Blockchain Intelligence: When Blockchain Meets Artificial Intelligence

Zibin Zheng, *Senior Member, IEEE*, Hong-Ning Dai, *Senior Member, IEEE*, Jiajing Wu, *Senior Member, IEEE*

***Abstract*—Blockchain is gaining extensive attention due to its provision of a secure and decentralized resource sharing manner. However, the incumbent blockchain systems also suffer from challenges in aspects of operational maintenance, quality assurance of smart contracts, and malicious behavior detection of blockchain data. The recent advances in artificial intelligence bring opportunities in overcoming the above challenges. The integration of blockchain with artificial intelligence can be beneficial to enhance current blockchain systems. This article presents an introduction to the convergence of blockchain and artificial intelligence (namely blockchain intelligence). This paper also gives a case study to further demonstrate the feasibility of blockchain intelligence and point out future directions.**

Keywords*—*Blockchain; Artificial Intelligence; Smart Contract; Machine Learning

I. Introduction

BLOCKCHAIN has received extensive attention recently due to its provision of secure data sharing services with traceability, immutability, and non-repudiation. Despite the merits of blockchain, the development of blockchain technologies has undergone challenges including poor scalability, difficulties in operational maintenance, detecting vulnerable codes in smart contracts, and identifying malicious behaviors in blockchain historical data.

The recent advances in artificial intelligence (AI) have greatly propelled the evolution of diverse business applications. The integration of AI with blockchain has the potentials to overcome the limitations of blockchain. We name the intelligent capability bestowed by AI to blockchain as *blockchain intelligence*. In particular, AI approaches may help to capture the abnormal behaviors in blockchain after analyzing the blockchain data, detecting, and identifying possible vulnerable program codes in smart contracts. Consequently, proactive and autonomic actions can be made to prevent blockchain from disruptive actions.

There are a number of surveys and reviews on blockchain technologies [1], [2], [3] while few studies focus on the integration of blockchain and AI. Despite recent attempts [4], [5] on the integration of blockchain and AI to enhance 5G communication systems or Internet of Things (IoT), they did not consider using AI technologies to improve blockchain systems themselves. In contrast to most recent studies on the integration of blockchain and AI, this article mainly concentrates on improving blockchain systems by applying AI technologies. This article aims at reviewing challenges of blockchain technologies, discussing opportunities brought by AI to blockchain, and introducing enabling technologies of blockchain intelligence.

Z. Zheng and J. Wu are with School of Data and Computer Science, Sun Yat-sen University, China (email: {zhzibin,wujiajing}@mail.sysu.edu.cn).

H.-N. Dai is with Faculty of Information Technology, Macau University of Science and Technology, Macau (email: hndai@ieee.org).

II. Opportunities of Blockchain Intelligence

This section briefly surveys blockchain challenges and discusses opportunities of blockchain intelligence.

A. Blockchain challenges

As a disruptive software technology, blockchain is reshaping diverse business sectors, such as healthcare [6], IoT [7], smart manufacturing [8], and intelligent transportation [9]. Blockchain is essentially a chain-like data structure storing transactions verified by majority of nodes throughout the whole network. The appearance of blockchain has promoted the development of smart contracts. A smart contract essentially consists of a number of computerized contractual agreements consented by multiple parties [1]. Blockchain and smart contract technologies can simplify the business process, improve the interoperability across different sectors, and reduce the administrative cost.

Although blockchain and blockchain-enabled smart contracts are promising in reshaping various industrial sectors, the intrinsic limitations of blockchain systems also lead to the following challenges.

1) Operational maintenance. Due to the decentralization and heterogeneity of blockchain systems, it is difficult to identify the potential factors affecting the performance of blockchain [10]. For example, the *transaction throughput* bottleneck of Hyperledger Fabric is different from that of Bitcoin and Ethereum since different consensus algorithms are adopted. Moreover, like other software systems, smart contracts consist of a number of computer programs, which may suffer from software bugs, malicious codes, and incompatibility of running environments [11].

2) Quality assurance of smart contracts. Smart contracts suffer from software vulnerabilities such as *re-entrancy* vulnerability, overcharging issue, randomness controlling and Decentralized Autonomous Organization (DAO) attack [1], [2]. In addition, contract correctness is also crucial to smart contracts since it is nearly impossible to make any revisions once they are deployed on top of blockchains. However, like software systems, smart contracts often contain programming bugs which may lead to crashes or misbehaviors while it is challenging to detect and identify these bugs due to the complexity of smart contracts.

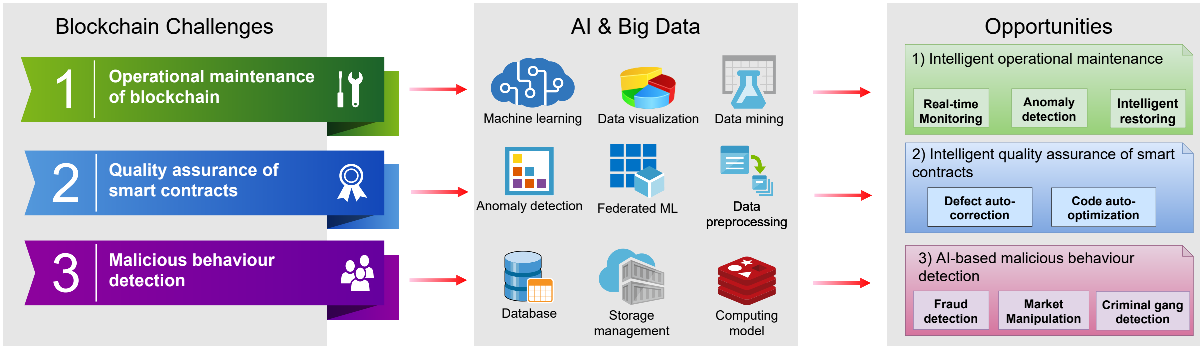

Fig. 1. Opportunities brought by artificial intelligence to address challenges of blockchain

3) Malicious behavior detection. Besides legal businesses, blockchain may be exploited for malicious activities that are nevertheless difficult to be detected due to the *pseudonymity* of blockchains (*i.e.*, anonymous blockchain addresses). On the other hand, the encrypted blockchain data also leads to the difficulty of detecting and identifying malicious behaviors via simply data analytics [12]. Moreover, the massive volume and heterogeneity of blockchain data as well as the diversity of user behaviors make the problem even worse. As a result, conventional classification-based methods (*e.g.*, machine learning methods) cannot be directly applied.

B. Opportunities of Blockchain Intelligence

AI as a broad discipline covering machine learning and cognitive computing is an ability of intelligent agents conducting intellectual tasks. Meanwhile, big data plays a critical role in propelling AI as well as AI applications. The appearance of diverse blockchain systems has generated the enormous volumes of blockchain data. Data analytics on the massive blockchain data can extract huge business values and valuable information. The recent advances in AI have also greatly driven the development of big data analytics on blockchain data. Thus, the integration of AI and blockchain technologies can potentially overcome the aforementioned challenges of blockchain systems, thereby forming intelligent blockchain systems. We name such integration of blockchain with AI as *blockchain intelligence*. It is worth mentioning that there are several research efforts [10], [13] on integrating blockchain with AI while they mainly concentrated on exploiting blockchain for AI so as to overcome security and privacy vulnerabilities of AI. In contrast to these studies, this paper mainly concentrates on solving the intrinsic issues of blockchains by using AI technologies.

Fig. 1 summarizes the opportunities of blockchain intelligence to be illustrated as follows.

1) Intelligent operational maintenance of blockchain. Blockchain generates a huge amount of data in a real-time manner. By analyzing blockchain data, we can detect the possible faults, forecast the failures, and identify the performance bottleneck so as to tune or adjust the performance of blockchain systems. There are four different levels of data analytics including: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. In particular, the descriptive analytics of blockchain log data can help to monitor the real-time performance of blockchain systems and identify the possible faults [14]. In addition to diagnostic analytics on blockchain data, predictive analytics is also necessary to anticipate the performance bottleneck of blockchain systems. Unlike diagnostic and predictive analytics, prescriptive analytics can simulate and optimize blockchain systems so as to improve the reliability of blockchain systems.

2) Intelligent quality assurance of smart contracts. Like computer software, smart contracts may contain bugs or faulty programming codes, which are vulnerable to crashes and malicious attacks. It is crucial to detect and identify bugs in smart contracts so as to achieve the ultimate goal of intelligent quality assurance of smart contracts. Meanwhile, smart contracts are essentially program codes that are sensitive to execution cost (*e.g.*, the execution of smart contracts is charged by gas in Ethereum). Thus, it is a necessity to identify the gas-costly patterns and correct these vulnerable smart contracts. Machine learning methods can be used to detect and recognize vulnerable bugs in smart contracts automatically. Moreover, the growing number of smart contracts also brings opportunities to automate the composition of multiple contracts.

3) Automated malicious behavior detection. The decentralized blockchain systems result in the difficulty in auditing malicious behaviors such as money laundering, phishing, gambling, and scams that occurred in blockchain platforms. Blockchain systems have generated massive transaction data, which are essentially available to everyone, whereas the historical transaction data are pseudonymous through anonymizing account addresses. The massive blockchain data brings opportunities in auditing and detecting malicious behaviors. Big data analytics on massive blockchain data can help to identify malicious users, recognize behavior patterns, analyze market manipulation, and detect scams.

III. Enabling Technologies of Blockchain Intelligence

Big data analytics of blockchain data is beneficial to fraud recognition of transactions [12] and vulnerability detection of smart contracts [11]. Fig. 2 shows a general procedure of big data analytics of blockchain data. This procedure consists of

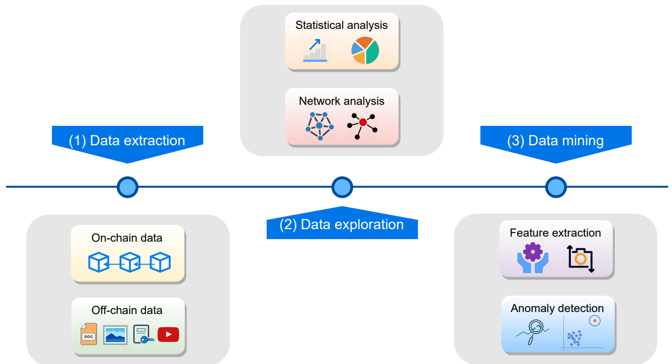

Fig. 2. General procedure of big data analytics of blockchain data.

three consecutive processes: 1) data extraction on both on-chain and off-chain data, 2) data exploration through statistical analysis and network (graph) analysis, and 3) data mining on blockchain data.

However, it is also challenging to conduct big data analytics of blockchain data. 1) It is extremely time-consuming to download the entire blockchain data due to the bulky blockchain size, *e.g.*, it took more than one week and over 500 GB storage space to fully synchronize (*i.e.*, download) the entire Ethereum at a newly-joined peer [15]. 2) It requires substantial efforts in extracting and processing blockchain data. First, blockchain data is stored at clients in heterogeneous and complex data structures, which cannot be directly analyzed. Meanwhile, the underlying blockchain data is either binary or encrypted. Thus, it is a necessity to extract and process binary and encrypted blockchain data so as to obtain valuable information while this process is non-trivial as conventional data analytic methods may not work for this type of data. 3) There is no general data extract tool for blockchain data. Although several open-source tools for blockchain data extraction are available, most of them can only support to extract partial blockchain data (not the entire data).

To address the above challenges, we propose a blockchain data analytics framework namely XBlock-ETH to analyze Ethereum data [15]. In particular, we extract raw data consisting of 8,100,000 blocks of Ethereum. Fig. 3(a) illustrates the typical Ethereum transaction execution flow from Block N to EVM through blockchain peer. During this procedure, we collect the three types of blockchain raw data: Block, Receipt, and Trace. Since the analysis on the raw blockchain data is difficult, we process and categorize the obtained Ethereum Blockchain data into six datasets: *(1) Block and Transaction*, *(2) Internal Ether Transaction*, *(3) Contract Information*, *(4) Contract Calls*, *(5) ERC20 Token Transactions*, *(6) ERC721 Token Transactions* as shown in Fig. 3(b). It is non-trivial to process the raw since it requires substantial efforts in extracting useful information from raw data and associating with six datasets. It is worth mentioning that the complete XBlock-ETH datasets as well as other data exploration tools have been released on the XBlock.pro website (http://xblock.pro/ethereum/) and [15] presents a technical description on those tools and datasets of XBlock-ETH.

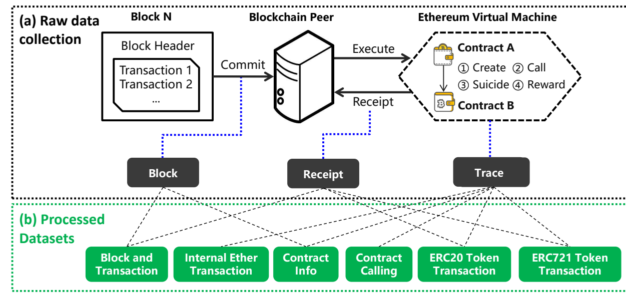

Fig. 3. Data processing during Ethereum transaction flow

IV. Conclusion

This article first reviews the blockchain technologies, analyze the challenges in blockchain systems, discusses opportunities brought by AI to blockchain systems. We name such integration of blockchain and AI as *blockchain intelligence*. We mainly discuss that AI bring benefits to blockchain in aspects of *intelligent operational maintenance of blockchain*, *intelligent quality assurance of smart contracts* and *automated malicious behavior detection*. We then discuss the enabling technologies of blockchain intelligence. We believe that the integration of AI with blockchain technology will further drive the benignant development of blockchain systems.

References

[1] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," *Future Generation Computer Systems*, vol. 107, pp. 841 – 853, 2020.

[2] Z. Zheng, S. Xie, H.-N. Dai, W. Chen, X. Chen, J. Weng, and M. Imran, "An overview on smart contracts: Challenges, advances and platforms," *Future Generation Computer Systems*, vol. 105, pp. 475–491, April 2020.

[3] M. Wu, K. Wang, X. Cai, S. Guo, M. Guo, and C. Rong, "A Comprehensive Survey of Blockchain: From Theory to IoT Applications and Beyond," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8114–8154, 2019.

[4] Y. Dai, D. Xu, S. Maharjan, Z. Chen, Q. He, and Y. Zhang, "Blockchain and Deep Reinforcement Learning Empowered Intelligent 5G Beyond," *IEEE Network*, vol. 33, no. 3, pp. 10–17, 2019.

[5] Y. Wu *et al.*, "Convergence of blockchain and edge computing for secure and scalable iiot critical infrastructures in industry 4.0," *IEEE Internet of Things Journal*, pp. 1–18, 2020.

[6] R. Wang, H. Liu, H. Wang, Q. Yang, and D. Wu, "Distributed security architecture based on blockchain for connected health: Architecture, challenges, and approaches," *IEEE Wireless Communications*, vol. 26, no. 6, pp. 30–36, 2019.

[7] C. Lin, D. He, X. Huang, and K.-K. R. Choo, "Secure and Privacy-preserving Smart Contract-based Solution for Access Control in IoT," *IEEE Blockchain Technical Briefs*, July 2018.

[8] Y. Zhang, X. Xu, A. Liu, Q. Lu, L. Xu, and F. Tao, "Blockchain-based trust mechanism for iot-based smart manufacturing system," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1386–1394, 2019.

[9] K. Xiao, W. Shi, Z. Gao, C. Yao, and X. Qiu, "DAER: A Resource Pre-allocation Algorithm of Edge Computing Server by Using Blockchain in Intelligent Driving," *IEEE Internet of Things Journal*, pp. 1–1, 2020.

[10] T. T. A. Dinh, R. Liu, M. Zhang, G. Chen, B. C. Ooi, and J. Wang, "Untangling blockchain: A data processing view of blockchain systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1366–1385, 2018.

[11] T. Chen, Y. Feng, Z. Li, H. Zhou, X. Luo, X. Li, X. Xiao, J. Chen, and X. Zhang, "Gaschecker: Scalable analysis for discovering gas-inefficient smart contracts," *IEEE Transactions on Emerging Topics in Computing*, pp. 1–14, 2020.

[12] W. Chen, Z. Zheng, J. Cui, E. Ngai, P. Zheng, and Y. Zhou, "Detecting ponzi schemes on ethereum: Towards healthier blockchain technology," in *Proceedings of the 2018 World Wide Web Conference*, 2018, pp. 1409–1418.

[13] M. Nassar, K. Salah, M. H. ur Rehman, and D. Svetinovic, "Blockchain for explainable and trustworthy artificial intelligence," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 1, p. e1340, 2020.

[14] P. Zheng, Z. Zheng, X. Luo, X. Chen, and X. Liu, "A detailed and real-time performance monitoring framework for blockchain systems," in *2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP)*. IEEE, 2018, pp. 134–143.

[15] P. Zheng, Z. Zheng, J. Wu, and H. Dai, "XBlock-ETH: Extracting and Exploring Blockchain Data From Ethereum," *IEEE Open Journal of the Computer Society*, vol. 1, pp. 95–106, 2020.