A Secure and Distributed Construction Document Management System Using Blockchain

Chapter · January 2021

DOI: 10.1007/978-3-030-51295-8_59

3 authors:

Moumita Das
The Hong Kong University of Science and Technology
26 PUBLICATIONS **649** CITATIONS
SEE PROFILE

Xingyu Tao
The Hong Kong University of Science and Technology
15 PUBLICATIONS **224** CITATIONS
SEE PROFILE

Jack C. P. Cheng
The Hong Kong University of Science and Technology
251 PUBLICATIONS **8,719** CITATIONS
SEE PROFILE

A Secure and Distributed Construction Document Management System Using Blockchain

Moumita Das, Xingyu Tao, and Jack C. P. Cheng(✉)

The Hong Kong University of Science and Technology,
Clear Water Bay, Hong Kong
cejcheng@ust.hk

Abstract. Construction Document Management Systems (CDMS) track, manage, and store a variety of documents such as 3D model files, schedules, specifications, and images that are large in size and have different ownerships. The primary objectives of CDMSs are to facilitate document approval workflows, document version management, and security properties such as integrity of data, audit-trail, and approvals. A popular approach for CDMS is to consolidate all the project documents into an on premise or cloud-based centralized location that is owned and managed by a project participant or a trusted third party. Such centralized approaches for electronic documents pose risks such as single points of failures causing loss or corruption of documents and deliberate blocking of access during disputes between the owner of CDMS and other project participants. This situation is particularly challenging in construction projects where project participants cannot fully trust each other due to its inherently fragmented project organizational structure. Therefore, in this paper, a distributed construction document management system using Blockchain and distributed content-addressable storage technologies is presented. Blockchain is a peer-to-peer technology that facilitates distributed computation and irreversible data recording through smart contract and blockchain ledger technologies respectively. Immutability of records and computational logic is facilitated through the unique cryptographic data structure of blockchain ledgers and probabilistic consensus algorithms. In this paper, smart contracts are deployed to facilitate document approval workflows to support processes such as design review and request for information in construction projects. A blockchain ledger data model for tracking workflows and document version management is proposed. Public-key cryptography is deployed to facilitate data confidentiality and integrity in endorsements. A cryptographic indexing structure to support blockchain ledger in document versioning and to validate the authenticity of document search results for CDMS is proposed. The proposed framework also deploys peer-to-peer content-addressable storage for preventing single points of failure and data integrity of documents through data partitioning, data replication, and cryptography. The proposed CDMS is a distributed yet unified platform for managing construction documents. A demonstration of the proposed CDMS is presented with a case-based scenario on request for information (RFI) management.

E. Toledo Santos and S. Scheer (Eds.): ICCCBE 2020, LNCE 98, pp. 850–862, 2021.
https://doi.org/10.1007/978-3-030-51295-8_59

Keywords: Blockchain · Construction Document Management System · Document workflow · Versioning · IPFS · Security

1 Introduction

Construction document management systems (CDMS) are systems that coordinate and control tracking, storage, retrieval, processing, and distribution of electronic documents in construction projects [1, 2]. The primary objectives of construction document management systems (CDMS) are to facilitate document approval workflows and support versioning and searching of documents [3, 4]. Construction documents contain sensitive information such as financial information, patented drawings, and participant's personal information. Therefore, security is also one of the primary requirements of construction document management systems [1, 5, 6]. Existing document management systems include frameworks that manage documents and workflows through a centralized approach [2]. Caldas and Soibelman [7] presented an SVM-based document management system that classifies construction documents according to standard construction information classification systems such as CSI MasterFormat [8]. Park et al. [9] deployed ontologies representing domain-specific knowledge of construction documents to improve document searchability, by inducing consistency in the vocabulary of search keywords among multi-disciplinary participants of construction projects. Commercial software such as Aconex [10] and Autodesk 360 [11] deploy a centralized cloud-based platform for the management of construction documents and workflows (shown in Fig. 1). However, cloud platforms are vulnerable to security risks such as data loss, denial of data access, and partial control over sensitive data which is well-documented in the existing literature [12, 13]. Moreover, due to the inherent contractual nature of construction projects, project participants do not fully trust each other and therefore are not confident about entrusting a central entity with the right to store and manage documents. Therefore, existing centralized systems for document management that require consolidation of project documents and workflows on a physical or cloud-based centralized platform that is owned by a central project participant or a trusted third party is inapt.

Therefore, this paper proposes a decentralized but unified construction document management system using blockchain and distributed content-addressable storage. Blockchain [14] is a peer-to-peer technology that facilitates recording and computation of data in a distributed and trustless manner via distributed ledger technology (DLT) and smart contracts respectively. In the existing literature, smart contracts and blockchain ledger in combination with other decentralized technologies such as IPFS storage systems and IoT networks have been explored for checking the authenticity of videos [15], securing AI applications [16], automatic monetization and security of IoT networks [17, 18]. For construction projects, blockchain has been deployed for several applications such as measuring construction productivity [19], contract management [20], pre-cast supply chains [21]. In this paper, smart contracts technology of blockchain that facilitates modeling of executable code for automating processes in a distributed manner is deployed to facilitate document approval workflows in construction projects. For developing the smart contract framework, standard components

and relations in document approval workflows in construction projects are identified from existing literature. Based on the finding, smart contracts capturing the logic (associating roles to their corresponding tasks) of various approval processes such as design review, are developed. As workflows may be unique to organizations and projects, smart contract logic to orchestrate custom workflows by re-using standard sub-processes in document approval workflows is proposed. A blockchain ledger data model for tracking workflows and document versioning is proposed. A cryptographic indexing structure to facilitate searchability and authenticity (in terms of completeness and consistency) of query results is proposed. The proposed indexing structure maintains the references to older versions of a document and requires no reconstruction when newer versions are added. Public-key cryptography is used to facilitate data confidentiality and integrity of endorsements. A distributed peer-to-peer content-addressable storage such as Interplanetary File System (IPFS) is deployed to facilitate data integrity and prevent single points of failure through cryptographic hashing and data replication. The proposed framework takes a distributed approach for document management to ensure that workflows and documents are not controlled by a single entity while at the same time deploys measures to counter security risks to data confidentiality, integrity, and authorization due to decentralization. The proposed framework is supported by a case-based scenario on request for information management that demonstrates its key features and advantages such as automatic workflow execution and tracking and security properties such as data authenticity, data confidentiality, and user integrity.

2 The Proposed Framework

In this section, a framework for decentralized construction document management using blockchain and distributed content-addressable storage is proposed (as shown in Fig. 1). As shown in Fig. 1, the proposed framework takes a distributed approach towards workflow deployment, document versioning, audit-trail recording, and document storage using smart contracts, distributed ledger technology, cryptographic data structures, and content-addressable storage as discussed in Sects. 2.1–2.5.

2.1 The General Structure of Document Approval Workflows

Figure 2(a) shows the general components of workflows for document management in construction projects [22–24] – (1) processes such as design review, request for information, and change order management, (2) roles such as architect, main-contractor, and owner, and (3) data such as instructions to initiate or documents to support processes. Processes may be further sub-divided into sub-processes like ‘creating a new document’, ‘providing endorsement’, and ‘feedback’ as shown in Fig. 2(a). Figure 2(b) shows the structure of a typical document management workflow which consists of a series of processes and sub-processes facilitated by roles through supporting data.

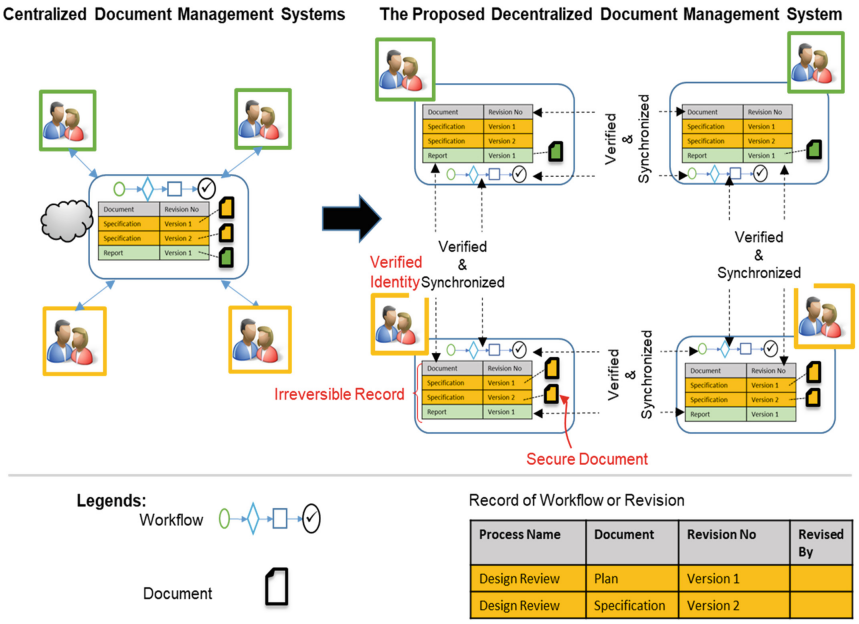


Fig. 1. The proposed framework for construction document management

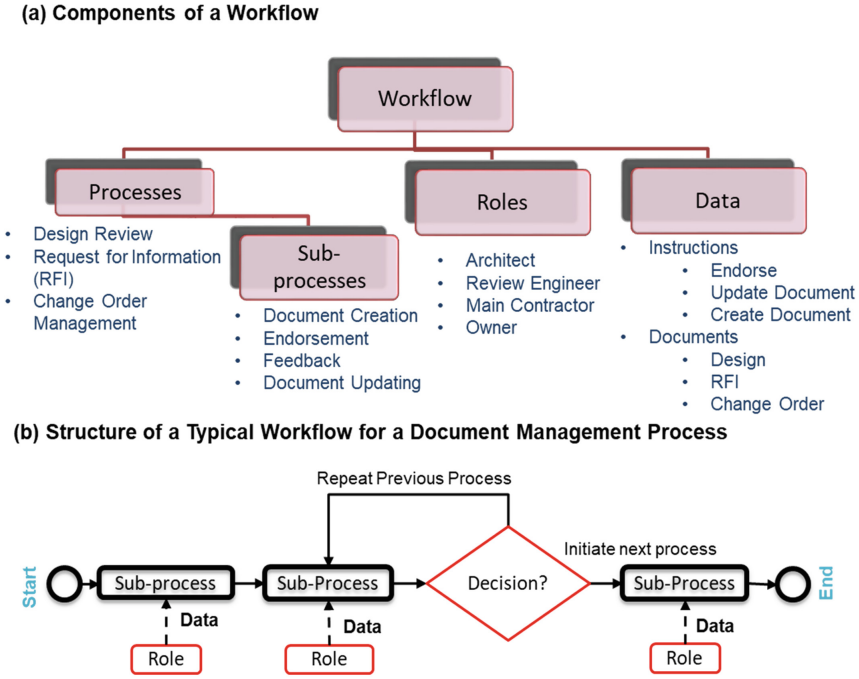


Fig. 2. The general structure of workflows for document management in construction projects

2.2 The Blockchain Ledger Data Model for CDMS

Figure 3 shows the data model of the blockchain ledger required for CDMS to record document approval workflows (discussed in Sect. 2.3) and versioning. The data model consists of metadata such as transaction ID, transaction type, project phase, document ID, the status of a document, time, and identity of roles in document approval workflows as shown in Fig. 3.

Attributes of the data model		Description	Example
Transaction ID		ID of a Blockchain Transaction	f4184fc596403b9d638783cf57adfe4c75c60 5f6356fbc91338530e9831e9e16
From Account		User or smart contract that sends a request to invoke a smart contract function	Process orchestrator Smart Contract/ End-User
To Account		A smart contract that receives a request from an end-user or another smart contract	Process orchestrator Smart Contract/ Sub-process Executor Smart Contract
Transaction Type		Type of sub-process	Add Document , Update Document, Endorse Document, Send Feedback
Payload	Project Phase	The phase of the construction project in which a process workflow is executed	Design/ Construction
	Document ID	Unique identifier of a document	17-111-12
	Status	The current status of a workflow process	RFI submitted, RFI endorsed
Time		Time at which a workflow process is executed	20191219 1220
Sender's Signature		Digital signature of the end-user who has invoked a sub-process smart contract	Role's Signature

Fig. 3. Data model of ledger in the proposed blockchain framework for CDMS

As shown in Fig. 3, 'Transaction ID' is a unique identifier for a smart contract execution that results in updating of the corresponding blockchain ledger and is generated automatically by the blockchain platform. 'Transaction type' in this data model refers to sub-process such as document creating, endorsement, or feedback corresponding to a phase of a document approval workflow. 'Project phase' is the phase of the construction projects in which document approval workflows are executed such as the design phase or the construction phase. 'Document ID' in the proposed data model refers to a unique document identifier composed of document type, document number, and document version number. For example, 17-111-12 represents a document ID where 17, 111, and 12 are document type, document number, and document version number respectively. 'Document type' refers to the type of a document such as 'notice to proceed', 'schedule', 'request for information', and 'field report' as per the Omniclass classification for construction information [25]. 'Document number' is an integer unique for the type of document. Unlike 'Document ID', the 'Document number' does not vary for different versions of the same document. Similarly, 'Document version number' is an integer unique for document number. Along with this, the proposed data model also records the status of workflows, time at which a workflow is executed, and the identity of the workflow executor as shown by 'Status', 'Time', and 'Sender's Signature', as shown in Fig. 3.

2.3 A Smart Contract Framework for Distributed Workflows for Document Management

Figure 4 shows the smart contract framework for facilitating document approval workflows for construction projects. The framework is designed by considering custom workflows to suit the needs of different organizations and project types. Therefore, two types of smart contract logics namely – (1) sub-process executor and (2) process orchestrator smart contracts are developed to facilitate reuse of sub-process for orchestrating different workflows.

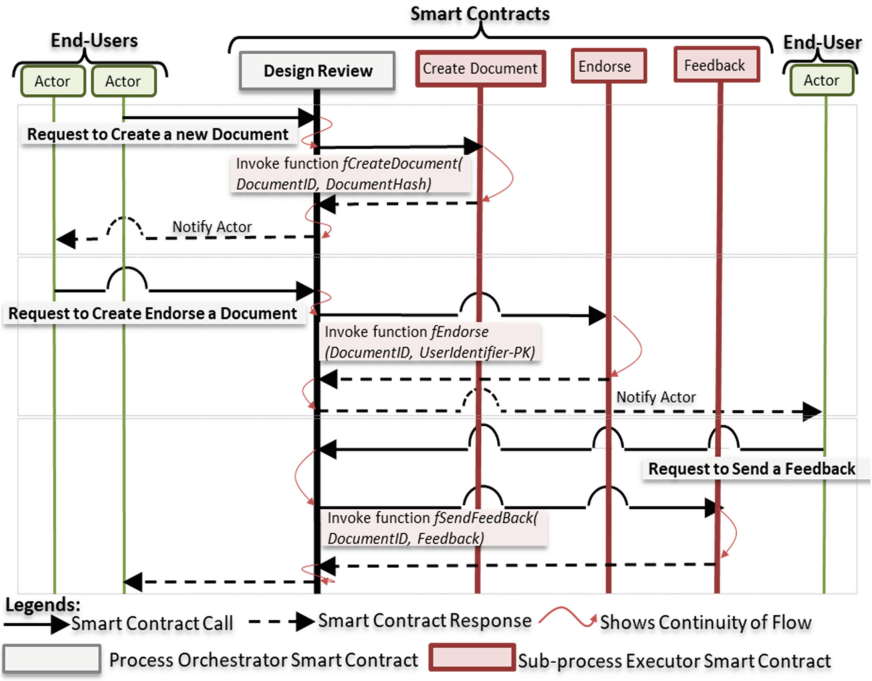


Fig. 4. Interactions between actors and smart contract for document approval workflows

2.3.1 Sub-process Executor Smart Contracts

As shown in Fig. 2, sub-processes such as document creation, document updating, document endorsement, and feedback are individual tasks that make a document approval workflow process. In this section, the logic to deploy sub-processes through smart contract functions is presented. Solidity which is the smart contract language of Ethereum [14] blockchain platform is used.

As shown in Fig. 4, *fCreateDocument* is a function in the '*Create Document'* smart contract that records the creation of a new document. The inputs to *fCreateDocument* function are Document ID (as discussed in Sect. 2.2) and a Document Hash (which is a unique identifier of the content of a document and will be discussed in Sect. 2.4). Upon execution, this function creates a record in the blockchain ledger (data model shown in

Sect. 2.2) and sends a notification to the project participant (as shown by 'actor' in Fig. 4) who is responsible for executing the subsequent sub-process of the document approval workflow. Sub-process executor smart contracts are created at the beginning of a construction project and deployed to the distributed blockchain network so that they may be verified for authenticity by all project stakeholders. Each sub-process is deployed as a separate smart contract to facilitate reuse and variation in orchestrations as required by different document approval workflows (discussed in Sect. 2.3.2). Furthermore, this approach imparts scalability to the proposed framework by permitting deployment of new smart contracts for special sub-tasks if required at a later stage or a project with special requirements.

2.3.2 Process Orchestrator Smart Contract

The objective of a process orchestrator smart contract is to orchestrate the calls to different sub-process executor smart contract functions as per the requirement of a document approval workflow. A process orchestrator smart contract facilitates orchestration of a document approval workflows via smart contract functionality to facilitate interactions between two smart contacts [26] by using ABIs (Application Binary Interface) of sub-process executor smart contracts. For example, the document approval workflow for the design review process consists of sub-processes such as uploading a new plan by the project architect, followed by endorsement from a review engineer, and feedback from the project owner. Therefore, as shown in Fig. 4, the logic of the process orchestrator smart contract for the design review process consists of sequential calls to functions *fCreateDocument* of '*Create Document*' smart contract, *fEndorse* of '*Endorse*' smart contract and *fSendFeedback* of '*Feedback*' smart contract facilitated by the architect, review engineer, and the project owner respectively. Figure 4 shows the logic of the process orchestrator smart contract for design review. Similarly, smart contracts for different document approval workflows such as a request for information and change order management may be deployed.

2.4 A Blockchain-Based Indexing Data Structure for Construction Document Management

In this paper, an indexing structure (as shown in Fig. 5) based on Radix tree [27] and Merkle tree [28] cryptographic data structures is presented for facilitating document querying according to document version and other attributes and enabling the security of consistency and completeness in the query results. A Radix tree is an ordered data structure for storing, indexing, and retrieving in-memory data that goes through many updates [27]. In a Radix tree, data is stored in a key-value format, where a "key" is a path (arranged in lexicographical order) in the tree data structure that leads to a "value" at a leaf node. As shown in Fig. 5, in the proposed indexing data structure, the 'Document ID' composed of "document type – document number – version number" (such as "14-13-1") is deployed as a branch (path) that leads to a leaf node. As shown in Fig. 5 the value is a document hash which is a unique identifier based on the content of a document generated by content-addressable storages like IPFS (Interplanetary File System) [29]. IPFS is a peer-to-peer distributed file system that stores documents in a key-value format such that the key is a cryptographic hash generated based on the

content of the document and the value is the document itself. Cryptographic hashes [30, 31] are deterministic encryption functions that convert a document or a text into a unique 256-bit hexadecimal number. This means that the hash of a document varies with very small changes in the document content and hence may be used as proof of authenticity.

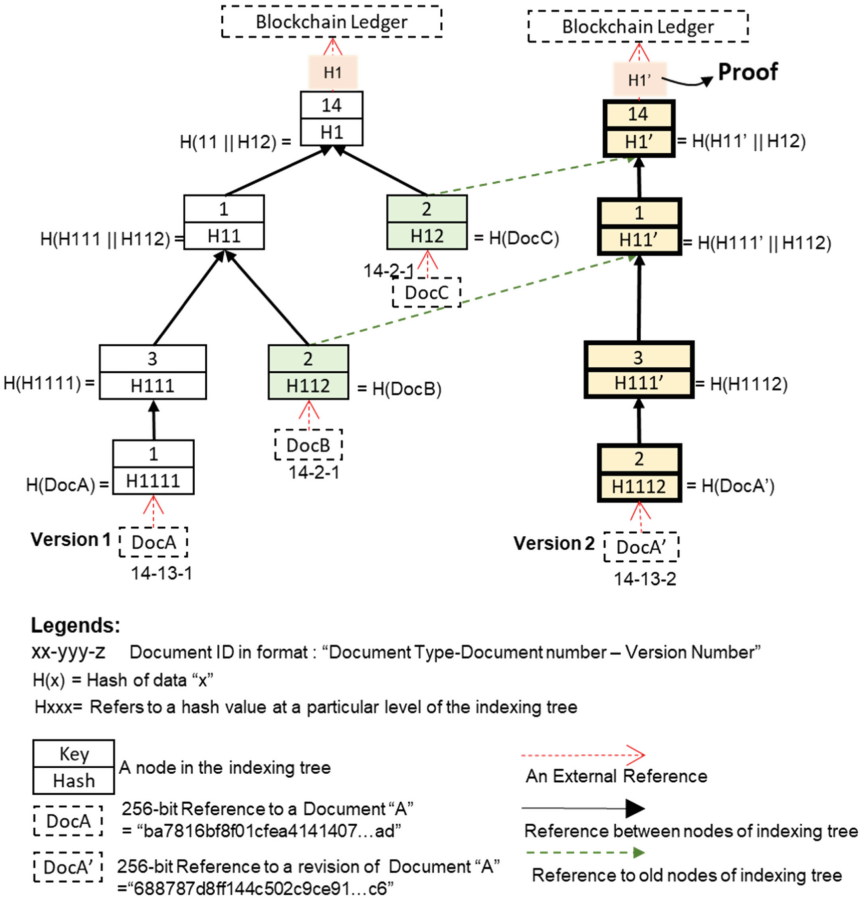


Fig. 5. A blockchain-based indexing data structure for document versioning

The second type of tree data structure used in the proposed framework is Merkle tree [28] to facilitate the integrity of datasets in distributed environments. Merkle trees are tree-like data structures in which the leaf nodes are deterministic cryptographic encryptions (Hashes) of the data stored in them and the non-leaf nodes are hashes of their child nodes. As shown in Fig. 5 the leaf node *DocA* of the proposed indexing data structure represents a 256-bit hexadecimal document hash. Its parent node contains the hash of this unique identifier as shown by *H1111*. Similarly, other parent nodes contain hashes generated from their corresponding child nodes. As shown in Fig. 5, when a

new version of *DocA* is added, it is included in the indexing tree through 4 new nodes and references to old nodes which prevents the need to reconstruct the whole tree as necessary in other indexing data structures such as B+ trees [27]. The authenticity of a document version may be verified using its corresponding root hash (proof) as shown in Fig. 5. For example, $H1'$ is the proof of completeness and consistency of the new document version, $DocA'$. It is difficult to modify $H1'$ as it depends on unique document hash and references to previous document hashes and is stored on blockchain ledger which is irreversible in nature. In Sect. 3, a case-based scenario is presented to demonstrate this property.

2.5 Public Key Cryptography for Data Confidentiality and Endorsement

In this paper, data confidentiality and endorsement is facilitated via public-key cryptography [32] for the proposed framework. To deploy public-key cryptography, every project participant should have a public-private cryptographic key pair (as shown as "$\boldsymbol{PK^i - SK^i}$" in Fig. 6 each. The public key, "PK^i" is the public identity of a user "i" and may be shared publicly while SK^i is the private key that should be kept confidential. Both public and private keys may be used to encrypt data to facilitate different types of security. Figure 6 shows two approaches where public and private keys are used in opposite manners to deploy digital endorsement and data confidentiality. For endorsement, a user A (as shown in Fig. 6) may encrypt data with his private key. Anyone may authenticate the validity of this endorsement if he can decrypt the data using the corresponding public key of user A. Similarly, data confidentiality is facilitated by encrypting data with a user's public key. Therefore, only those users who have the corresponding private key are able to access the encrypted data hence facilitating data confidentiality.

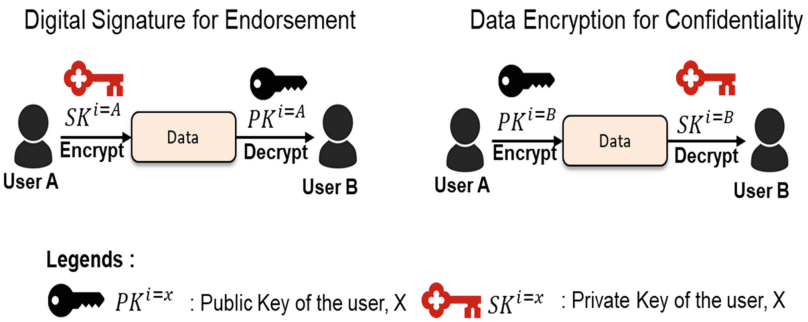


Fig. 6. Data confidentiality and endorsement security using public-key cryptography

3 Validation

In this section, the key properties of the proposed framework which are workflow automation, document searching, consistency and completeness (data integrity), and integrity of endorsement are validated using a case-based scenario of request for

information. Figure 7 shows a standard document approval workflow for a request for information [11, 33]. This workflow primarily consists of four sub-processes that are – (1) creation of an RFI document, (2) endorsement of the RFI document, (3) addition of comments thereby updating the RFI document, and (4) sending a notification to the project owner.

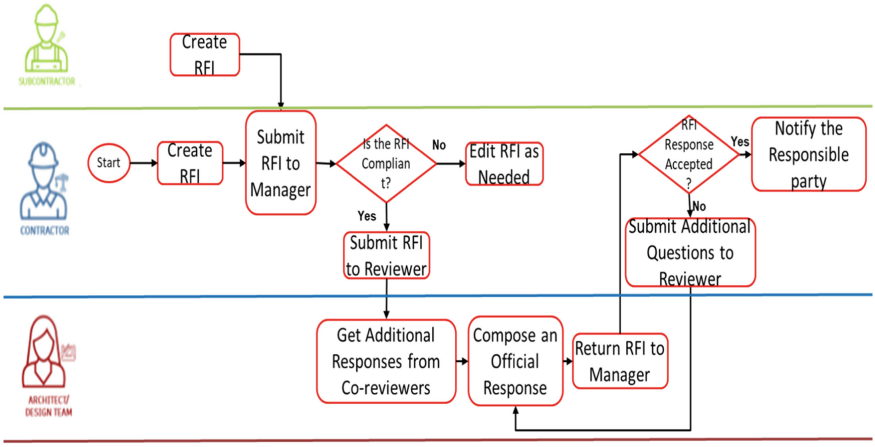
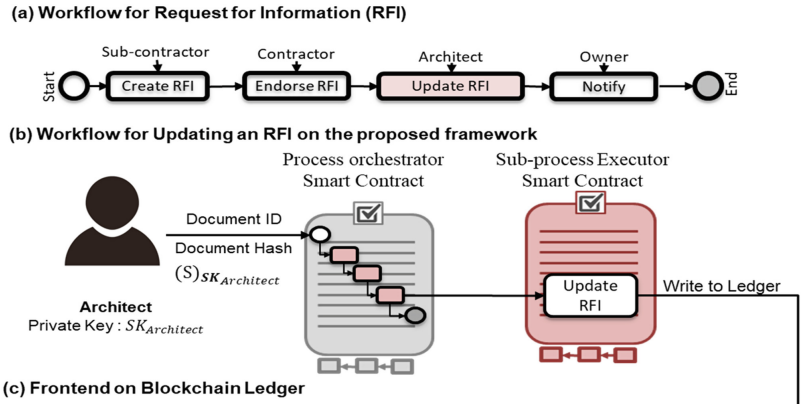


Fig. 7. A standard 'request for information' (RFI) workflow

Figure 8(a) shows the pictorial depiction of the RFI workflow that is deployed using the proposed blockchain-based framework. Figure 8(c) shows how the RFI workflow is tracked on the blockchain ledger by recording information such as document ID, name of the sub-process, and a digital endorsement of the authorized end-user. Endorsements are provided through digital signatures using the private keys of project participants. This facilitates integrity in endorsements as a digital endorsement once recording on an irreversible blockchain ledger cannot be refuted at a later stage.

Figure 9 shows the simplified blockchain ledger and the indexing data structure deployed by the proposed framework to facilitate the searching and integrity of search results. A particular version of a document may be searched by first parsing the blockchain ledger indexed with a local indexing structure such as Skip List [34] to find the document ID followed by traversing the indexing data structure shown in Fig. 9. In order to validate the authenticity of a document say, $DocA'$ (as shown in Fig. 9), the nodes (Shown in green in Fig. 9) of the indexing data structure should be requested from the network. The receiver may be assured of the correctness of a document if he can recreate the root node corresponding to a document. As shown in Fig. 5, the new state root, $H1'$ can be generated by an end-user if he has $DocA'$ and the elements (as shown in green) in Fig. 5. To verify the integrity, he can compare it with the state root stored in the blockchain ledger. If he has the right version of $DocA'$, he should be successful.



(c) Frontend on Blockchain Ledger

Document ID	Process Name	Method Description	Endorsed By / Created By	Signature	Input Awaited
17-13-1	Request for Information	Create RFI	Sub-contractor	$(S)_{SK_{sub-contractor}}$	Contractor
17-13-1	Request for Information	Endorse RFI	Contractor	$(S)_{SK_{contractor}}$	Architect/ Design Team
17-13-2	Request for Information	Update RFI	Architect /Design Team	$(S)_{SK_{Architect}}$	Contractor
17-13-2	Request for Information	Notify Owner			

Fig. 8. Document approval workflow for RFI on the proposed framework

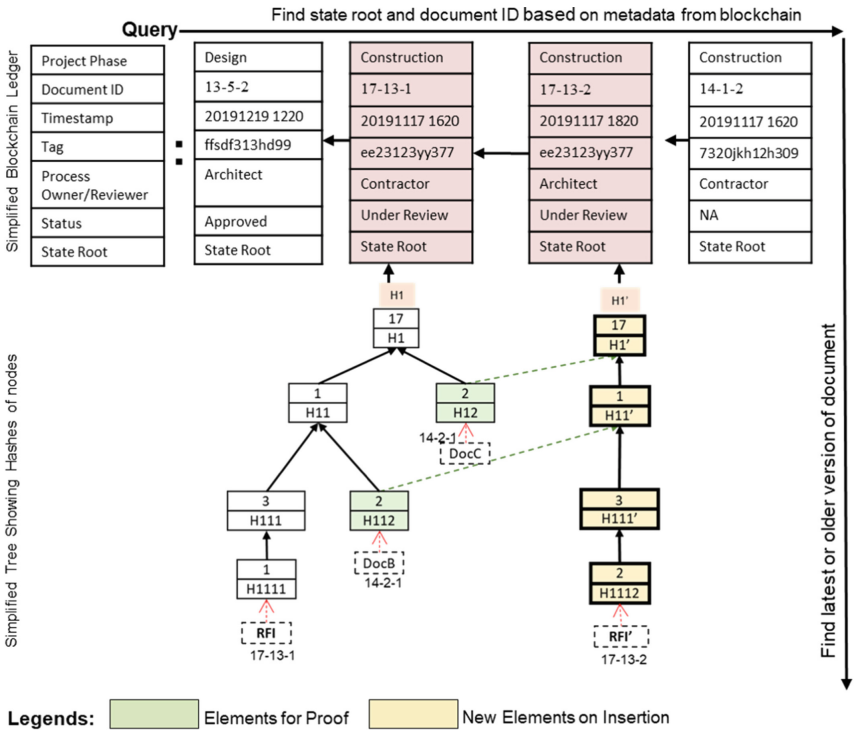


Fig. 9. The proposed blockchain ledger and indexing data structure for RFI management

4 Conclusion

Construction Document Management Systems (CDMS) facilitate coordination among project participants through document approval workflows and facilitating versioning, and searching for documents. Effective document management facilitates a smooth flow of information for various processes among different project participants thereby improving the quality of construction projects. Existing document management systems in construction projects are centralized which are unsuitable due to the fragmented contract-based project organization structure of construction projects. Furthermore, centralized approaches pose several risks such as loss and corruption of data. Therefore in this paper, a distributed construction document management system using blockchain and distributed content-addressable storage is proposed. This framework deploys blockchain smart contracts to facilitate distributed but synchronized or unified document approval workflows and document indexing. A blockchain ledger is deployed to track workflows and versioning of documents to create a credible and irreversible audit-trail. Future work will be done to extend this framework to include complex document approval workflows. The data model of the blockchain ledger and the indexing data structure will be extended to include document metadata to facilitate querying based on multiple criteria.

References

1. Fernando, H., Hewavitharana, T., Perera, A.: Evaluation of electronic document management (EDM) systems for construction organizations, pp. 273–278 (2019)
2. Asılı, H., Tanrıover, O.O.: Comparison of document management systems by meta modeling and workforce centric tuning measures. J. Comput. Sci. Eng. Inf. Technol. **4**(1), 57–67 (2014)
3. Opitz, F., Windisch, R., Scherer, R.J.: Integration of document-and model-based building information for project management support. Procedia Eng. **85**, 403–411 (2014)
4. Cha, H., Lee, D.: Framework based on building information modelling for information management by linking construction documents to design objects. J. Asian Archit. Build. Eng. **17**(2), 329–336 (2018)
5. Xie, L.: Evaluation of electronic document and record management program in Canadian municipality. The University of British Columbia (UBC) (2006)
6. Bazlamit, S.M., Ahmad, H., Ayoush, M.: Document management systems in small and medium size construction companies in Jordan. In: Proceedings of the 6th International Conference on Engineering, Project, and Production Management (EPPM 2015). Griffith University, Gold Coast (2015)
7. Caldas, C.H., Soibelman, L.: Automating hierarchical document classification for construction management information systems. Autom. Constr. **12**(4), 395–406 (2003)
8. CS Institute: MasterFormat (1995)
9. Park, M., Lee, K.-W., Lee, H.-S., Jiayi, P., Yu, J.: Ontology-based construction knowledge retrieval system. KSCE J. Civ. Eng. **17**(7), 1654–1663 (2013)
10. ACONEX: Cloud-based information management: a guide for IT. WhitePaper, Oracle (2018)
11. Autodesk: Autodesk collaboration for Revit security overview. Whitepaper (2016)

12. https://blogs.cisco.com/smallbusiness/the-top-5-security-risks-of-cloud-computing. Accessed 15 Jan 2019
13. Studnia, I., Alata, E., Deswarte, Y., Kaâniche, M., Vincent, N.: Survey of security problems in cloud computing virtual machines, pp. 61–74 (2012)
14. Wood, G.: Ethereum: a secure decentralised generalised transaction ledger. Ethereum Project Yellow Paper (2014)
15. Hasan, H.R., Salah, K.: Combating deepfake videos using blockchain and smart contracts. IEEE Access **7**, 41596–41606 (2019)
16. Khaled, S., Habib, M., Nishara, N., Ala, F.: Blockchain for AI: review and open research challenges. IEEE Access (2019)
17. Mistry, I., Tanwar, S., Tyagi, S., Kumar, N.: Blockchain for 5G-enabled IoT for industrial automation: a systematic review, solutions, and challenges. Mech. Syst. Sig. Process. **135**, 106382 (2020)
18. Chaer, A., Salah, K., Claudio, L., Ray, P.P., Sheltami, T.R.: Blockchain for 5G: opportunities and challenges. In: Proceedings of IEEE GLOBECOM 2019, Waikoloa, USA (2019)
19. Heiskanen, A.: The technology of trust: how the Internet of Things and blockchain could usher in a new era of construction productivity. Constr. Res. Innov. **8**(2), 66–70 (2017)
20. Wang, J.: The outlook of blockchain technology for construction engineering management. Front. Eng. Manag. **4**(1), 67 (2017)
21. Wang, Z., Wang, T., Hu, H., Gong, J., Ren, X., Xiao, Q.: Blockchain-based framework for improving supply chain traceability and information sharing in precast construction. Autom. Constr. **111**, 103063 (2020)
22. Froese, T.M.: The impact of emerging information technology on project management for construction. Autom. Constr. **19**(5), 531–538 (2010)
23. Craig, N., Sommerville, J.: Information management systems on construction projects: case reviews. Rec. Manag. J. **16**(3), 131–148 (2006)
24. Borrmann, A., König, M., Koch, C., Beetz, J.: Building information modeling (2018)
25. CSI: OmniClass (2012)
26. https://medium.com/coinmonks/creating-smart-contracts-with-smart-contract-d54e21d26e00. Accessed 27 Jan 2020
27. Leis, V., Kemper, A., Neumann, T.: The adaptive radix tree: ARTful indexing for main-memory databases, pp. 38–49 (2013)
28. Szydlo, M.: Merkle tree traversal in log space and time, pp. 541–554. Springer, Heidelberg (2004)
29. https://arxiv.org/pdf/1407.3561.pdf. Accessed 19 Mar 2019
30. Gilbert, H., Handschuh, H.: Security analysis of SHA-256 and sisters, pp. 175–193. Springer, Heidelberg (2004)
31. NSA: SHA256 hash generator (2012)
32. Rivest, R.L., Shamir, A., Adleman, L.: A method for obtaining digital signatures and public-key cryptosystems. Commun. ACM **21**(2), 120–126 (1978)
33. http://www.hka-tech.com.au/. Accessed 26 Jan 2020
34. Xu, C., Zheng, C., Xu, J.: vChain: enabling verifiable boolean range queries over blockchain databases, pp. 141–158