Docschain: Blockchain-based IoT Solution for Verficiation of Degree Documents

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

Abstract—Degree verification is the process of verifying the academic credentials of successfully graduated students. It is a time-consuming and costly process as universities annually spent millions of dollars for handling the degree verification requests. Hence, there is a dire need to improve the degree verification process and the 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 the degree verification process, it also alters the existing workflow of degree issuance. This is because blockcerts is primarily focused on facilitating the students and there is a room for improvement from the perspective of educational institutes.
In this paper, we have introduced the docschain to tackle three mentioned limitations of the blockcerts. docschain seamlessly incorporates within the existing workflow of degree issuance by operating over the hard copies of the degree documents. This is achieved through OCR (Optical Character Recognition) and the record of each degree document is stored along with the details of the corresponding OCR template to understand the semantics of the data stored at different sections of the degree document. In contrast to blockcerts, docschain also supports the bulk submission of degree details for both the previously and newly graduated students.

Index Terms—Docschain, Blockchain, PoE, OCR, Degree Verification, Document Verification, atDegree

1 Introduction

A degree document is issued after the completion of an academic program and is considered as the proof of its completion. Hence, companies or higher educational institutes can use the degree documents for confirming the educational history of the applicants. This pivotal role of degree documents attracts the scammers and motivates them to try securing jobs or admissions based on the fake or forged degree documents. Hence, the organizations can submit the verification request to the issuer of the received documents for confirming the originality and universities annually spent millions of dollars for handling the verification requests of degree documents [1]. This results in an important challenge of reducing cost and simplifying the verification process of the degree documents.

Blockchain is a popular concept which is being actively practised for solving many challenging problems of different domains. Massachusetts Institute of Technology (MIT) has introduced the blockcerts as an open standard for academic credentials on the blockchain and it can be freely used for verifying the academic credentials. However, we have identified the three limitations of blockcerts viz. 1) blockcerts changes the existing workflow of the degree issuance and is difficult to adopt for the degree awarding institutes, 2) it only operates over the degrees or certificates individually issued in the digital form and 3) it does not offer any solution for the degrees that have been already issued to the previously graduated students.

- *S. Rasool, A. Saleem, S. Rasool and A. Rasool has co-founded atDegree and works with atDegree and University of Gujrat, Gujrat, Pakistan. E-mail: Saqib@ieee.io and Saqib.Rasool@gmail.com*
- *M. Iqbal and T. Dagiuklas work with London South Bank University, UK*
- *S. Mumtaz works with Instituto de Telecomunicaes, Portugal*
- *Z. Qayyum is Vice Chancellor at Allama Iqbal Open University, Pakistan*

Manuscript received December 19, 2018; revised February 11, 2020.

This paper presents the docschain[1] as a solution to all three mentioned limitations of the blockcerts. The first limitation of the blockcerts is the change in existing workflow of degree issuance and it occurs due to its dependence on the graduating students. Blockcerts requires a student to create an account and share it with its degree-awarding institute and only then the degree can be issued through the blockcerts. This limits the universities to issue a single degree per transaction which is against their existing workflow of printing the degree documents in bulk. Docschain exposes the REST APIs through which universities can perform the data submission in bulk. Docschain also requires no student intervention during the degree issuance process and thus it can be easily incorporated within the existing workflow of degree issuance by the academic institutes.

The second limitation of blockcerts is to operate only over the digital data available only through the blockcerts. This confines the students to apply for a job or higher education only through the blockcerts credentials. On the other hand, docschain allows the students to follow the conventional way of applying by submitting the scanned or hard copies of degree documents as docschain uses the OCR (Optical Character Recognition) for extracting the data of degree documents. A degree awarding institute first defines an OCR template which is used by docschain to understand the semantics of data given at different sections of a document. We have patented a way of storing the OCR template along with the data of corresponding hard copy in the blockchain. The stored OCR template can correctly point to the different data containing sections of a document and the same approach is used in docschain for correctly extracting the accurate information from the degree documents.

1. www.DocsChain.org

The third limitation of the blockcerts is that the process of degree issuance needs to be initiated by the degree-awarding institutes by sending invitations to the students which are then accepted by the students that are willing to join the blockcerts. However, this is particularly difficult to reach out to the already graduated students for inviting them to participate in the degree issuance through the blockcerts. In contrast, docschain allows the bulk submission of data, for multiple degree documents, without involving the graduating or already graduated students. Hence, the participating institutes can very easily list the data of all the degree documents that have been awarded till date and it further simplifies the adaptation of blockchain-based degree verification solution by the academic institutions.

Blockcerts mainly focuses on the simplification of the degree handling process for the students while docschain simplifies it both for the students and the academic institutes. However, both solutions face the same challenge of ensuring the privacy of the data stored on the blockchain. According to the US federal law of FERPA (Family Educational Rights and Privacy Act) [2], it is not allowed to share the students' data without their permission. However, blockchain is inherently based on a distributed ledger that is shared among all the participants and to make a FERPA compliant solution, it must restrict the visibility of data to the corresponding students. PoE (Proof of Existence) is a popular solution used for handling the privacy of the data stored in the blockchain. Both blockcerts [3] and our proposed solution of docschains are using the PoE for ensuring the data integrity without sacrificing the privacy of the actual data of the stored degree documents.

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 [3], [4] that are using the blockchain-based PoE for various use cases. In the case of docschain, we are using the SHA-256 hashing algorithm [5]. In contrast to the CRUD operations of the database, blockchain only supports the operations of create and read and we have used the same one-way hashing algorithm for both of these operations. REST APIs are used for the create operation while the IoT/WoT cameras are used for the read operation in the docschain. The image captured through the cameras is first passed through the OpenCV [6] for the purpose of preprocessing and is then converted to the digital text using the tesseract [7], [8]. This digital text is then converted to a one-way hash using the SHA-256 algorithm and is then searched from the ledger of the docschain.

The contribution of this paper is threefold. Section three covers the first contribution of incorporating a blockchain-based solution within existing workflow of degree issuance. Section four covered the details of the second contribution of embedding OCR template within the blockchain and section five explains the third contribution of supporting the verification of degree documents that have been issued till date by the participating degree awarding institutes. Section six enlists the optimizations that have been made to improve the effectiveness of the degree verification through the docschain and the last section concludes this paper along with presenting some future research directions.

2 Related Work

Millions of fake academic documents and certificates have been reported around the globe [9] and it motivated the researchers to use blockchain for verification of degree documents. University of Nicosia was the first academic institute to issue blockchain-based verifiable certificates, in 2014, for its online course [10], [11]. However, blockcerts is the first bitcoin-based blockchain solution, for verification of academic certificates, that gained the traction of masses and therefore, we have presented it for the detailed comparison of the docschain. Researchers have also built many other blockchain-based degree verification solutions using different blockchain platforms like bitcoin [10], [11], [12], [13], ethereum [14], [15], [16], [17], [18], [19], hyperledger [20], [21], multichain [22], [23], tangle [24] etc.

Researchers have compared the two popular platforms, of bitcoin (through blockcerts by MIT) and ethereum (through TrueRec by SAP), to evaluate these for degree verification solutions [25]. Efforts have also been made for using multiple platforms like ethereum and hyperledger simultaneously for building a degree verification solution [26]. Custom blockchain implementation [27] has also been used for the same purpose and we have also followed the similar approach of using our own custom implementation of blockchain for docschain. Researchers have presented the comparison of the consortium blockchain based solution with public and private blockchain based degree verification solutions [28], [29] and thus we have implemented the docschain as a semi-private blockchain solution, which is an extension of the consortium blockchain.

To this point, all the discussed blockchain-based degree verification solutions are using the data of degree documents in digital form. However, few projects have been proposed for the verification of digitally signed PDFs [30] or the hardcopies of degree documents equipped with bar codes [31] or QR codes [32]. All of these approaches require the academic institutes to update the already issued degree documents. However, docschain operates both on digitally scanned images or physical photocopies of the degree documents, without requiring any change in these copies and we have achieved this through OpenCV (for preprocessing) [6] and tesseract (for OCR) [8]. Tesseract [8] has already been used for storing grades in database after extracting these from copies of report cards. We have used it with a template of a degree document that contains the details of localized text in the form of bounding boxes as it tremendously improves the accuracy of tesseract [7]. To the best of our knowledger, docschain is the first-ever solution to store the information of this template on the ledger of blockchain and we have also patented it.

A study has been conducted on 25 blockchain-based degree verification solutions which shows that 10 projects are facilitating both institutes and students as compare to 11 and 4 projects that only facilitates students and institutes respectively. Researcher have emphasized on improving the usability [33] and have also used the parameter of usability [32] for evaluation. Hence, we have also proposed the docschain to support both students and institutes and have tried to improve its usability through automatic verification of degree documents from an IoT/WoT camera.

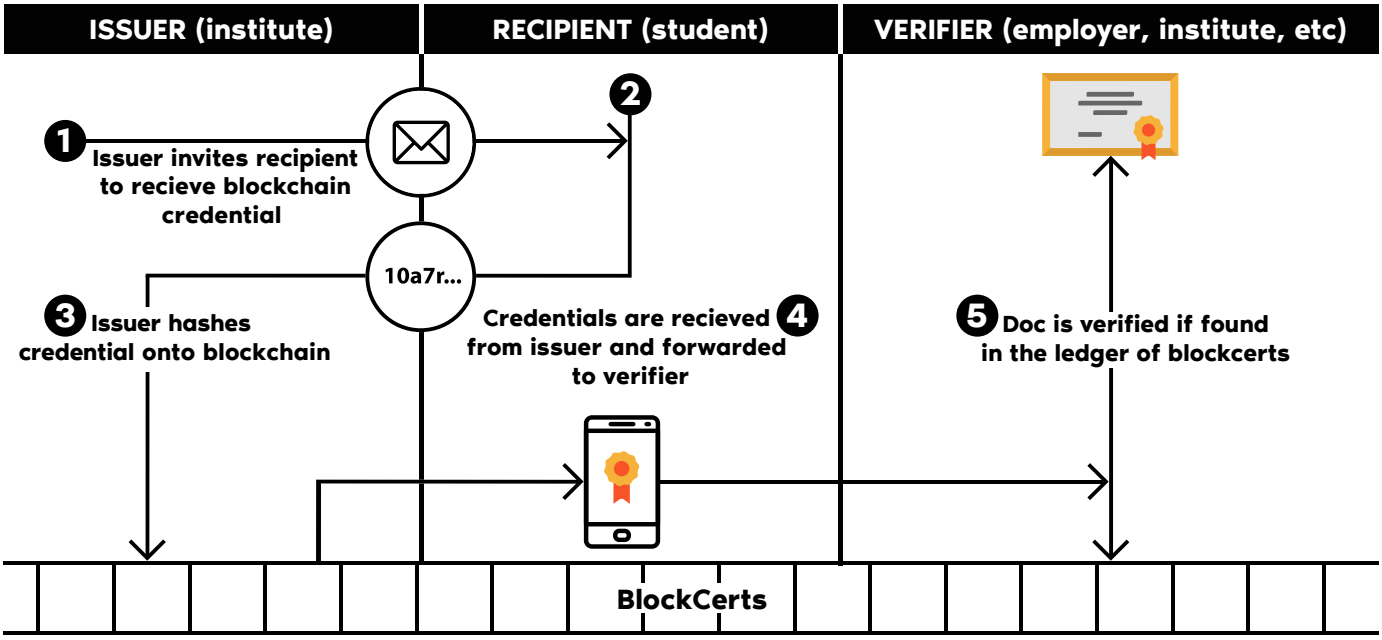

Fig. 1. Workflow of degree issuance and verification through the existing solution of blockcerts

3 Incorporation of Blockchain within the Social Workflow of Degree Processing

A social workflow originates when a group of people interacts with each other to perform a common task [34]. This section presents the first out of three contributions of the paper by presenting a blockchain-based degree verification solution that can be easily adopted within the existing social workflow of degree issuance and verification.

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 its mobile application. 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 the given credentials and this information is then stored on the ledger of blockchain.
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 mobile application of blockcerts.

3.2 Limitations of Blockcerts

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

- **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. The effect of this dependency on students becomes more severe 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 graduated 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 only place the data of 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 limitations of blockcerts for institutes by eliminating the need of involving students during the process of submitting the data of the 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. Docschain solves this problem by adding the support of degree verification through the hard copies of degree documents. Hence, the 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 submit its photocopy in hard form.
- **Employees** do not have the option of verifying the scanned copies or photocopies of documents through the blockcerts and are only bound to accept the digital credentials of the applicants generated by the mobile application of the blockerts. However, docschain eliminates this need by operating over the hard copies only.

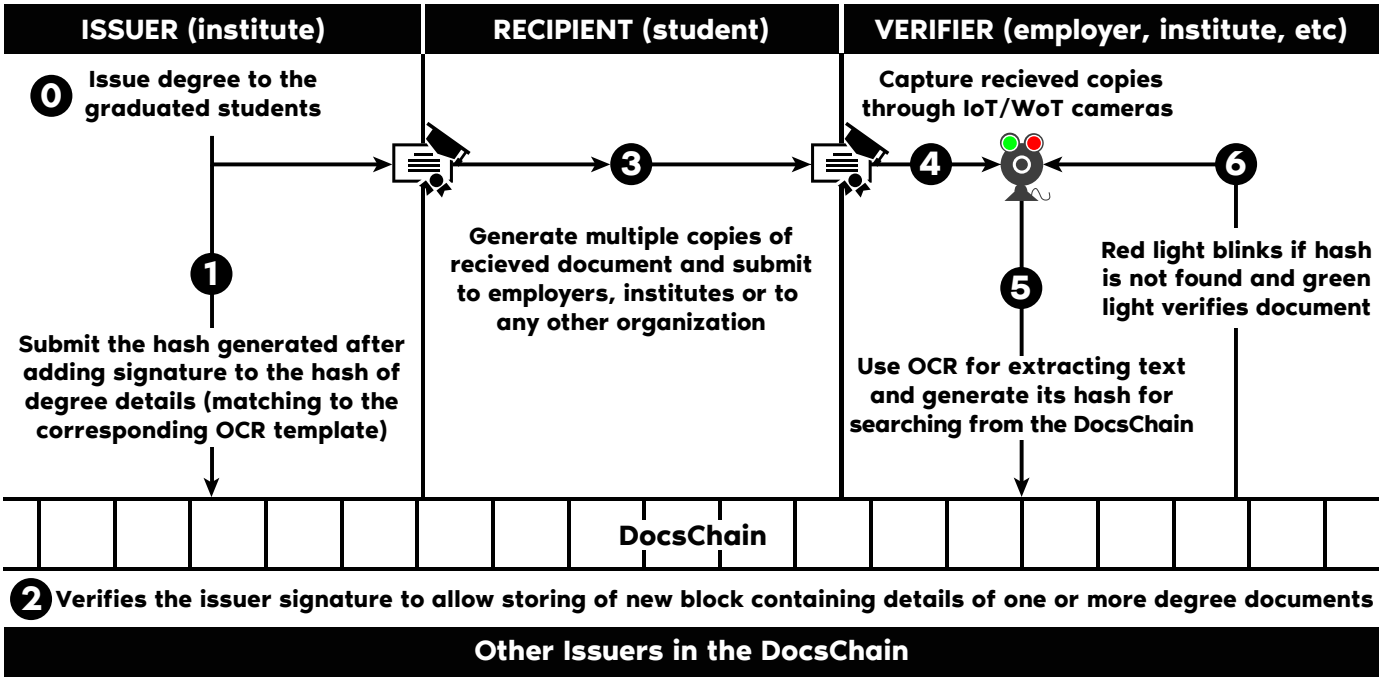

Fig. 2. Workflow of degree issuance and verification through the proposed solution of docschain

3.3 Workflow of Degree Issuance and Verification through Proposed Solution of docschain

Fig. 2 shows the workflow of degree processing through docschain and each of its steps are elaborated below:

0) Degree documents are printed for the students by the degree awarding institute and the zero number of this step shows its decoupling from docschain as it can be performed before or after storing data on the docschain.
1) The degree awarding institute can use the REST API of docschain (covered in sub-section of 5.2) for storing data of a single degree or multiple degrees in bulk at the docschain. The submitted data must be hashed using the private signature of the institutes and other requirements of the format of this data are given in sub-section of 4.2.
2) All other degree-awarding institutes identify the public key of the degree awarding institute from its name in the block given in Fig. 3 and uses this to validate if the block is submitted by the authorized degree awarding institute or not. In the ideal scenario, all the participants will accept or reject the request of adding new block to the docschain and a more advanced consensus algorithm needs to be proposed for tackling the edge cases.
3) Students can take multiple copies of the received degree and can submit scanned copies/photocopies to multiple organizations for the purpose of higher studies, job etc.
4) The verifier receives the hard copy of the degree document from the student or can take print of the scanned copy submitted in the digital form and places that copy in front of the IoT camera of the docschain. A button of IoT camera is currently used for capturing the image of the hard copy of the degree document. However, we are working on a solution that can automatically detect and capture the image of the degree document, directly from the video stream of the IoT camera, for improving the usability of the docschain.
5) The sub-section 5.3 explains the technical details of searching the captured image from the docschain for finding the legitimacy of that document.
6) Two lights of green and red are added on the IoT camera for showing the success or failure of degree verification respectively.

3.4 First Contribution of Docschain

The first contribution of docschain is its seamless integration within the existing workflow of degree issuance and verification. Many of the existing blockchain-based degree verification solutions [26], [35], [36] have been proposed as a complete solution for different steps of the degree processing. A recent study has compared the 25 existing blockchain-based degree verification solutions from the prospective of facilitating different stakeholders of the degree processing but none of these can be adopted within the existing workflow of degree processing.

The workflow of degree processing typically involves three types of stake holders; viz. 1) issuer, 2) recipient and 3) verifier. 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. Fig. 2 shows the first contribution of the docschain in step zero which depicts the independence of degree issuance from the docschain and it facilities the degree awarding institutes to follow the existing workflow of degree issuance, without facing any friction from the docschain. Students and verifiers can follow the conventional ways of submitting and receiving the documents respectively. However, the verifiers just need to replace the conventional way of manual verification of documents with an automatic document verification process through an IoT camera of the docschain.

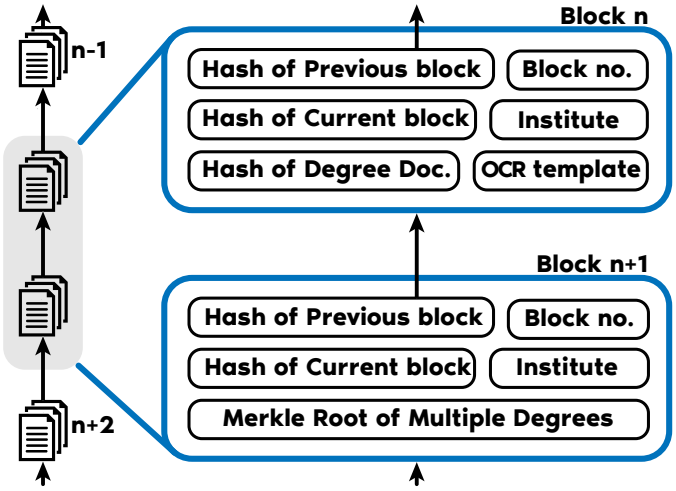

Fig. 3. Docschain with two blocks of different configurations

4 Blockchain of Docschain

This section presents the second contribution of the docschain along with the block structure (given in Fig. 3) of the ledger of docschain. It also explains the semi-private blockchain nature of the docschain and elaborates the read and write access for the ledger of the docschain.

4.1 Second Contribution of Docschain

The second contribution of the docschain is to introduce the concept of storing an OCR template, along with the hash generated from the content of the hard copy of the document, in the blockchain ledger. That OCR template contains the information similar to the ODL (OCR Description Language) [37] and is used for more accurately reaching to the ground truth [38]. A patent has been filed for storing images in the blockchain [39]. However, we are just storing the data of an image along with the OCR template that can be used for accurately extracting the data to get the exact same matching hash in the blockchain. We have also filed an application for patent to use this novel approach of storing OCR template in the ledger of the blockchain.

4.2 Two Types of Block Structures for the Distributed Ledger of the Docschain

Fig. 3 shows two different types of blocks for the distributed ledger of the docschain. The *Block n* is for a single degree document while *Block n+1* is for multiple degree documents. Blockcerts stores the data of a single degree document per transaction while the researchers have highlighted the requirement of bulk submission [29], [40] of data of multiple degree documents. Hence, we have also added support of both single degree document per transaction (Block n) and multiple degree documents per transaction (Block n+1).

Each transaction of submitting the data of degree documents is initiated through the REST API of the docschain and the received data is passed through a one-way hashing algorithm, after adding the signature (private key [41]) of a particular degree awarding institute. Hence, each block of the docschain can store the data of the degree documents belonging to only a single degree awarding institute.

Following is the description of each of the components of both blocks shown in Fig. 3:

- **Block number** is an auto-incremental number used for identifying different blocks of the blockchain ledger. It may also refer to the block height which defines the number of blocks between the targeted block and the genesis block which is the very first block of the distributed ledger of any blockchain.
- **Hash of the previous block** is used to link multiple blocks of docschain in an immutable manner. Each hash is generated using the one-way hashing algorithm of SHA-256. As its name states that it generates a string of 256 characters and we are using that string for representing the data of the whole degree document.
- **Hash of the current block** is generated in two different ways. In order to generate the hash for a block similar to the *Block n*, a signature of private key is added with five values, viz. 1) *hash of previous block*, 2) *block no*, 3) *insititute*, 4) *hash of degree document* and *OCR template*. In case of hash generation for the block similar to the *Block n+1*, both *hash of degree document* and the *OCR template* are replaced with the *merkle root of multiple degree documents*.
- **Institute** contains the title of the degree awarding institute and it must be unique for every institute. This unique title of the institute is used by the other participants of the docschain to find the public key of the institute which is trying to add a new block in ledger of the docschain. Each participant will be using that public key to verify if the submitted block is signed with the valid private key, acting as the signature of that degree awarding institute.
- **OCR template** is a custom representation of multiple bounding boxes on the template of a degree document of a degree awarding institute. Each bounding box refers to an exact location of the text at the degree document and the set of all bounding box is referred to as an OCR template. Each institute must define an OCR template while submitting the data of a degree document and this OCR template is used for accurately reaching the ground truth [38] during the degree verification process.
- **Hash of Degree Document** is collected by generating the one-way hash of data collected from all the bounding boxes of the corresponding OCR template. This hash is stored in the block of distributed ledger and the concept of replacing original data with its one-way hash is known as proof-of-existence (PoE) [4]. This generated hash not only ensures the privacy but also maintains the integrity of the data. PoE uses the one-way hashing algorithm for producing the Irreversible output for ensuring the privacy of the data and even a minor change in the data results in the change of the generated hash which helps in maintaining the integrity of the data. used for verifying the existence of a degree document by any third-party verifier using the concept of PoE.
- **Merkle root of multiple degrees** is required only when a block contains the data of multiple degree documents. An institute can select different same or different OCR templates for each of the nodes of a merkle tree. However, the information of an institute is not stored in the merkle tree and hence only one institute can store the documents in a single block of the docschain.

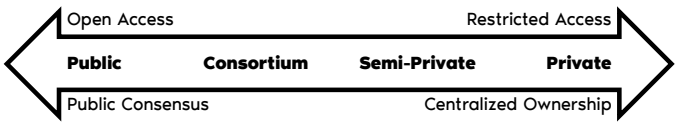

Fig. 4. Categorization of blockchain platforms with respect to the access and ownership

4.3 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 (often referred as miners). However, the read operation can be performed by both of the special devices and the normal users of the blockchain. Based on the access of creating new data and reading existing data, blockchains can be broadly categorized into the permissioned and permissionless blockchain platforms.

Permissionless [42] blockchain platforms are open for everyone to join and are known as public blockchain platforms. Permissioned [43] blockchain platforms need some sort of validation before allowing the access and these can be further divided into consortium [44] and private [45] blockchain platforms. Hence the division of permissioned and permissionless blockchain can be discussed in terms of three types of platforms [46], 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 [47], [48], [49]. 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 of the blockchain solution can ensure the transparency of the semi-private blockchain.

We have implemented the docschain as a semi-private blockchain. Fig. 4 compares the four discussed categories of the blockchain solutions against the level of access and ownership. It shows that a semi-private blockchain solution is required to restrict the read access and also the ownership to limited users. Semi-private blockchain of docschain uses PoE to restricts the read access to ensure the privacy of the stored data and authorize the write access to only the institutes that have submitted the correct OCR template against their degree documents.

5 Verification of Hard Copies of Degree Documents from the Digital Data

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.

5.1 Third Contribution of Docschain

Third contribution of the docschain is to receive the data of degree documents in the digital form and use it for the verification of the hard copies of degree documents. This contribution is very important for the degree awarding institutes and their alumni because it allows the institutes to shift the data of all the degree documents, that have been issued till date, to the docschain so that all of the already graduated students can also verify their already collected degree documents.

5.2 Storing Digital Data 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 the next section describes the data retrieval process which is used for the verification of hard copies of the degree documents.

- **Operation of adding a new institute** does not add any data in the ledger and is only completed after the approval from the admin of the semi-private blockchain of docschain. Admin generates a pair of public-private key and shares the private key with the newly added institute while distributes the public key with all the other degree awarding institutes in docschain. Admin will only approve the addition of a new degree awarding institute after confirming the submission of a valid OCR template by the particular institute.
- **Operation of adding a new OCR template** also does not add any information into the distributed ledger of the docschain and this operation is only accomplished after the verification, of the submitted 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 other participating degree awarding institutes of the docschain.
- **Operation of adding new degree documents** only adds the data to the distributed ledger of the docschain (mentioned in step 1 and 2 of the Fig. 2). Whenever a new block of data of degree document/documents is submitted, it is first verified from the signature of the degree awarding institute. If the signature is found valid, only then the newly submitted block is added in the ledger of the docschain. However, if the newly submitted block is not hashed with the valid signature of the degree awarding institute then the request for the addition of a new block, in the docschain ledger, is declined by the other degree awarding institutes of the docschain.

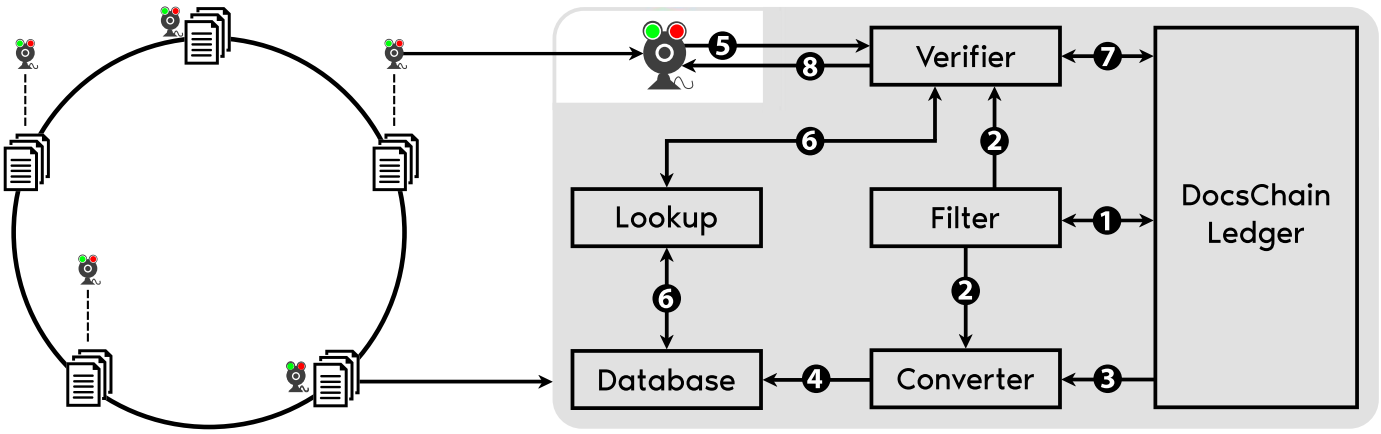

Fig. 5. Steps of degree verification through IoT and WoT Cameras

5.3 Retrieving Data for Verification of Hard Copies of Degree Documents

Fig. 5 elaborates different steps of retrieving data for docschain ledger for performing the degree verification. Following is the explanation of each of these steps:

1) Initially the *filter* module is invoked through two different REST endpoints. First REST endpoint operates without any input and returns all the institutes available in the docschain. The second REST endpoint takes a list of institutes (1 to N) as input and returns the OCR templates of all of the given institutes.
2) The *filter* module then forwards the collected information to both the *converter* and the *verifier* modules. Populating both of these modules through the same *filter* modules and with similar configurations ensures the synchronized working of both modules. This is important because the *filter* leads to efficiently finding the required data and the *verifier* module performs the verification of the same data from the ledger.
3) The *converter* module focuses on reducing the lookup time for *step 6*. It uses the list of institutes received from the *filter* module to pick the data from the *ledger* and efficiently stores it to the *database*, after performing the proper indexing, to optimize the lookup time for the data retrieved from the ledger of the docschain.
4) The *database* is focused on reducing the lookup time for *step 7*. The *converter* module not only stores the data of degree documents from the *ledger* but also mark the block number for each of the degree document. That block number then helps in *step 7* to quickly reach at the exact block for performing the verification through the ledger of the docschain.
5) As soon as the image of the degree document is captured by the *camera*, it forwards it to the *verifier* module which has already collected the list of available institutes along with their OCR templates in step 2. It is also equipped with a generic OCR which tries to find the title of the degree awarding institute from the captured document with the list of institutes returned from the filter module. If the generic OCR is failed in finding the institute then it skips both 6 and 7 steps and directly reaches to step 8, by displaying red light, to declare the non-verified degree document.
6) This step only executes if the *verifier* module can successfully find the degree awarding institute of the captured degree document. Before reaching to the lookup module, the OCR applies all of the available templates of the identified institute and selects the information of a single template which matches the maximum constraints of each of the bounding box stored in the template. 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 searches that hash from the *database*. Hash will not be found even if a single character is modified from the degree document and the *verifier* module will keep on trying with all the OCR templates. However, if the lookup is failed for all the templates then the verifier will skip step 7 by directly moving to step 8 and displays the red light for showing the non-verified degree document.
7) This step will only execute if the *lookup* module returns true and it will verify the same from the *ledger*.
8) It will trigger the green light, for showing the successful verification of the captured document, only if the *verifier* returns true after executing step 7. For all the remaining scenarios, the red light will be displayed for showing the non-verifiability of the captured degree document.

6 Optimizations for Efficiently Retrieving Data from the Ledger of the Docschain

A blockchain ledger stores data in the blocks that are connected in a linked list like structure [50] and therefore, it is a time-consuming process to search data from the blockchain ledger. Hence, we have implied different optimization techniques to reduce the data searching time from the docschain and this section covers these techniques in detail.

6.1 Using Database for Efficient Searching

It has been already proven that a database operates many times faster than a blockchain [51]. Hence, we have used the database as a caching layer for the distributed ledger of the docschain. Fig. 5 highlights the interactions between the database and the blockchain ledger. Initially, database cache the data of the ledger and once an entry is found in the database, docschain ledger acts as a single source of truth by providing the final confirmation of verification.

6.2 Frequency of Lookups from Distributed Ledger

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.

6.3 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.

6.4 WoT vs IoT Camera and the Ad-hoc Cloud

Fig. 5 contains three IoT cameras that are connected with ledger through dotted lines and two WoT cameras that are directly connected with the distributed ledger of the docschain. Both IoT and WoT can be characterised by the availability of computational resources. The WoT devices are resource-rich devices that can afford to perform the CPU intensive tasks while the IoT devices are resource-constrained devices that need special architectures and protocols for reducing the resource and power consumption [52]. Hence, we have used two different execution approaches for both IoT and WoT cameras.

Fig. 5 shows two arrows that are extending from the network of five distributed ledgers hosting nodes of the docschain and are used for highlighting the execution difference of both IoT and WoT cameras. Bottom arrow shows the processing of a WoT camera that hosts all the components with dark background and can perform all the required computation by itself. However, the same figure depicts the working of the IoT camera from the above arrow which shows that none of the computing components are hosted by the IoT camera. It just captures the degree image and forwards it to some other device which is hosting all the components, with dark background, for performing the computation on behalf of the IoT camera.

The other machine can be a dedicated system or some already running machine/machines that can also be used for offloading the computation of IoT camera. A group of devices that are used for performing the computation offloading, but are not dedicated for that purpose, can collectively form an ad-hoc cloud [53] and hence, the working of an ad-hoc cloud enabled IoT camera can be replaced with a WoT camera alone.

7 Conclusion and Future Work

In this paper, we have presented a prospective architecture of a blockchain-enabled degree verification solution that seamlessly fits within the existing workflow of degree issuance. It not only supports the verification of currently issued degrees but can also verifies the degree documents that have been issued till date and the degree awarding institutes can store the data of multiple degrees, ,in bulk, to the docschain. Hence, docschain support all the stakeholder of the degree processing including the degree awarding institutes, degree verifiers and both alumni and the currently graduating students of an institute.

Docschain has introduced a novel way of blockchain-based degree verification solution, therefore, it can be improved in many different ways. Following are the ten most significant enhancements that can further improve the effectiveness of the docschain:

1) We have evaluated the docschain for the verification of degree documents only. However, the same idea of blockchain-based OCR enabled solution can be applied for other use cases of hard copy verification.
2) Docschain is based on the idea of PoE [4]. A more advanced attribute level access control mechanism can also be adopted in future.
3) Docschain is based on REST but interfaces like GraphQL [52], RQL [54], gRPC etc. can also be offered in future.

4) Docschain only operates over grayscale images and support of coloured degree documents can also be added in future versions of the docschain.
5) The proposed solution of docschain only operates over the textual information as its OCR template only supports the text-based bounding boxes. It can be extended for facilitating the degree awarding institutes to store the logos on the docschain and image processing techniques can be used for improving the verification process by identifying an institute through its logo.
6) The performance of different components of docschain needs to be evaluated through a testbed as it will help in improving the overall performance of the system.
7) A mobile application can also be implemented for using the camera of the mobile phone for capturing the image of the degree document and utilizing the mobile resources for performing the degree verification.
8) A customized consensus algorithm needs to be implemented for reducing the network traffic and transaction time for storing the data on the docschain.
9) Currently docschain collectively uses the data from all the bounding boxes of the OCR template for finding the accuracy of degree documents. However, if data of all the bounding boxes are stored separately on the ledger then the verification module can independently validate the data of each bounding box. In this way, the exact bounding box with the fabricated data can be identified and can be reported for the fake degree documents.
10) Serverless computing has been evaluated for different use cases especially for the stateless applications [55]. Computation during the verification phase of docschain is a practical example of a stateless computation and the effectiveness of serverless cloud services can be evaluated in reducing the verification time of the docschain. In contrast to the current solution of horizontal resource sharing in the ad-hoc cloud, serverless will be based on the concept of vertical resource sharing [56].

ACKNOWLEDGMENT

The authors would like to thank Dr Muhammad Faheem Malik for his support. He is former Vice Chancellor at University of Gujrat and is currently serving as Pro-Vice Chancellor at the University of Gujrat, Gujrat, Pakistan.

REFERENCES

[1] N. K. Bajwa, "Modelling and simulation of blockchain based education system," Ph.D. dissertation, Concordia University, 2018.
[2] C. A. Ramirez, *FERPA clear and simple: The college professional's guide to compliance*. John Wiley & Sons, 2009.
[3] J. Hope, "Issue secure digital credentials using technology behind bitcoin," *The Successful Registrar*, vol. 17, no. 11, pp. 1–4, 2018.
[4] J. B. Bernabe, J. L. Canovas, J. L. Hernandez-Ramos, R. T. Moreno, and A. Skarmeta, "Privacy-preserving solutions for blockchain: review and challenges," *IEEE Access*, vol. 7, pp. 164 908–164 940, 2019.
[5] 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.
[6] A. Kaehler and G. Bradski, *Learning OpenCV 3: computer vision in C++ with the OpenCV library*. " O'Reilly Media, Inc.", 2016.
[7] S. Charjan, R. Mante, and P. Chatur, "Comparing tesseract results with and without character localization for smartphone application," *Technology*, vol. 2, no. 5, 2013.
[8] M. M. Bautista and B. E. V. Comendador, "Adoption of an open source optical character recognition (ocr) for database buildup of the students' scholastic records," *International Journal of Information and Electronics Engineering*, vol. 6, no. 3, p. 206, 2016.
[9] H. O. Ochieng, "A mobile based application for verification of legitimacy of degree certificates in kenya," Master's thesis, Strathmore University, 2016.
[10] 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.
[11] C. Brunner, F. Knirsch, and D. Engel, "Sproof: A platform for issuing and verifying documents in a public blockchain," in *Int. Conf. on Information Systems Security and Privacy (ICISSP)*, 2019, pp. 15–25.
[12] P. Schmidt, "Blockcertsan open infrastructure for academic credentials on the blockchain," *MLLearning (24/10/2016)*, 2016.
[13] A. S. d. P. Crespo and L. I. C. García, "Stampery blockchain timestamping architecture (bta)-version 6," *arXiv preprint arXiv:1711.04709*, 2017.
[14] T. Kanan, A. T. Obaidat, and M. Al-Lahham, "Smartcert blockchain imperative for educational certificates," in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*. IEEE, 2019, pp. 629–633.
[15] F. Brinkkemper, "Decentralized credential publication and verification: a method for issuing and verifying academic degrees with smart contracts," Master's thesis, University of Twente, 2018.
[16] E. F. G. Dias, "Ethereum smart contracts for educational certificates," Ph.D. dissertation, 2018.
[17] S. Kolvenbach, R. Ruland, W. Gräther, and W. Prinz, "Blockchain 4 education," in *Proceedings of 16th European Conference on Computer-Supported Cooperative Work-Panels, Posters and Demos*. European Society for Socially Embedded Technologies (EUSSET), 2018.
[18] Z. Zaccagni, A. Paul, and R. Dantu, "Micro-accreditation for matching employer e-hire needs," in *2019 IEEE International Conference on Blockchain (Blockchain)*. IEEE, 2019, pp. 347–352.
[19] B. Boeser. (2017, July) Meet truerec by sap: Trusted digital credentials powered by blockchain — sap news center. https://news.sap.com/2017/07/meet-truerec-by-sap\-trusted-digital-credentials-powered-by-blockchain/. (Accessed on 01/29/2020).
[20] A. Badr, L. Rafferty, Q. H. Mahmoud, K. Elgazzar, and P. C. Hung, "A permissioned blockchain-based system for verification of academic records," in *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE, 2019, pp. 1–5.
[21] J. A. Otuya, "A blockchain approach for detecting counterfeit academic certificates in kenya," Master's thesis, Strathmore University, 2019.
[22] G.-A. Dima, A.-G. Jitariu, C. Pisa, and G. Bianchi, "Scholarium: Supporting identity claims through a permissioned blockchain," in *2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI)*. IEEE, 2018, pp. 1–6.
[23] R. Arenas and P. Fernandez, "Credenceledger: a permissioned blockchain for verifiable academic credentials," in *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. IEEE, 2018, pp. 1–6.
[24] A. Wahab, M. Barlas, and W. Mahmood, "Zenith certifier: A framework to authenticate academic verifications using tangle," *Journal of Software and Systems Development*, vol. 2018, no. 370695, p. 13, 2018.
[25] S. Kazakzeh, E. Ayoubi, B. K. Muslmani, M. Qasaimeh, and M. Al-Fayoumi, "Framework for blockchain deployment: The case of educational systems," in *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*. IEEE, 2019, pp. 1–9.
[26] K. Batzavalis, R. Bala, A. Norta, and O. Norta Partners, "A platform for leveraging blockchain technology for the storage, issuance and authentication of academic credentials."
[27] R. P. Uhlig, R. Yonts, B. W. Cashman, R. S. Clark, B. Nieman, R. Landa, L. B. M. Elizalde, C. V. G. Cordova, K. Wright, P. E. Slaboch *et al.*, "Blockscripts–a blockchain system for university tran-scripts."
[28] K. Nikolskaia, D. Snegireva, and A. Minbaleev, "Development of the application for diploma authenticity using the blockchain

technology," in *2019 International Conference" Quality Management, Transport and Information Security, Information Technologies"(IT&QM&IS)*. IEEE, 2019, pp. 558–563.
[29] E. E. Bessa and J. S. Martins, "A blockchain-based educational record repository," *arXiv preprint arXiv:1904.00315*, 2019.
[30] E. Martiri and G. Muca, "Dms-xt: A blockchain-based document management system for secure and intelligent archival." in *RTA-CSIT*, 2018, pp. 70–74.
[31] S. Mthethwa, N. Dlamini, and G. Barbour, "Proposing a blockchain-based solution to verify the integrity of hardcopy documents," in *2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC)*. IEEE, 2018, pp. 1–5.
[32] A. Tariq, H. B. Haq, and S. T. Ali, "Cerberus: A blockchain-based accreditation and degree verification system," *arXiv preprint arXiv:1912.06812*, 2019.
[33] V. Garcia-Font, "Blockchain: Opportunities and challenges in the educational context," in *Engineering Data-Driven Adaptive Trust-based e-Assessment Systems*. Springer, 2020, pp. 133–157.
[34] S. Görg and R. Bergmann, "Social workflowsvision and potential study," *Information Systems*, vol. 50, pp. 1–19, 2015.
[35] J. D. Judd, "Cryptocollege: how blockchain can reimagine higher education," *International Journal on Innovations in Online Education*, vol. 2, no. 2, 2018.
[36] D.-H. Nguyen, D.-N. Nguyen-Duc, N. Huynh-Tuong, and H.-A. Pham, "Cvss: A blockchainized certificate verifying support system," in *Proceedings of the Ninth International Symposium on Information and Communication Technology*, 2018, pp. 436–442.
[37] K. Luo, J. Lu, K. Q. Zhu, W. Gao, J. Wei, and M. Zhang, "Layout-aware information extraction from semi-structured medical images," *Computers in biology and medicine*, vol. 107, pp. 235–247, 2019.
[38] B. Santra and D. P. Mukherjee, "A comprehensive survey on computer vision based approaches for automatic identification of products in retail store," *Image and Vision Computing*, vol. 86, pp. 45–63, 2019.
[39] K. Bakalis, "Systems and methods for using codes and images within a blockchain," Oct. 3 2019, uS Patent App. 16/156,570.
[40] J. Gresch, B. Rodrigues, E. Scheid, S. S. Kanhere, and B. Stiller, "The proposal of a blockchain-based architecture for transparent certificate handling," in *International Conference on Business Information Systems*. Springer, 2018, pp. 185–196.
[41] M. Pilkington, "11 blockchain technology: principles and applications," *Research handbook on digital transformations*, p. 225, 2016.
[42] Z. Jiang, B. Krishnamachari, S. Zhou, and Z. Niu, "Senate: A permissionless byzantine consensus protocol in wireless networks," *arXiv preprint arXiv:1803.08694*, 2018.
[43] 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.
[44] 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.
[45] 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.
[46] 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.
[47] "Startup management understanding semi-private blockchain applications," http://startupmanagement.org/2016/11/06/understanding-semi-private-blockchain-applications/, (Accessed on 04/21/2018).
[48] D. Guegan, "Public blockchain versus private blockhain," 2017.
[49] E. C. Ferrer, T. Hardjono *et al.*, "Robochain: A secure data-sharing framework for human-robot interaction," *arXiv preprint arXiv:1802.04480*, 2018.
[50] S. Gupta and M. Sadoghi, "Blockchain transaction processing." 2019.
[51] S. Chen, J. Zhang, R. Shi, J. Yan, and Q. Ke, "A comparative testing on performance of blockchain and relational database: Foundation for applying smart technology into current business systems," in *International Conference on Distributed, Ambient, and Pervasive Interactions*. Springer, 2018, pp. 21–34.
[52] 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.
[53] 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.
[54] S. Rasool, A. Saleem, and A. N. Mian, "Poster: Rql: rest query language for converting firebase to a mobile cloud computing platform," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 2017, pp. 567–569.
[55] T. Asghar, S. Rasool, M. Iqbal, Z. ul Qayyum, A. N. Mian, and G. Ubakanma, "Feasibility of serverless cloud services for disaster management information systems," in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 2018, pp. 1054–1057.
[56] S. Rasool, M. Iqbal, T. Dagiuklas, Z. ul Qayyum, and A. N. Mian, "Towards reliable computation offloading in mobile ad-hoc clouds using blockchain," in *International Conference on Broadband Communications, Networks and Systems*. Springer, 2018, pp. 180–188.

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/fog services.

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 Tasos received his Engineering Degree from the University of Patras, Greece, in 1989. He completed an MSc at the University of Manchester in 1991 and a PhD at the University of Essex-UK in 1995, all in Electrical Engineering. His research interest includes 5G Networking technologies (SDN, NFV, MEC), Cloud Computing Technologies and Machine Learning in 5G Networks.

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 and energy-efficient framework for 4G.

Dr. Zia ul Qayyum has completed his doctoral degree in Artificial Intelligence from university of Leeds, UK. Research interests in the areas of Knowledge Representation, Data Mining, Recommendation Systems, and Semantic Information Retrieval. Currently working as Vice Chancellor, Allama Iqbal Open University, Islamabad, Pakistan