-private Blockchain of DocsChain

In contrast to the database, which supports the CRUD (Create, Read, Update, Delete) operations, blockchain only supports the create and read operations. Blockchain also restricts the create operations to special devices known as miners. However, read operation can be performed by both miners and normal users of the blockchain. Based on the access to create and read operations, blockchains can be broadly categorized into permissioned and permissionless blockchain platforms.

Permissionless [29] blockchain platforms are open for everyone to join and are known as public blockchain platforms. Permissioned [30] blockchain platforms need some sort of validation before allowing the access and these can be further divided into consortium [31] and private [32] blockchain platforms. Hence the division of permissioned and permissionless blockchain can be discussed in terms of three types of platforms [33], viz. 1) public blockchain, 2) private blockchain, and 3) consortium blockchain. In private blockchain platforms, only a single member controls the access to the blockchain. It is very much similar to the sharing of a read-only copy by the owner of the database with multiple parties. However, in the consortium blockchain platforms, a group of preselected members or the co-founders control the blockchain through the mutual consensus.

Consortium blockchains are also known as federated blockchains and by default, the co-founding members of a consortium control the blockchain. They also control the access of other members to the blockchain. However, there is a variation in consortium blockchain where more consortium members can be added by a centralized entity and this is known as a semi-private blockchain [34], [35], [36]. This name of semi-private is derived from the fact that in private blockchain a centralized entity controls the blockchain. Although in semi-private blockchain, the control is given to the consortium members but these members are selected by the central entity, therefore, the main control of blockchain is given to the central entity but all other participants can ensure the transparency of the semi-private blockchain.

We have implemented the DocsChain as a semi-private blockchain. Fig. 4 summarizes the categories of blockchain platforms with respect to two important factors of access and ownership. All listed categories in Fig. 4 are further evaluated against both of the mentioned factors below:

- **Public** blockchain platforms give open access to all and also give equal preference to all participating members through a public consensus.
- **Consortium** blockchain platforms give variable access to the members depending upon the custom policies. However, the ownership is only limited to the co-founders of the consortium and the addition of new consortium members is not allowed.
- **Semi-private** blockchain platforms also give variable access to the members based on the implementation of customs policies. Ownership is also limited to the consortium members but new members can be added by the central authority. Although, semi-private blockchains are not directly owned by the central authority more control is given to the central authority through the right of selecting the consortium members.
- **Private** blockchain platforms provide the ownership to the central entity and the same entity is responsible for allowing access to the individuals. It is more likely a cryptographically secured and shared database with limited access to increase the trust level.

4.3 Data Handling in DocsChain

Blockchain of DocsChain supports only create and read operations. Create operation is restricted only to the degree awarding institutes while the read operation is publicly available to everyone. Both of these read and write operations are exposed through the REST services. Following are the REST services for adding new data in DocsChain while next section describes the degree verification process through the data retrieval operations in detail.

- **Operation of adding a new institute** does not add any data in the ledger and is only done with the permission of admin (see semi-private blockchain for details). Admin generates a pair of public-private key and shares the private key with the newly added institute while distribute the public key to all the miners for validating the creation of new degree documents by the newly added institute. Admin only allows the addition of new institute after confirming the availability of a valid OCR template submitted by the institute along with the corresponding degree formats. This is to ensure the existence of a degree awarding institute only with a valid OCR template.
- **Operation of adding a new OCR template** also does not add any information into the distributed ledger of DocsChain and this operation is only accomplished after the verification of that OCR template by the admin. This is to ensure that the provided OCR template can accurately retrieve the information during the degree verification process. Once the admin verifies the accuracy of the OCR template against the provided formats of the corresponding degree document, this OCR template is forwarded to all the miners that stores the newly added OCR templates in a temporary queue.
- **Operation of adding new degree documents** only adds the data to the distributed ledger of the DocsChain (mentioned in step 3 of Fig. 2). Whenever a new degree document is submitted, it is first verified from the signature of the degree awarding institute. If the signature is found valid then the verification of the provided OCR template/templates is done by searching for the same template/templates in the previously submitted blocks or from the temporary array of templates.

5 Performance of Degree Verification

This section explains the experimental setup for finding the details of verification results by each IoT camera. It also elaborates on the results collected from the experiments.

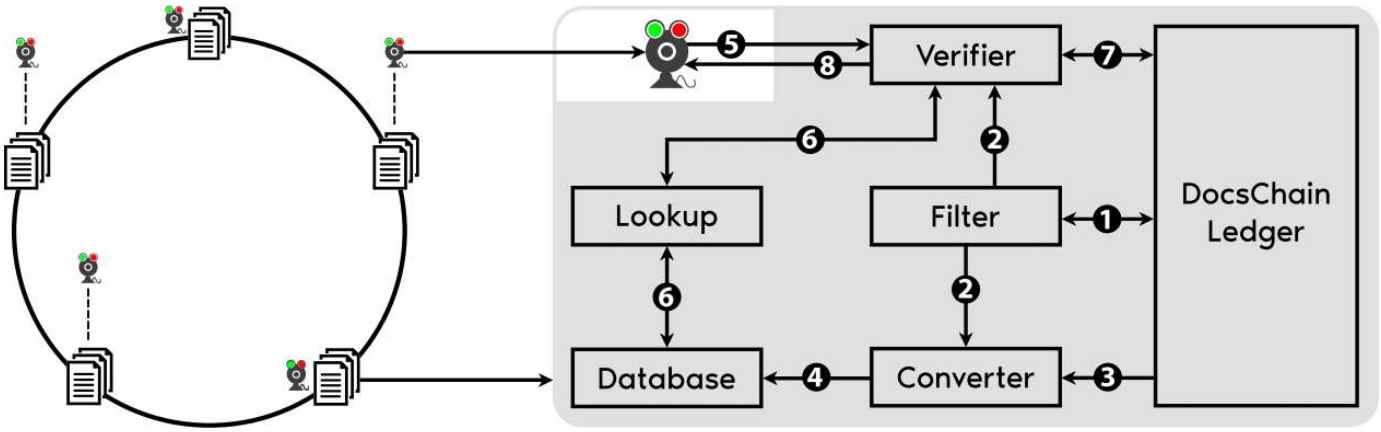

Fig. 5. Steps of degree verification through IoT and WoT Cameras

5.1 Process of Degree Verification

Fig. 5 elaborates different steps of degree verification process and each of these steps are covered below in detail:

1) Initially the *filter* module is invoked and it contains two REST endpoints. First REST endpoint returns all the institutes available in the DocsChain and second REST endpoints takes a list of institutes (1 to N) as input and returns the OCR templates of all of the given institutes.
2) *Filter* then forwards the collected information to both *converter* module and the *verifier* module. Populating both of these modules through same *filter* configurations ensures the correct work of these dependent modules.
3) *Converter* module picks the data from the *ledger* and perform the proper indexing to optimize the lookup time for the collected data.
4) The *Converter* then stores the indexed data into a *database* for reducing the lookup time. It also stores the block number of distributed ledger against each value to help in reducing the lookup time in step number 7.
5) Whenever the *camera* captures the image of a degree document, it forwards it to the *verifier* module for the verification of that degree document. The *verifier* module has already collected the list of available institutes along with their template libraries in step 2. The *verifier* is also equipped with an OCR which tries to match the name of degree awarding institute from the already maintained list of institutes. If the institute is not found then it skips the steps 6 and 7 and directly reach to the step 8, which indicates the non-verified degree document through the blinking of red light.
6) This step executes after the identification of degree awarding institute from the degree document. Before reaching to the lookup module, the OCR applies all of the available templates of the identified institute and select the information of a single template which matches the maximum constraints of each template component (details of these constrains have already been covered in sub-section 5.1). The *verifier* module passes the information collected through the OCR template to the *lookup* module which generates the one-way hash of the collected information and then search that hash from the database. Hash will not be found even if a single character is modified from the degree document and the *verifier* will skip the step 7 by directly moving to step 8 and the camera will show the non-validity of degree document through its red light.
7) This step will only execute if the hash (generated from the data collected by the OCR) has been found in the database. This step will reverify the existence of the found hash from the DocsChain ledger.
8) If the generated hash is found in the DocsChain ledger then the green light of camera will blink to show the successful verification status of degree document. If the hash is not found in the ledger then red light of camera indicates the failure in degree verification.

5.2 WoT Camera vs IoT Camera and Ad-hoc Cloud

Fig. 5 five copies of the distributed ledger of DocsChain and each of these copies contains one of two types of cameras, viz. IoT cameras and WoT cameras. The cameras that are connected with the dotted lines are the IoT cameras and there are three IoT cameras in Fig. 5. The cameras that are directly connected with the DocsChain ledger are the WoT cameras and there are two WoT cameras in the Fig. 5.

IoT and WoT can be characterised by the availability of the computational resources. The WoT devices are resource rich devices that can perform the CPU intensive tasks while the IoT devices are resource constrained devices and need special architectures and protocols for reducing the resource consumption [37]. We have also used different architectures for both IoT and WoT cameras.

There are two arrows extending from the network of five distributed ledgers of DocsChain to the eight steps of verification process. Bottom arrow is for the WoT camera that hosts all the components shown with dark background. Above arrow is for the IoT camera and it does not host any of the components with the dark background. Hence, IoT camera only captures the image and forwards it to some other machine for computation offloading. This other machine can be a dedicated system or some already running machine/machines can also be used for offloading the computation of IoT camera. A group of devices that are used for computation but are not dedicated for that purpose can be termed as an ad-hoc cloud [38] [39] [40] and IoT camera can use the ad-hoc cloud for offloading its computation.

5.3 Frequency of Lookups from Distributed Ledger of DocsChain

Fig. 5 indicates three steps that collects the data from the distributed ledger of the DocsChain. Execution frequency for each of these steps vary and following is the detail of the lookup frequence by each of these steps:

- **Step 1** picks the list of targeted degree awarding institutes and it only needs to be executed once or only after the addition of a new degree awarding institute in the DocsChain.
- **Step 3** picks all the data of DocsChain and store it into the database. It needs to be executed periodically so that the maintained database can remain synchronized with the ledger of DocsChain and the rate of this periodic execution is selected by the organization deploying the degree verification solution. Whenever the *converter* module finds a newly added degree awarding institute, it also executes the *filter* module.
- **Step 7** reverifies the hash of a degree document found in the database. Hence, it executes every time when a degree document has been verified by the database and the final confirmation is required from the DocsChain ledger.

5.4 Frequency of OCR Iterations in the Verifier Module

Frequency of OCR iterations is the main contributor in determining the performance of the *verifier* module. Hence, the verification module of DocsChain has been designed in a configurable way so that the OCR iterations can be controlled according the requirements. For the said purpose, the organization which is planning to use DocsChain for the verification of degree documents can chose the targeted institutes and the *filter* module will load only the names of the targeted degree awarding institutes along with the OCR templates of each of these institutes.

Initially, the single iteration of OCR is used for finding the degree awarding institute and then one OCR iteration is required for each of the OCR template of the identified degree awarding institute. Based on the these details, total number of OCR iterations can be summarized through following three scenarios:

- **1 OCR iteration** is required when the targeted degree does not match any of the degree awarding institutes that are picked by the *filter* module. Hence, degree will not be verified.
- **N OCR iterations** are required for extracting the digital information from the image if a degree document if the DocsChain is configured for verifying the degree documents of a single degree awarding institute. Here, N referes to the total number of templates of the targeted degree awarding institute. As all the provided degrees are from the same degree awarding institute therefore, no first iteration is required for finding the institute and only one OCR iteration is required against each of the template of the degree awarding institute.
- **N + 1 OCR iterations** are required when DocsChain is running for verifying the degree documents of multiple degree awarding institutes where one iteration is used for finding the institute of targeted degree and N refers to the number of OCR templates of the identified degree awarding institute.

6 Conclusion and Future Work

This paper presents a semi-private blockchain based degree verification solution which fits within the existing social fabric of students and degree awarding institutes. In contrast to the existing solution of blockcerts, it allows the universities for bulk submission of degrees in DocsChain without involving the students. It also enables the verification from the photocopies of all the degrees that have been issued by an institute till date. To the best of our knowledge, there is no existing blockchain based degree verification system that operates over the hardcopies of the degree documents.

DocsChain uses an IoT camera for converting the hard copies of degree documents into a digital equivalent. It then uses the OCR for extracting data from that digital representation of hard copies of degree documents. OCR returns the data of the degree document which is then passed through one-way hashing algorithm for preserving the privacy of the data and the resultant hash is then matched from the DocsChain for confirming the authentication of targeted degree document. As DocsChain has started a new trend in the blockchain based degree verification solution therefore, it can be enhanced in many different ways. Following are some of the proposed future extensions for DocsChain:

1) As the DocsChain only operates over black and white images therefore, if a colored photocopy of a degree document is provided to the DocsChain, it first coverts it to the black and white image and then performs the verification process. An option of verification of colored photocopy can be provided which will result in including more options (like logo) for the segments of OCR templates.
2) Proposed solution of DocsChain only operates over the textual information. It can be extended to allow the degree awarding institutes to store the logos on the DocsChain and image processing techniques can be used for identifying the institute through its logo.
3) Current version of DocsChain is based on REST services but it can be extended to offer other interfaces like GraphQL, gRPC etc.
4) We have evaluated DocsChain for the verification of degree documents only. However, the same idea of OCR based blockchain can be applied for other use cases as well.
5) DocsChain is based on the idea of PoE. A more advanced attribute level access control mechanism can also be added in the DocsChain.
6) A detailed evaluation study needs to be conducted, after involving the real stakeholders, for reporting the different aspects of DocsChain in detail.
7) A mobile application can be implemented which uses the camera of mobile phone to replace the dependency on IoT camera in DocsChain and allows to perform the degree verification directly from a mobile device.
8) Currently DocsChain collectively uses the data from all the segments of the OCR template for finding the accuracy of degree documents. However, if data of all segments of the OCR template is preserved in the DocsChain and the verification module can independently validate each segment then the exact information can be identified which is being fabricated to fake the documents.

REFERENCES

[1] N. K. Bajwa, "Modelling and simulation of blockchain based education system," Ph.D. dissertation, Concordia University, 2018.

[2] J. Hope, "Issue secure digital credentials using technology behind bitcoin," *The Successful Registrar*, vol. 17, no. 11, pp. 1–4, 2018.

[3] C. A. Ramirez, *FERPA clear and simple: The college professional's guide to compliance*. John Wiley & Sons, 2009.

[4] P. Snow, B. Deery, J. Lu, D. Johnston, and P. Kirby, "Factom business processes secured by immutable audit trails on the blockchain," *Whitepaper, Factom, November*, 2014.

[5] M. Araoz, "Proof of existence," *En ligne]. Available: proofofexistence. com*, 2013.

[6] B. Gipp, N. Meuschke, and A. Gernandt, "Decentralized trusted timestamping using the crypto currency bitcoin," *arXiv preprint arXiv:1502.04015*, 2015.

[7] T. Xue, Y. Yuan, Z. Ahmed, K. Moniz, G. Cao, and C. Wang, "Proof of contribution: A modification of proof of work to increase mining efficiency," in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1. IEEE, 2018, pp. 636–644.

[8] A. Kaehler and G. Bradski, *Learning OpenCV 3: computer vision in C++ with the OpenCV library*. " O'Reilly Media, Inc.", 2016.

[9] M. Sharples and J. Domingue, "The blockchain and kudos: A distributed system for educational record, reputation and reward," in *European Conference on Technology Enhanced Learning*. Springer, 2016, pp. 490–496.

[10] R. Guo, H. Shi, Q. Zhao, and D. Zheng, "Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems," *IEEE Access*, vol. 776, no. 99, pp. 1–12, 2018.

[11] I. Sukhodolskiy and S. Zapechnikov, "A blockchain-based access control system for cloud storage," in *Young Researchers in Electrical and Electronic Engineering (EIConRus), 2018 IEEE Conference of Russian*. IEEE, 2018, pp. 1575–1578.

[12] X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat, and L. Njilla, "Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability," in *Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*. IEEE Press, 2017, pp. 468–477.

[13] D. D. F. Maesa, P. Mori, and L. Ricci, "Blockchain based access control," in *IFIP International Conference on Distributed Applications and Interoperable Systems*. Springer, 2017, pp. 206–220.

[14] H. O. Ochieng, "A mobile based application for verification of legitimacy of degree certificates in kenya," Ph.D. dissertation, Strathmore University, 2016.

[15] G.-h. Wu, L. He, X.-w. Wang, and K. Wei, "Study on integrated method of files management based on cooperative oa," in *Internet Technology and Applications, 2010 International Conference on*. IEEE, 2010, pp. 1–5.

[16] B. Hall, M. Thompson, T. Huynh, and W. Kern, "Method and system for online third-party authentication of identity attributes," Sep. 19 2013, uS Patent App. 13/799,997.

[17] R. VS, V. C. Raj, S. Eswaran, and S. RU, "Optimization of digitalized document verification using e-governance service delivery platform (e-sdp)," *International Journal of Applied Engineering Research*, vol. 11, no. 4, pp. 2531–2539, 2016.

[18] D. Saha, S. Sonar, P. Telore, and L. Jadhav, "Secured document generation and authentication mechanism using vss and qr code," 2016.

[19] S. Eskenazi, P. Gomez-Krämer, and J.-M. Ogier, "When document security brings new challenges to document analysis," in *Computational Forensics*. Springer, 2015, pp. 104–116.

[20] R. Jain, "Searching heterogeneous document image collections," Ph.D. dissertation, 2015.

[21] F. Cruz, N. Sidere, M. Coustaty, V. P. D'Andecy, and J.-M. Ogier, "Local binary patterns for document forgery detection," in *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, vol. 1. IEEE, 2017, pp. 1223–1228.

[22] S. Görg and R. Bergmann, "Social workflowsvision and potential study," *Information Systems*, vol. 50, pp. 1–19, 2015.

[23] S. Underwood, "Blockchain beyond bitcoin," *Communications of the ACM*, vol. 59, no. 11, pp. 15–17, 2016.

[24] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[25] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin, and S. Chen, "Vehicular fog computing: A viewpoint of vehicles as the infrastructures," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3860–3873, 2016.

[26] P. Garcia Lopez, A. Montresor, D. Epema, A. Datta, T. Higashino, A. Iamnitchi, M. Barcellos, P. Felber, and E. Riviere, "Edge-centric computing: Vision and challenges," *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 5, pp. 37–42, 2015.

[27] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in *Big Data (BigData Congress), 2017 IEEE International Congress on*. IEEE, 2017, pp. 557–564.

[28] M. Pilkington, "11 blockchain technology: principles and applications," *Research handbook on digital transformations*, p. 225, 2016.

[29] Z. Jiang, B. Krishnamachari, S. Zhou, and Z. Niu, "Senate: A permissionless byzantine consensus protocol in wireless networks," *arXiv preprint arXiv:1803.08694*, 2018.

[30] T. Sato and Y. Himura, "Smart-contract based system operations for permissioned blockchain," in *New Technologies, Mobility and Security (NTMS), 2018 9th IFIP International Conference on*. IEEE, 2018, pp. 1–6.

[31] J. Gu, B. Sun, X. Du, J. Wang, Y. Zhuang, and Z. Wang, "Consortium blockchain-based malware detection in mobile devices," *IEEE Access*, vol. 6, pp. 12 118–12 128, 2018.

[32] E. Mengelkamp, B. Notheisen, C. Beer, D. Dauer, and C. Weinhardt, "A blockchain-based smart grid: towards sustainable local energy markets," *Computer Science-Research and Development*, vol. 33, no. 1-2, pp. 207–214, 2018.

[33] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in *Big Data (BigData Congress), 2017 IEEE International Congress on*. IEEE, 2017, pp. 557–564.

[34] "Startup management understanding semi-private blockchain applications," http://startupmanagement.org/2016/11/06/understanding-semi-private-blockchain-applications/, (Accessed on 04/21/2018).

[35] D. Guegan, "Public blockchain versus private blockhain," 2017.

[36] E. C. Ferrer, T. Hardjono *et al.*, "Robochain: A secure data-sharing framework for human-robot interaction," *arXiv preprint arXiv:1802.04480*, 2018.

[37] S. Rasool, R. Khan, and A. N. Mian, "Graphql and dc-wsn-based cloud of things," *IT Professional*, vol. 21, no. 1, pp. 59–66, 2019.

[38] S. Rasool, M. Iqbal, T. Dagiuklas, Z. Ul-Qayyum, and S. Li, "Reliable data analysis through blockchain based crowdsourcing in mobile ad-hoc cloud," *Mobile Networks and Applications*, pp. 1–11, 2019.

[39] D. M. Shila, W. Shen, Y. Cheng, X. Tian, and X. S. Shen, "Amcloud: Toward a secure autonomic mobile ad hoc cloud computing system," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 74–81, 2017.

[40] M. A. Al Mamun, K. Anam, M. F. A. Onik, and A. Esfar-E-Alam, "Deployment of cloud computing into vanet to create ad hoc cloud network architecture," in *Proceedings of the world congress on engineering and computer science*, vol. 1, 2012, pp. 24–26.

Saqib Rasool holds an MS degree in Computer Science from the National University of Science and Technology (NUST), Islamabad, Pakistan. He is currently pursuing PhD studies and is also serving as senior lecturer at Department of Computer Science, University of Gujrat (UoG), Gujrat, Pakistan. His research interests are Blockchain, Internet/Web/Cloud of Things, Reflection and Meta-programming, Declarative DSLs, DevOps, and scalable cloud services using Kubernetes etc.

Afshan Saleem holds MSc in IT and is currently working as a project manager at Nexthon Technologies, Gujrat, Pakistan

Dr. Muddesar Iqbal is Senior Lecturer in Mobile Computing in the Division of Computer Science and Informatics, School of Engineering. He is an established researcher and expert in the fields of 5G networking technologies, multimedia cloud computing, mobile edge computing, fog computing, Internet of Things, software-defined networking, network function virtualization, quality of experience, and cloud infrastructures and services.

Tasos Dagiuklas is the group leader of the SuITE research group and also heading the Division of CS and Informatics at London South Bank University, UK.

Dr. Shahid Mumtaz received his M.Sc. degree from the Blekinge Institute of Technology, Sweden, and his Ph.D. degree from the University of Aveiro, Portugal. He is now a senior research engineer at the Instituto de Telecomunicaes, Plo de Aveiro, Portugal, working in EU funded projects. His research interests include MIMO techniques, multi-hop relaying communication, cooperative techniques, cognitive radios, game theory, energy-efficient framework for 4G, position information assisted communication, and joint PHY and MAC layer optimization in the LTE standard. He is the author of several conferences, journals, and books