Log Data Integrity Solution based on Blockchain Technology and IPFS

https://doi.org/10.3991/ijim.v16i15.31713

Yassine Azizi(✉), Mostafa Azizi, Mohamed Elboukhari
Lab. MATSI, ESTO, University Mohammed 1st, Oujda, Morocco
Y1.azizi@ump.ac.ma

Abstract—Threats towards information systems have continued to increase and become more sophisticated, making security approaches a necessity for all types of organizations to ensure their protection. To implement an appropriate computer security policy, it is necessary to efficiently exploit the data that has become a valuable asset for these security systems, provided it is well used, controlled and monitored. In this paper, we focus on developing a decentralized solution based on Blockchain technology and IPFS (Inter Planetary File System) that can maintain and ensure the integrity of log files and sensitive information. The obtained results are promising, we obtained a distributed ledger of all log file transactions in a chronological sequence, which was shared among all Ethereum participants, allowing us to verify the log files' integrity, validity, and auditability throughout their life cycle.

Keywords—blockchain, Ethereum, security, IPFS, log files, smart contract

1 Introduction

In the current context of the Internet of Things, Cloud Computing, Big Data and the interconnection of systems, the frequency and impact of cybersecurity incidents are strongly increasing [1]. Each organization has a considerable amount of data and information, and the major challenge is to find the most reliable security solutions, hopefully to guarantee the traceability, accessibility, availability and integrity of data transfers.

Among these data, we can mention those of log files, which present a mine of information allowing to measure the security level [2], to detect possible threats and to trigger possible actions to be taken, all in real time or not. The extraction of useful knowledge using log files raises the issue of data integrity, because with each operation on the system, the risks of altering log data, whether voluntary or involuntary, are numerous. As an example, the post-intrusion steps during an attack are the deletion of the traces that the hacker may have left. This is done, in part, by deleting the logs and histories that can provide insight into the actions taken during the intrusion and possibly into a network identity, so ensuring a methodology that can guarantee the exhaustiveness, accuracy, precision and validity of the log files throughout their life cycle is our real challenged concern.

To address this specific vulnerability, we propose a solution based both on Blockchain and on IPFS (Inter Planetary File System). The blockchain with its immutable and irreversible character of the content, combined with a decentralized file storage system such as an IPFS, can provide a technical solution that allows maintaining the integrity of log files.

The rest of this paper is organized as follow. The second section presents the theoretical background. In the third section, we explain in details our proposed approach based on Ethereum networks. We present and discuss the obtained results in the fourth section. The fifth section concludes this paper and gives some perspectives to this work.

2 Theoretical background

2.1 Log files

Event logs are text files that record chronologically the events executed by a server or a computer application. The log files are a very important source of information [3]; they retrace all the events that occur during the activity of the system. These are often of great volume and come from everywhere, operating systems, application servers, sensors, etc.

Several research studies consider log files as a very useful data source in several areas: e-commerce [4], security [5, 6], cloud computing [7].

2.2 Blockchain

Blockchain is a modern secured technology for storing and transmitting information. It operates without a central control organ and provides transparency and security through the validation of transactions by the network nodes [8].

The blockchain contains the history of all data exchanges made since its creation. As these data are shared between all its users and without intermediaries, everyone can verify their validity and confirm their integrity.

The three main principles inherent in this blockchain technology are [9]:

- Distributed ledger technology : All network participants have access to the distributed ledger and its unalterable record of transactions. With this shared ledger, transactions are recorded only once, eliminating duplication of tasks.
- Unalterable transactions : No participant can change or alter a transaction once it has been posted to the shared ledger. If a transaction record has an error, a new transaction must be added to reverse the error, and both transactions are then visible.
- Smart contracts : To speed up transactions, a set of rules and predefined instructions, called a smart contract, is deployed in the blockchain and executed automatically.

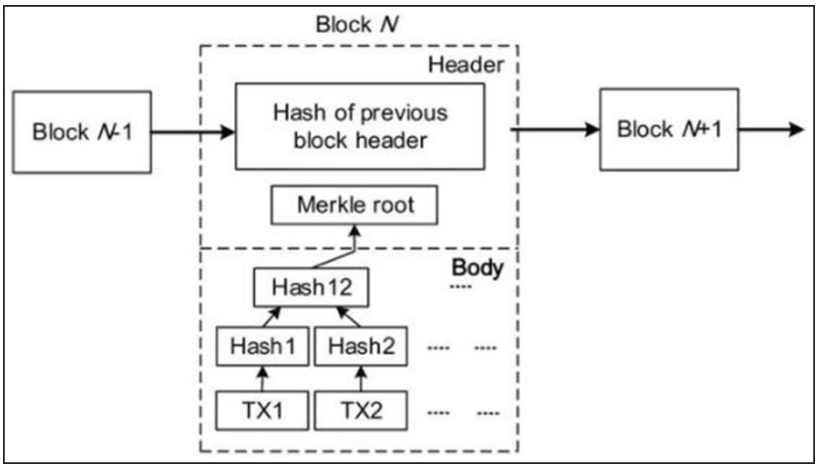

Fig. 1. Blockchain structure [10]

2.3 Ethereum

Ethereum is an online exchange protocol, based on the blockchain, and open to decentralized applications [11].

The Ethereum network is composed of thousands of computers connected. They share a set of unalterable chained blocks called blockchain, where all transactions are recorded forever.

Ethereum is also the name of a cryptocurrency based on this technology, which is the second widely used cryptocurrency after BTC.

2.4 IPFS

IPFS (Inter Planetary File System) is a decentralized file system that seeks to guarantee security, privacy and resistance to data censorship.

This peer-to-peer file sharing technique removes the possibility of censorship or unilateral deletion, because no single party owns or controls the data once it has been uploaded to the network, IPFS is a system that works according to the "search by content" scheme, i.e. every time we perform a search in IPFS, we have to tell the system "what we are looking for" instead of telling it "where to look for it" [12].

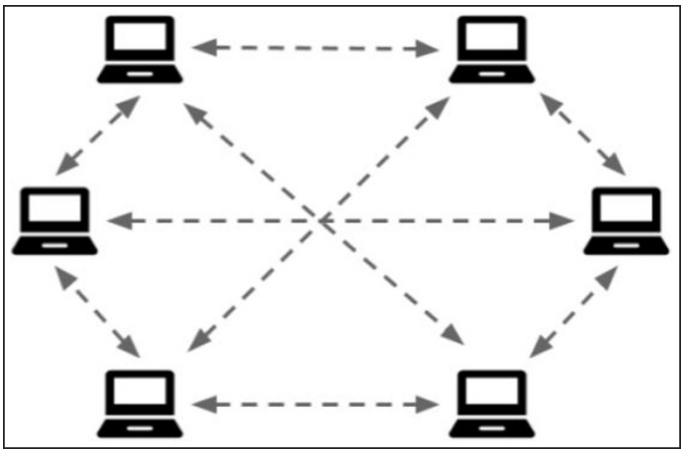

Fig. 2. Movement of data in IPFS [13]

3 Literature review

Since the invention of the Blockchain, its applications have multiplied. Today, the use cases are very varied, and blockchain is becoming essential for many sectors. In the literature, several studies have been conducted based on blockchain technology, one of these research works is [14], in which the authors use the blockchain technology to build a secure electronic voting system that offers the equity and privacy of current voting systems, while providing transparency and flexibility. This system addresses some of the limitations of other existing systems, improves security, and reduces the cost of running a national election. In [15], Tanwar et al. propose an access control policy algorithm to improve data accessibility between healthcare providers, to implement the Hyperledger-based electronic medical record sharing system that uses the blockchain concept, this solution can reform the interoperability of healthcare data-bases, providing increased access to patient medical records, device tracking, prescription databases, and hospital assets.

In [16], Smart cities support mission-critical applications that require protecting data and functionality from malicious and unauthorized use. Equipping the supporting platforms with appropriate means for access control is demanding. To this end, they have proposed a new solution for distributed management of identity and authorization policies based on blockchain technology.

The field of education is also impacted by blockchain technlogy, papers [17, 18] have evaluated the advantages of blockchain technology and advocate a decentralized trust model for transactions based on an academic cryptocurrency. Blockchain is used in their approaches to manage content, instructional, and skill transactions, which are evaluated by students, instructors, and employers through consensus., this proposal has the potential to eliminate the gap between academia and the world of work, and aims to address the current challenges of an increasingly dispersed, open and pervasive higher education.

In the age of Industry 4.0, most of digital payments are made through applications supported by a variety of payment gateways. Attackers can use these heterogeneous payment gateways to carry out harmful operations like as wallet account hacking. In [19], authors proposed a solution named BloHosT (Blockchain Enabled Smart Tourism and Hospitality Management). It is a framework that allows tourists to communicate with several stakeholders using a single wallet identifier linked to a cryptocurrency server to initiate payments. BloHosT uses an immutable ledger that eliminates the need for proofs during travel, allowing tourists to enjoy a stress-free vacation.

4 Proposed methodology

The major goal of this contribution is to offer an appropriate and efficient architecture for securing log data, ensuring their integrity, and ensuring the log files' survival in the scenario of destruction or unauthorized modifications.

This proposal is based on blockchain technology, which provides the best protection against data loss or corruption by duplicating all transactions and making them available on each network node. They cannot be changed or removed without the other members of the network agreeing.

The system architecture is presented in the following Figure 3:

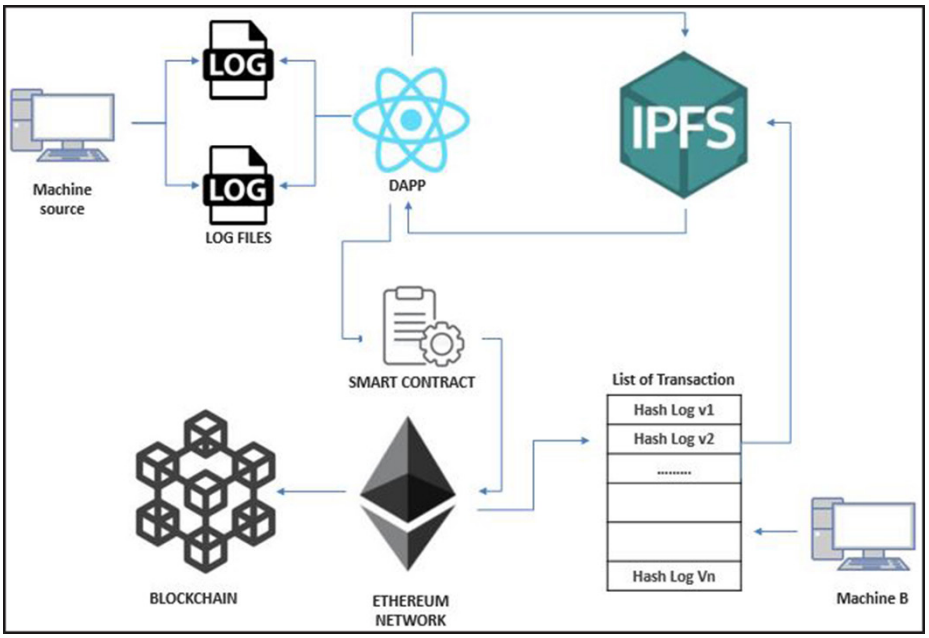

Fig. 3. Proposed approach

In our proposed methodology, we combine two innovative distributed technologies: a peer-to-peer network for file storage (IPFS) and the Ethereum blockchain to record the unique fingerprint (hash) of each version of a log file saved in the storage system.

Cryptographic hashing algorithms can make it easier to verify that each block's transaction data cannot be changed and that linked blocks in the blockchain can't be deleted. The reason for this combination is to overcome the storage limit in a blockchain network, since IPFS enables an entirely decentralized data storage with high scalability and uses distributed hash table technology (DHT).

A DApp (Decentralized application) created specifically for this project facilitates the interface between machine source that contains the log file to be secured, smart contract, and IPFS.

Our DApp provides true traceability of changes made. It continuously monitors our system log files to provide instant, accurate and complete information on access events and attempts. Once the log file is changed, it is uploaded to the IPFS via Dapp along with complete log events.

When a log file is uploaded to IPFS, IPFS generates hash of the log data version and returns it to the DApp. The smart contract will receive the IPFS hash and create a new transaction on the blockchain network with the IPFS hash as input data. The ethereum network will store smart contract transactions with log file hash data.

On the Ethereum network, all members (nodes) share the same copy of the logs database. The blockchain is composed of blocks, each of which contains a list of transactions. Each block has a timestamp, its own cryptographic hash and a hash reference of the previous block listed in the block header on the blockchain structure. The entire log file will be saved on IPFS, whereas the Blockchain will just store the IPFS log file's hash. The hash log files that are recorded on the Blockchain, are the links to log files that are saved on IPFS, which are later accessible by network nodes.

5 Implementation and results

In order to implement our proposed solution, we have to go through several essential steps described as follows:

The first step in constructing a system is to install and open a Metamask account in order to obtain a wallet and manage Ethereum Blockchain transactions, Metamask is a cryptocurrency wallet based on the Ethereum Virtual Machine. It is therefore compatible with any EVM-based blockchain. It is a browser extension that allows you to interface with Dapps (decentralized applications).

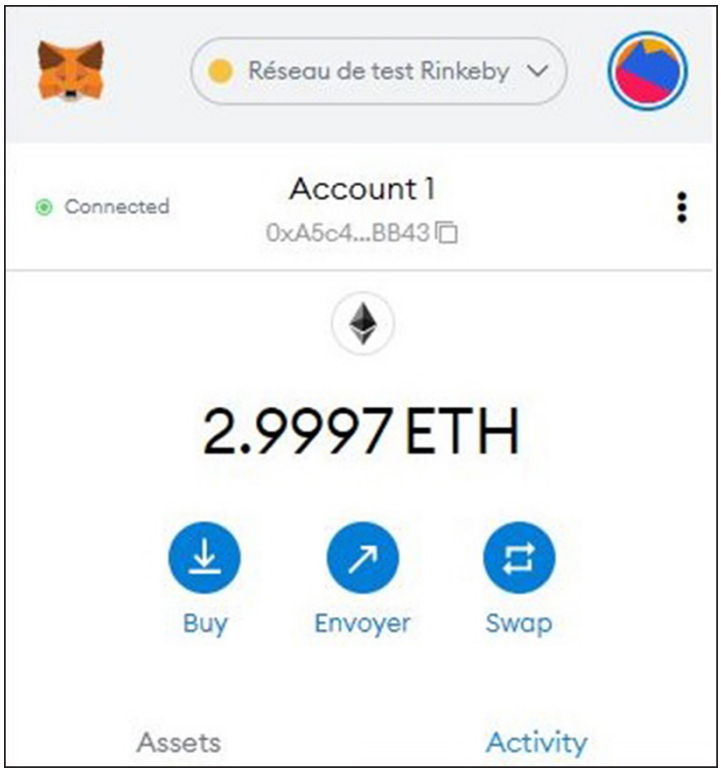

Fig. 4. Metamask interface

On Metamask (Figure 4), we select the Rinkebey test network which is a network used by protocol or smart contract developers to test, in a production environment, both protocol upgrades and smart contracts prior to their deployment on the main network.

Then, we build our DApp using node.js and the necessary NPM (Node Package Manager) dependencies like (Figure 5).

```
npm i ethereumjs-tx
npm i fs
npm i log-timestamp
npm i react
npm i web3-eth
npm i web3
npm i eth-provider
npm i ipfs-http-client
```

Fig. 5. NPM dependencies

DApp will manage the interaction between the log file, IPFS and Ethreum blockchain and it contains information about the user's private Ethereum wallet address, log file path, public contract address and others.

The next step is to create and deploy a smart contract on Ethreum Remix IDE using Solidity programming language, we compile the smart contract and use injected Web3 to deploy it.

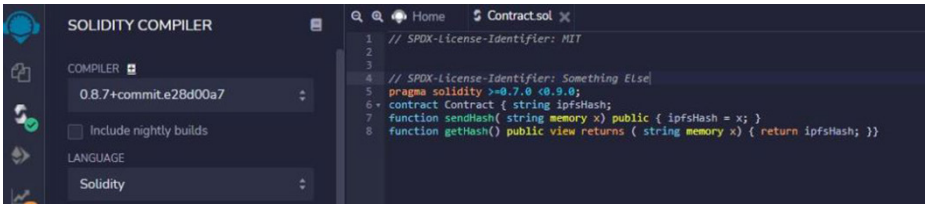

Fig. 6. Ethereum smart contract deployment

We use Etherscan, an Ethereum BlockExplorer, to verify the contract deployment. A BlockExplorer is a search engine that allows users to find transactions on the Ethereum blockchain easily and accurately. Etherscan improves Blockchain transparency by indexing and making all Ethereum Blockchain transactions accessible in the most transparent and easy way possible.

When we have a change in our log file, the DApp will load the new version of the log in the buffer and will send it to IPFS by apfs-api, IPFS will save the log file and return the hash from the file by IPNS to the web DApp and smart contract Ethereum.

Ethereum will use the smart contract with input data (IPFS hash), by sending a transaction, the hash of the log file was saved in blockchain.

On the Blockchain, smart contract transactions with file hash data will be stored and and transaction recipts will be transmitted to the DApp web via web3. Transaction receipt is smart contract transaction information such as transaction hash, nonce, Timestamp, hash block, block number, input data (ipfs hash) gas used, gas price, and others.

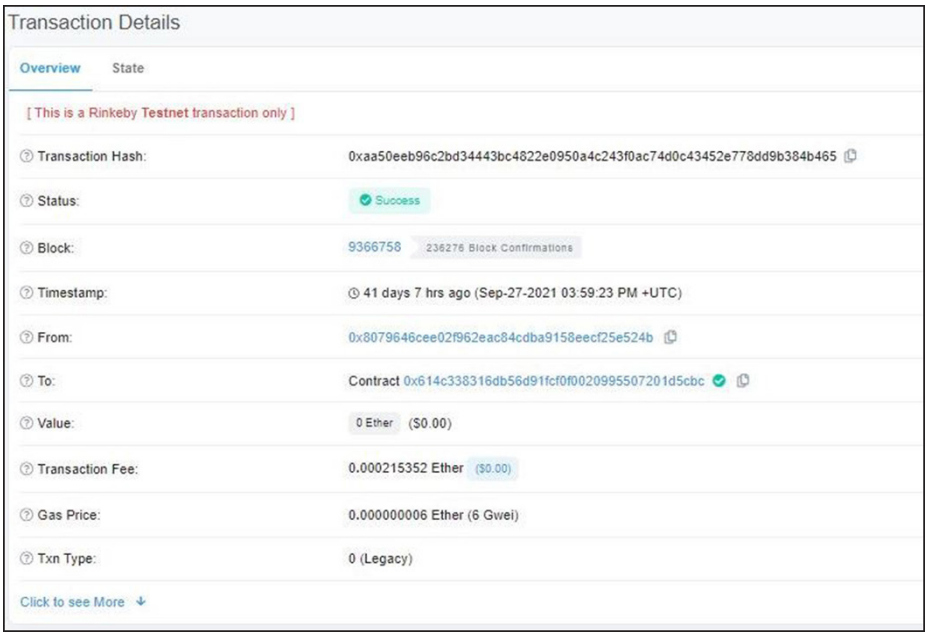

Fig. 7. Transaction details

On Etherscan you can track the different transactions and consult the versions of the log file.

Fig. 8. List of transactions

Each transaction contains input data in the form of an IPFS hash that refers to a dated version of the log file.

Fig. 9. Transaction details: IPFS hash

Using this hash stored with the IPFS gateway URL we can retrieve the full version of our file on IPFS.

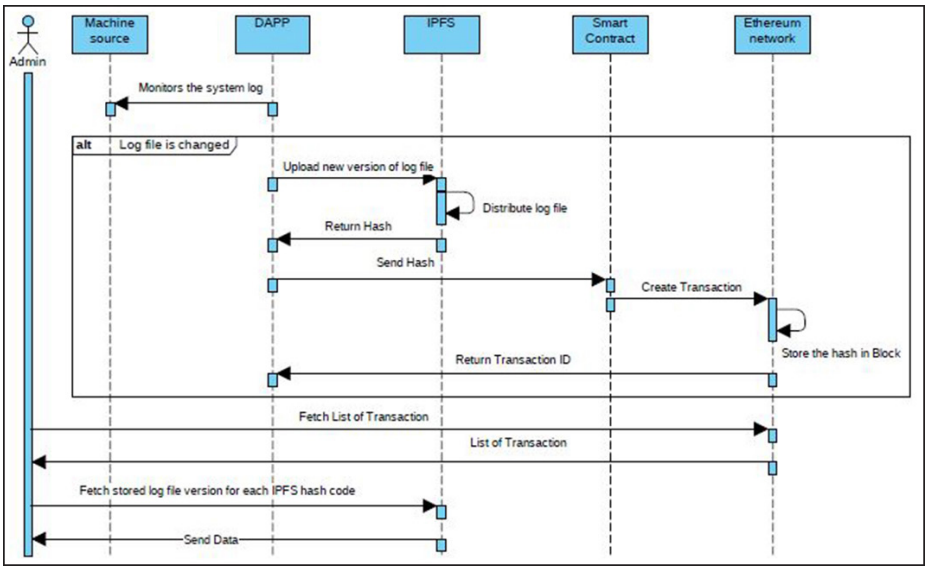

Fig. 10. Sequence diagram for different interaction types

As a results, we have obtained a register of all the log files transactions in chronological order, this register is shared between all the users of the Ethereum network and they can access the different versions of the log files. If the log source machine is attacked and a hacker tries to modify the file in question, he would not be able to modify the hash stored on the Blockchain. Another member of our Ethereum network would then be able to compare the legitimate hash with the hash of the compromised file and would immediately know that the data has been modified and can access the original version. This blockchain integrity and historization feature helps us to ensure the integrity, validity, and auditability of log files throughout their lifecycle.

6 Conclusion

In the today's context of big data, where the volume of data processed and stored is greater than ever, it is critical to take measures to protect the integrity of the data collected for analysis and intrusion detection systems. To preserve log data of our context, we present in this paper a methodology to secure log files and sensitive data in order to solve the problem of assuring the survival and tracking of log file modifications.

Our proposed solution is mainly based on blockchain technology, which provides for the decentralized, secure, and transparent recording of a set of transactions in the form of a chain of blocks. Since the cost of storing huge files in the blockchain is prohibitively expensive, our proposed methodology saves only the file's fingerprint on the chain rather than the file itself. We did this by storing the log files on the IPFS and generating a hash code that would be saved to the blockchain.

Furthermore, we can see that our hybrid approach, which uses Blockchain and IPFS, and interacts with a decentralized application dedicated to coordination, has provided us with an accurate, transparent and immutable view of the time-stamped version history of the log files.

Our future work will focus on the integration of Blockchain technology in an IoT environment in order to secure the log files of IoT devices in the hope of preventing some cyberattacks that target the availability and integrity of logs.

7 References

[1] M. I. Alghamdi, "Survey on Applications of Deep Learning and Machine Learning Techniques for Cyber Security", International Journal of Interactive Mobile Technologies, Vol 14, No 16, 2020. https://doi.org/10.3991/ijim.v14i16.16953

[2] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, "A Survey of Network-Based Intrusion Detection Data Sets", Computers & Security, Vol 86, pp. 147–167, 2019. https://doi.org/10.1016/j.cose.2019.06.005

[3] Y. Azizi, M. Azizi, and M. Elboukhari, "Tracking Attacks Data Through Log Files Using Mapreduce", In International Conference Europe Middle East & North Africa Information Systems and Technologies to Support Learning, pp. 331–336, 2018. https://doi.org/10.1007/978-3-030-03577-8_36

[4] A. Muneer, S. Razzaq, and Z. Farooq, "Data Privacy Issues and Possible Solutions in E-commerce", Journal of Accounting & Marketing, Vol 7, no 3, pp. 1–3, 2018.

[5] Y. Azizi, M. Azizi, and M. Elboukhari, "Log Files Analysis Using MapReduce to Improve Security", Procedia Computer Science, Vol 148, pp. 37–44, 2019. https://doi.org/10.1016/j.procs.2019.01.006

[6] Y. Azizi, M. Azizi, and M. Elboukhari, "Anomaly Detection from Log Files Using Multidimensional Analysis Model", In International Conference on Digital Technologies and Applications. pp. 515–524. 2021. https://doi.org/10.1007/978-3-030-73882-2_47

[7] M. Jelidi, A. Ghourabi, and K. Gasmi, "A Hybrid Intrusion Detection System for Cloud Computing Environments," 2019 International Conference on Computer and Information Sciences (ICCIS), pp. 1–6. 2019. https://doi.org/10.1109/ICCISci.2019.8716422

[8] B. K. Mohanta, D. Jena, S. S. Panda, and S. Sobhanayak, "Blockchain Technology: A Survey On Applications And Security Privacy Challenges", Internet of Things, Vol 8, p. 100107, 2019. https://doi.org/10.1016/j.iot.2019.100107

[9] E. Bertino, A. Kundu, and Z. Sura, "Data Transparency With Blockchain And Ai Ethics", Journal of Data and Information Quality (JDIQ), Vol 11, No 4, pp. 1–8, 2019. https://doi.org/10.1145/3312750

[10] J. Wang, Q. Wang, N. Zhou, and Y. Chi, "A Novel Electricity Transaction Mode Of Microgrids Based On Blockchain And Continuous Double Auction", Energies, Vol. 10, No. 12, p. 1971, 2017. https://doi.org/10.3390/en10121971

[11] K. R. Benita, S. G. Kumar, M. B, and M. A, Authentic Drug Usage and Tracking with Blockchain Using Mobile Apps. International Journal of Interactive Mobile Technologies (iJIM), Vol 14, No 17, pp. 20–32, 2020. https://doi.org/10.3991/ijim.v14i17.16561

[12] Q. Zheng, Y. Li, P. Chen, and X. Dong, "An Innovative IPFS-Based Storage Model for Blockchain," *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 704–708, 2018. https://doi.org/10.1109/WI.2018.000-8

[13] Á. Tenorio-Fornés, S. Hassan, and J. Pavón, 'Peer-to-Peer System Design Trade-Offs: A Framework Exploring the Balance between Blockchain and IPFS', Applied Sciences, Vol. 11, No. 21, p. 10012, 2021. https://doi.org/10.3390/app112110012

[14] F. Þ. Hjálmarsson, G. K. Hreiðarsson, M. Hamdaqa, and G. Hjálmtýsson, "Block-Chain-based e-Voting System", In 2018 IEEE 11th international conference on cloud computing (CLOUD), pp. 983–986, 2018. https://doi.org/10.1109/CLOUD.2018.00151

[15] S. Tanwar, K. Parekh, and R. Evans, "Blockchain-Based Electronic Healthcare Record System For Healthcare 4.0 Applications", Journal of Information Security and Applications, Vol 50, p. 102407, 2020. https://doi.org/10.1016/j.jisa.2019.102407

[16] C. Esposito, M. Ficco, and B. B. Gupta, "Blockchain-Based Authentication and Authorization For Smart City Applications", Information Processing & Management, Vol 58, No 2, p. 102468, 2021. https://doi.org/10.1016/j.ipm.2020.102468

[17] D. Lizcano, J. A. Lara, B. White, and S. Aljawarneh, "Blockchain-Based Approach To Create A Model Of Trust In Open And Ubiquitous Higher Education", J. Comput. High. Educ., Vol 32, No 1, pp. 109–134, Apr 2020. https://doi.org/10.1007/s12528-019-09209-y

[18] Y. Ma and Y. Fang, "Current Status, Issues, and Challenges of Blockchain Applications in Education", Int. J. Emerg. Technol. Learn., Vol. 15, No. 12, pp. 20–31, Jun. 2020. https://doi.org/10.3991/ijet.v15i12.13797

[19] U. Bodkhe, P. Bhattacharya, S. Tanwar, S. Tyagi, N. Kumar, and M. S. Obaidat, "BloHosT: Blockchain Enabled Smart Tourism And Hospitality Management", In 2019 international conference on computer, information and telecommunication systems (CITS), pp. 1–5, 2019. https://doi.org/10.1109/CITS.2019.8862001

8 Authors

Yassine Azizi is a Ph.D. student in Computer Science, at the faculty odf Sciences, Mohammed First University in Oujda, Morocco. He holds a Master degree in Information Systems Engineering from Faculty of Sciences Semlalia, Cadi Ayyad University in 2016, and a B.Sc. degree in Computer Engineering from The Faculty of Science of Tetouan, Abdelmalek Essaadi University in 2013. His research focuses on big data analytics, security, Data Science and Software Engineering. He can be contacted at azizi.yass@gmail.com

Mostafa Azizi received a State Engineer degree in Automation and Industrial Computing from the Engineering School EMI of Rabat, Morocco in 1993, then a Master degree in Automation and Industrial Computing from the Faculty of Sciences of Oujda, Morocco in 1995, and a Ph.D. degree in Computer Science from the University of Montreal, Canada in 2001. He earned also tens of online certifications in Programming, Networking, AI, and Computer Security. He is currently a Professor at the ESTO, University Mohammed 1st of Oujda. His research interests include Security and Networking, AI, Software Engineering, IoT, and Embedded Systems. His research findings with his team are published in over 100 peer-reviewed communications and papers. He also served as PC member and reviewer in several international conferences and journals.

Mohamed Elboukhari received the MSc degree in computer science in 2005 from the University of Science, Oujda, Morocco. He is currently a Professor in the University of Oujda in the field of computer science. His research interests include computer security, web tracking and wireless network security.

Article submitted 2022-04-15. Resubmitted 2022-05-29. Final acceptance 2022-06-03. Final version published as submitted by the authors.