Proceedings of the 27th International Conference on Automation & Computing, University of the West of England, Bristol, UK, 1-3 September 2022

Proof of learning: Two Novel Consensus mechanisms for data validation using Blockchain Technology in Water Distribution System

Haitham H. M. Mahmoud
School of Engineering and Built Environment
Birmingham City University
Birmingham, UK
Haitham.mahmoud@bcu.ac.uk

Wenyan Wu
School of Engineering and Built Environment
Birmingham City University
Birmingham, UK
Wenyan.wu@bcu.ac.uk

Yonghao Wang
School of Computing and Digital Technology
Birmingham City University
Birmingham, UK
Yonghao.wang@bcu.ac.uk

***Abstract*—This paper proposes an architecture of a data validation system for Water distribution system (WDS) utilising machine learning as two consensus mechanisms instead of the typical consensus mechanism based on the hashing function. The two consensus mechanisms are called Proof-of-single-learning (PoSL) and Proof-of-multiple-learning (PoML) in which the data is validated based on the learning. These two novel methods are compared with the other five hashing-based consensus mechanisms: Proof-of-Work (PoW), Proof-of-Trust (PoT), Proof-of-Vote (PoV), Proof-of-Assignment (PoA), and proof-of-Authentication (PoAuth) for evaluation. Five case studies of WDS are applied and three performance metrics, and two data conversion methods are utilised. Throughput, latency and operations per transaction (OpT are investigated to evaluate the proposed system.**

Keywords-component; Data validation; Blockchain technology, Crowdsourcing, IoT, Consensus mechanism.

I. INTRODUCTION

Recent research and development of blockchain technologies have led to the realization that blockchain has great potential to reinforce the next generation of water distribution systems (WDS) with redundancy and immutability of the stored data [1]. Also, it verifies the data before transmission to ensure that the original data have not been corrupted. In addition to improving automation in water distribution systems, blockchains facilitate peer-to-peer trading systems to reduce water losses and deliver a transparent and fair water distribution system. In Blockchain, data is transferred in a decentralised network, aiming to facilitate information sharing. These data have been aggregated and timestamped into linked, chain-like blocks, known as ledgers [2]. The ledger consists of digital transactions, records of data, and executions by smart contracts, along with blockchain algorithms and consensus mechanisms as described in [3]. An algorithm for securing digital transactions is the blockchain algorithm, which is said to be a digital self-executing agreement between network peers (e.g., sensors). As discussed earlier, smart contracts run on top of a blockchain algorithm which is a virtual machine to perform further functions [4]. It is thought that Ethereum Virtual Machine (EVM) systems are the most popular type of smart contract supporting infrastructure and Industrial Internet of Things (IIoT) systems. With the consensus mechanism, information is validated using the hash function to reach the required agreement. It is possible to prove this agreement through one or multiple nodes during the validation process, and the selection of these nodes should also be considered during the algorithm development process. A hash function involves the mapping of arbitrary data to fixed-size values, and it is also not possible to invert or reverse the computation [5]. As a simple definition, the blockchain is a data structure that includes interconnections and nodes distributed throughout the network. Changes to data are regarded as a new block in a data structure. Accordingly, the existing data block is not affected. Furthermore, blockchain nodes will be able to store data in a distributed manner. Blocks are generated following a consensus algorithm. Among the consensus algorithms, it provides mechanisms to insert blocks into the blockchain, which have been accepted by the distributed nodes. As an example, a blockchain node initiates a transaction that is broadcast to other nodes in the blockchain for validation according to the consensus mechanism. It is proposed to eliminate the role of the trusted third parties (TTP) by implementing it within the blockchain algorithm. This would remove the need for intermediaries. Thus, the blockchain system is both tamper-resistant and self-enforcing, since it is installed on partners that have sufficient computing resources. As an entity, TTP assists in the exchange of information between parties that trust one another.

Blockchain can be implemented as a means of automating water management and trading systems, thereby supporting operational functions without relying on centralized providers. In wastewater applications, the use-cases are similar; they offer fair trading of untreated water and a seamless, secure process for measuring water quality [6]. Despite the fact that these applications in the water sector are still in the conceptual stage, only a few have succeeded in developing prototypes. Blockchain has the potential to improve water scarcity, fairness, and system security in a very significant way, and both markets and

management systems related to water are promising applications. The Water Market is an extension of the existing concept of a tradable commodity to enable the exchange of water without centralized intervention or centralisation. Another topic discussed in the article is the fair distribution of water based on demand called water rights. It is through the development of water management systems that bursts in water can be detected and operational and information technologies are improved. With these two use cases in mind, it is possible to implement secure communication and data exchange using blockchain technologies and other standard security measures (for example, two-factor authentication, and white and blacklisting). We proposed in our previous work an open-source simulation system for the validation system of data held in WDS using blockchain [7] and proposed a data aggregation system using the blockchain-based on the hash function and bloom filter to maintain the anonymity of the transferred data by representing network peers using pseudonyms as described in [8]. To the best of our knowledge, no study focused on developing a data validation system using blockchain for WDS except our paper [6].

On the other hand, learning machine learning (ML) models can be computationally and memory intensive [9], which may require hardware acceleration. Hence, the concept of crowdsourcing has been developed in the literature to let the users that have the capabilities to process data for benefit tasks online. Data validation in the blockchain can be considered as an example of those data that requires computations and analyses. Therefore, the weak coin concept is introduced in the literature to utilise the crowdsourcing technology for validating the transmitted cryptocurrency called Proof-of-Learning (PoL) [8][9]. The PoL approves the transmitted data by learning, in the contrast Proof-of-work (PoW) matches the hash puzzle to achieve consensus, and Proof-of-vote (PoV) matches the hash puzzle and computes voting to confirm the validation. By the same approach as PoL, the crowdsourcing technology can be implemented on validating the transmitted data and information in the water distribution system. Therefore, proposing an architecture that utilises machine learning in the data validation system instead of the hashing function is essential for the water distribution system. As mentioned, this proposal is inspired by Wekacoin [9] [10] which utilises machine learning in validating and approving cryptocurrency. In this study, two consensus mechanisms utilizing machine learning are proposed, called Proof-of-single-learning (PoSL) and Proof-of-multiple-learning (PoML). These two novel methods are compared with the other five hashing-based consensus mechanisms for evaluation. These consensus mechanisms are Proof-of-Work (PoW), Proof-of-Trust (PoT), Proof-of-Vote (PoV), Proof-of-Assignment (PoA), and proof-of-Authentication (PoAuth). Five case studies of WDS are utilised, three performance metrics, and two data conversion methods are utilised. Throughput, latency and operations per transaction (OpT), in addition to others, are investigated to evaluate the proposed system. This work is considered the extension of our previous work called WDSchain [6]. This paper is organised as follows: Section II discusses and proposes the methodology including the proposed architecture, considered case studies, and the performance metrics realized in this work. Section III analyses and discusses the results. Section IV concludes the work.

II. METHODOLOGY

A. *Proposed Architecture*

By using machine learning methods to validate the transferred data, the PoL approach differs from other hashing mechanisms (See Figure 1). Four steps are discussed in the validation process; initiating a transaction, selecting the dataset and ML algorithm, processing the dataset, and comparing the hashed value (See Figure 1). In a validation, either one validator would be used as the Proof of Single-node Learning (PoSL) or multiple validators would be used as the Proof of Multiple-node Learning (PoML). In PoSL, one blockchain node is selected at random to validate the data with a machine learning model by learning the dataset: in PoML, several nodes are used for validation with the same technique as in Proof-of-stake (PoS) or Proof-of-trust (PoT) with a voting system (see Algorithms 1 and 2). A random dataset and machine learning model are selected by the initiating node to analyse the data, and the hash value of the results of the analysis is sent. In addition to the hashed results, the selected datasets, and the machine learning model are also included as a number along with the data and preceding block hash value. Based on the machine learning algorithm, the validator(s) compare the hash value of the transmitted results with the hash value of the selected dataset. The transaction is authorised to join the chain if the hash values of both transactions are identical. Taking advantage of earlier linked water facility datasets, the processing of these models aims to train and fine-tune the attack detection model in the cloud (assuming there is a second layer of security that has an attack detection system in the cloud). Additionally, it ensures that the original transaction has not been changed during transmission between blockchain nodes.

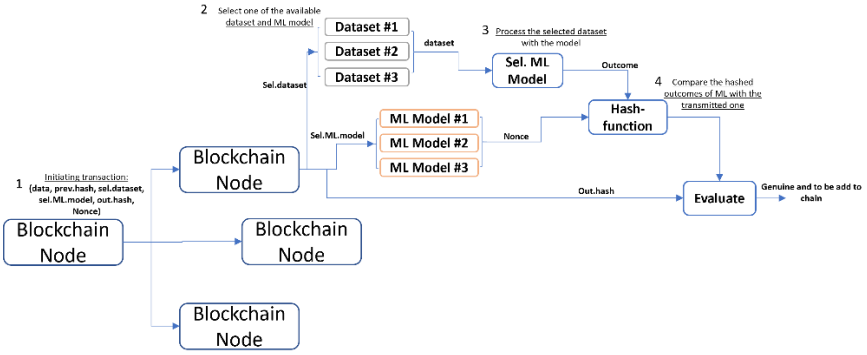

FIGURE 1 PROPOSED FRAMEWORK OF THE PROOF-OF-LEARNING (POL)

Algorithm 1: VerifyNewblock using PoSL.

```
Random selection of the blockchain node
if (selData==1)
    Select dataset1
elseif (selData==2)
    Select dataset2
elseif (selData==3)
    Select dataset3
end
if (selModel==1)
    labels=Generate RandomForest
elseif (selModel==2)
    labels=Generate SVM
elseif (selModel==3)
    labels=Generate KNN
end
Out=DataHash(labels, nonce)
If (outhash, Out)
    tf=true
else
    tf=false
end
```

Algorithm 2: VerifyNewblock using PoML.

```
While (count 1 to Networksize) do
if (selData==1)
    select dataset1
elseif (selData==2)
    select dataset2
elseif (selData==3)
    select dataset3
end
if (selModel==1)
    labels=Generate RandomForest
elseif (selModel==2)
    labels=Generate SVM
elseif (selModel==3)
    labels=Generate KNN
end
Out=DataHash(labels, nonce)
If (outhash, Out)
    res=res+1
else
res=res+0
end
If (res/Networksize)¿=0.6 then
tf=true
else
tf=false
end
end
```

B. Case Studies

There have been two critical uses of blockchain in water systems discussed according to research on the topic: water trading and water management. Water management systems are divided into three subsystems depending on the supply and demand aspects of the system: water supply, water treatment, and water distribution. The study discusses the process of distributing water on the supply side. The treated water and reservoirs are supplied with water via water storage tanks (with level sensors), valves, pumps, and pipelines. To obtain hydraulic data for water distribution on the supply side, the EPANET software is employed. Five case studies are used to assess the validity of data in two types of static and dynamic blockchains (See Table I). An example of C-Town is presented (see Figure 2). These analyses are run on an i5-6200U processor that runs at 2.4 GHz with 8192 MB of RAM. Based upon a top-down simulation, the results are verified by peers on the network. Similarly, the system does not simulate peer-to-peer communication. This means that the time spent transmitting data is ignored.

Identify applicable sponsor/s here. *(sponsors)*

TABLE I. SPECIFICATIONS OF THE CONSIDERED CASE STUDIES.

#	WDS	Description	Specifications
1	D-Town	A residential district in the Eastern part of Exeter city.	407 nodes, 443 pipelines, 11 pumps, 7 tanks, and 1 reservoir.
2	C-Town	A residential district in the Eastern part of Exeter city.	396 nodes, 429 pipelines, 11 pumps, 7 tanks, and 1 reservoir.
3	Net3	EPANET Example.	97 nodes, 119 pipelines, 2 pumps, 3 tanks, and 2 reservoirs.
4	Richmond	District in Town in the UK	872 nodes, 957 pipelines, 7 pumps, 6 tanks, 1 valve, and 1 reservoir.
5	BWSN	A real WDS is "twisted" to preserve their anonymity.	129 nodes, 169 pipelines, 2 pumps, 2 tanks, 46 valves, and 1 reservoir.

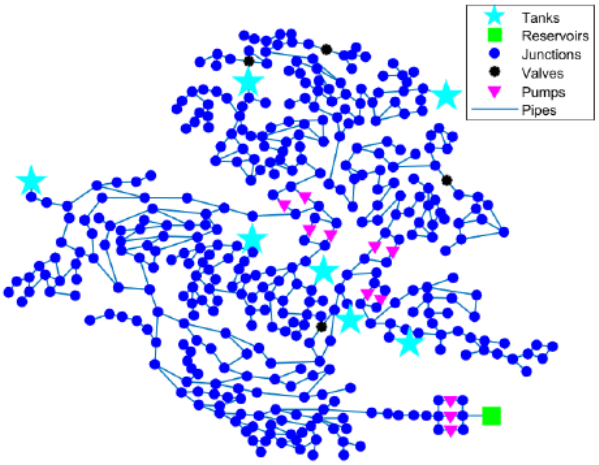

FIGURE 2 C-TOWN DISTRIBUTION MODEL; AN EXAMPLE OF THE CASE STUDIES.

C. Performance Metrics

Several performance indicators are considered to assess the complexity of the system: latency, the number of operations per transaction (*OpT*), and throughput. A transaction takes a certain amount of time to become irreversible and verified, which is called latency. It is based on two coefficients: time spent producing a data block t^G and time spent verifying the data block (T^δ) where δ denotes the consensus mechanism used $\delta \epsilon$ [0, 1, 2, 3, 4] for the five consensus mechanisms. It is indicated by:

$$t^L = T^G + T^\delta \quad (1)$$

OpT indicates the number of operations that need to be performed to verify the data. Measurement of OpT helps determine the complexity of the consensus mechanism. Transactional throughput (S^T) is estimated by counting the number of transactions per second. By definition, throughput is based on both the number of transactions (N^T) and the latency (T^L). It is defined as:

$$S^T = \frac{N^T}{T^L} \quad (2)$$

When it comes to security, most consensus mechanisms such as Proof-of-Work (PoW) and PoT are designed to give a high probability of security, since the network might be vulnerable if a disproportionate amount of the mining power is possessed. Moreover, if a significant proportion of the mining power is biased (more than 33.3%), there might be clear biasing for a Proof-of-Vote (PoV) method. The following steps can be employed, therefore, to ensure that any consensus method (δ) is secure from malicious verifiers (v^M):

$$v^M \leq v^{T,\delta} \quad (3)$$

$$v^{T,\delta} = \left\lfloor \frac{N-1}{3} \right\rfloor \quad (4)$$

Based on the definition described above, $v^{T,\delta}$ denotes the number of true validators for certain consensus mechanisms (δ), and N signifies the number of nodes or validators. Despite the relatively small possibility ($\frac{1}{N}$) that a malicious verifier will utilize Proof-of-Assignment (PoA), it is a fair idea to develop this method as it requires one of the least amounts of computing resources.

III. Results and Discussion

However, even though mining with POSL and POML requires more time, and the system is more sophisticated, compared to current PoW systems, these consensus techniques give significant job opportunities in mining calculations. These two methods are assessed by Latency, OpT and throughput in five case studies in water models (see Table II). PoSL and PoML reduce latency and throughput by 25% and 63%, respectively, in comparison with PoW (See Figures 3-6). In the same approach, the throughput of other consensus mechanisms such as PoW and PoV is better than PoSL and PoML. The PoML method requires the machine learning model to be run twice and doubles the number of blockchain nodes, so the mining time is generally larger. Also, the mining time of the PoML and PoSL may vary, since the dataset is randomly selected, and the machine learning model is chosen at random. Our datasets and the complexity of the machine learning models are designed to enable rapid processing; however, the size of the datasets and the size of the models vary from small to medium-sized. Furthermore, the datasets employed for the computation puzzles are identical to those used for the detection model, enabling further tuning of the attack detection model.

Among datasets, the OpT measure is the same since it is fixed for all PoML and PoSL by 23 transactions and from 23 to 89 depending on the number of blockchain nodes, respectively. There can be demonstrated that the OpT of PoSL is equal to that of PoW, however, the PoML method requires more operations in each transaction (see Figure 5). Depending on the randomly selected dataset and machine learning method, the miners' lowest and highest average mining times for each block range from 0.15 to 1.9 minutes and 0.22 to 2.9 minutes, respectively (see Figure 6). These two techniques are less efficient than the others, but they entail that mining computations are spent on meaningful tasks in the form of training and fine-tuning the detection model and ensuring that there is no one point of trust in the system. Moreover, if the link between the initiating node and the other blockchain peers is compromised, the transaction will be rejected. PoML offers lower performance while providing greater security, since each node of the blockchain utilizes its own randomly selected datasets and machine learning algorithm. In order to achieve these objectives, it can use a crowd-sourcing model in which blockchain nodes communicate with an Internet public database to solve computational puzzles, such as Kaggle.

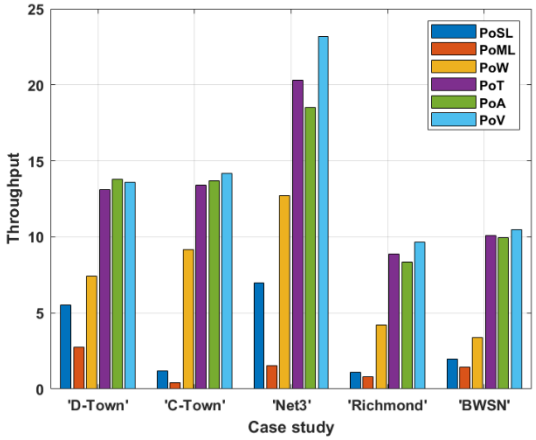

Figure 3 Evaluation of the Throughput For PoSL and PoML with the Others.

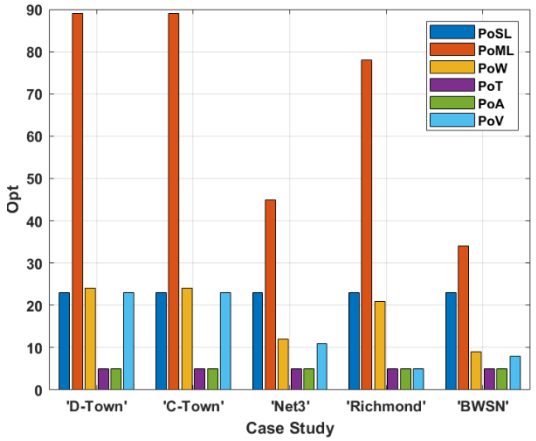

Figure 4 Evaluation of the OpT For PoSL and PoML with the Others.

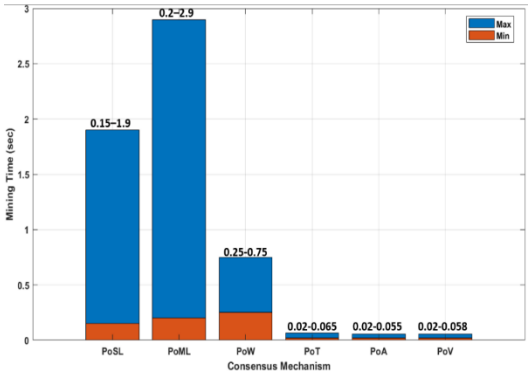

Figure 5 The Minimum and Maximum Mining Time for one Transaction in the Data

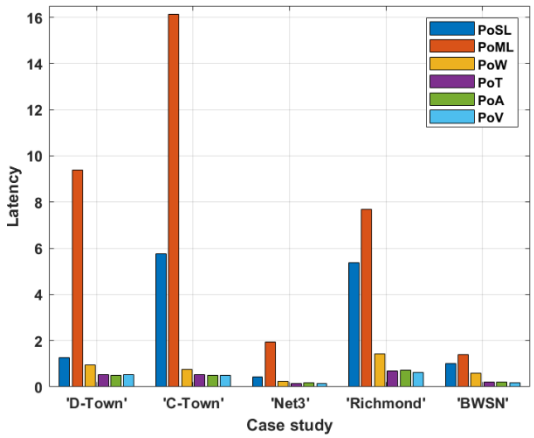

FIGURE 6 EVALUATION OF THE LATENCY FOR POSL AND POML WITH THE OTHERS.

TABLE II. PERFORMANCE EVALUATION OF THE POSL AND POML MECHANISMS.

WDS	Consensus Mechanism	N^T	T^G (s)	T^δ (s)	T^L (s)	S^T (TPS)	OpT
D-town	PoSL	7	0.07	1.1	1.2	5.52	23
	POML			9.3	9.3	2.74	89
C-Town	PoSL	7	0.07	5.6	5.7	1.21	23
	POML			16.0	16.1	0.43	89
Net3	PoSL	3	0.03	0.3	0.4	6.96	23
	POML			1.9	1.9	1.55	45
Richmond	PoSL	6	0.06	5.3	5.3	1.11	23
	POML			7.6	7.6	0.78	78
BWSN	PoSL	2	0.02	0.9	1.0	1.99	23
	POML			1.3	1.3	1.42	34

IV. CONCLUSION

In validation of data transmission for WDS, two consensus mechanisms having machine learning features such as categorization are developed in this study. The objective of this procedure is to convert the effort invested in the mining computations into something useful for training and tuning the attack detection model in the cloud. Further, it is designed to prevent the creation of a single point of trust throughout all systems.

Among the suggested techniques are PoSL and PoML, which involve delivering a randomly selected dataset, a machine learning model, and a hashed value of the processing result. Regarding PoSL and PoML, the calculation is completed by one random node or by several random nodes. Despite the low performance of the two previously mentioned consensus methods compared to the present consensus mechanism, this mechanism has the advantage of enabling a significant amount of work to be accomplished through mining computing. In addition to the immutability, decentralisation, and transparency that blockchain technology provides, it also enhances the automation of the WDS. Automating the process of water delivery will allow the monitoring and minimization of water bursts and pollution. To evaluate system complexity, three performance metrics are analysed: latency, OpT, and throughput, as well as two additional coefficients to quantify system complexity by comparing mining consensus mechanisms.

ACKNOWLEDGEMENT

This research has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Training Networks (ITN)-IoT4Win grant agreement No [765921].

References

[1] Dogo, E.M.; Salami, A.F.; Nwulu, N.I.; Aigbavboa, C.O. Blockchain and Internet of things-based technologies for the intelligent water management system. In Artificial Intelligence in IoT; Springer: Antalya, Turkey, 2019; pp. 129–150.

[2] Suciu, G.; N˘adrag, C.; Istrate, C.; Vulpe, A.; Ditu, M.C.; Subea, O. Comparative analysis of distributed ledger technologies. In Proceedings of the 2018 Global Wireless Summit (GWS), Chiang Rai, Thailand, 25–28 November 2018; pp. 370–373.

[3] Andoni, M.; Robu, V.; Flynn, D.; Abram, S.; Geach, D.; Jenkins, D.; McCallum, P.; Peacock, A. Blockchain technology in the energy sector: A systematic review of challenges and opportunities. Renew. Sustain. Energy Rev. 2019, 100, 143–174.

[4] Fontein, R. Comparison of static analysis tooling for smart contracts on the EVM. In Proceedings of the 28th Twente Student Conference on IT, Twente, The Netherlands, 2 February 2019.

[5] Mingxiao, D., Xiaofeng, M., Zhe, Z., Xiangwei, W. and Qijun, C., 2017, October. A review of the consensus algorithm of blockchain. In 2017 IEEE international conference on systems, man, and cybernetics (SMC) (pp. 2567-2572). IEEE.

[6] Campbell, R. The Genesis System Wants to Record Cleaned Fracking Water on the Blockchain; Bitcoins Magazine: Nashville, TN, USA, 2017.

[7] Mahmoud, H.H., Wu, W. and Wang, Y., 2021. WDSchain: A Toolbox for Enhancing the Security Using Blockchain Technology in Water Distribution System. Water, 13(14), p.1944.

[8] Mahmoud, H.H.M.; Wu, W.; Wang, Y. Secure Data Aggregation Mechanism for Water Distribution System using Blockchain. In Proceedings of the 2019 25th International Conference on Automation and Computing, Lancaster, UK, 5–7 September 2019; pp. 1–6.

[9] Jia, H., Yaghini, M., Choquette-Choo, C.A., Dullerud, N., Thudi, A., Chandrasekaran, V. and Papernot, N., 2021, May. Proof-of-learning: Definitions and practice. In 2021 IEEE Symposium on Security and Privacy (SP) (pp. 1039-1056). IEEE.

[10] Bravo-Marquez, F., Reeves, S. and Ugarte, M., 2019, April. Proof-of-learning: a blockchain consensus mechanism based on machine learning competitions. In 2019 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPCON) (pp. 119-124). IEEE.