Hindawi
Security and Communication Networks
Volume 2022, Article ID 4233238, 11 pages
https://doi.org/10.1155/2022/4233238

Research Article

Blockchain Empowered Federated Learning for Distributed Network Security Behaviour Knowledge Base in 6G

Kun Li, Huachun Zhou, Zhe Tu, Feiyang Liu, and Hongke Zhang

School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Huachun Zhou; hchzhou@bjtu.edu.cn

Received 13 January 2022; Accepted 29 March 2022; Published 21 April 2022

Academic Editor: Hao Peng

The malicious flow originating from massive access devices in 6G network will increase sharply. In order to effectively reduce malicious flow, we hope to establish a new framework for coordination of security monitoring and malicious behaviour control in 6G network. Federated learning provides data and privacy protection for the distributed network security behaviour knowledge base. However, since the equipment of its participants needs to upload the original data to the central server for model training, this may lead to data leakage in the knowledge base. Therefore, in this article, we first use the knowledge graph to describe network security behaviours, then build a universal network security malicious behaviour knowledge base, and discuss its application scenarios. Then, we propose a blockchain empowered federated learning (BeFL) for distributed network security malicious behaviour knowledge base architecture to ensure the security of knowledge transmission. Finally, we deployed the designed distributed knowledge base in the prototype system and compared it with the other two baseline methods to verify the performance. Relevant results show that our method outperforms other methods in terms of user identification, flow detection, and attack source tracing.

1. Introduction

With the rapid development of communication technology, artificial intelligence (AI) technology, especially machine learning and big data analysis technology, plays a key role in designing and optimizing 6G architecture, protocols, and operations [1]. Many works discuss potential technologies for 6G to enable mobile AI applications, as well as AI-enabled methodologies for 6G network design and optimization [2, 3]. However, potential security and privacy issues in 6G networks are gradually exposed, mainly in the following two aspects:

(i) Malicious operations caused by massively connected devices [4, 5]. Large-scale device interconnection can facilitate attackers to quickly find protocol vulnerabilities, and then locate weak links and initiate malicious operations. The IoT scenario of massively connected devices puts forward higher requirements on the network infrastructure and preventive mechanism of 6G networks. It is necessary to design a distributed preventive mechanism to support the secure access and identity verification of massive access devices.

(ii) Security and privacy issues caused by AI [6]. AI raises concerns about security and privacy problems, including data security, AI model and algorithm security, vulnerabilities in AI software systems and frameworks, and malicious utilization of AI technologies [7]. Various AI methods require the collection of large amounts of data to train the model, which may contain sensitive user information [8]. Although many researchers make up for the shortcomings of the AI model by optimizing algorithms, there are still some loopholes that can lead to potential data pollution attacks [9]. How to solve the attack vulnerabilities of the AI model itself to better serve network applications should also be a problem we consider.

As a classic solution for security control and privacy exposure, the distributed network security knowledge base can complete the knowledge accumulation between parties and the scheduling of malicious behaviour monitoring

methods in a decentralized network scenario [10]. This distributed solution does not need to disclose the private information of each domain to prevent network attacks and changes the traditional passive defense to active interception to block malicious user behaviour. The knowledge graph models the real network through a graph composed of nodes and relationships [11]. It can efficiently and intuitively analyse and process complex relationships in network security data and extract knowledge of network security behaviours, so that user identities and behaviours can be verified and controlled. It is a good foundation for building a large-scale distributed behaviour knowledge base. In order to further meet the needs of user privacy and network security, we use federated learning that can help multiple parties build a shared high-performance model to solve the problem of data islands [12]. The model combining federated learning and distributed knowledge base can well guarantee the privacy of local data and effectively deal with the malicious operation problems caused by massively access devices.

But as mentioned in our second question, the AI method represented by federated learning is dedicated to solving data privacy and security issues. However, when its central server or participants are maliciously manipulated, the aggregation process of the global model will be greatly affected, and even lead to misjudgement of security in the entire network environment [13, 14]. Therefore, we use blockchain [15] to empower federated learning and use its anonymity and credibility mechanism to build a safe and trusted third-party environment for each distributed network security behaviour knowledge base participating in federated learning. This method can guarantee the security of the knowledge of the malicious behaviour knowledge base of the distributed network security.

In summary, in order to deal with the security and privacy challenges in 6G networks, we use blockchain empowered federated learning to build a distributed network security behaviour knowledge base framework. Our contribution is clearly documented as follows:

(i) We use knowledge graphs to describe network security behaviours and build a universal network security malicious behaviour knowledge base. After that we discussed its application scenarios.

(ii) We discussed the technical issues of the combination of blockchain and federated learning and proposed a distributed network security malicious behaviour knowledge base architecture that empowers federated learning with blockchain.

(iii) Finally, we deployed the prototype system and compared the above architecture with the other two baseline methods to verify performance.

The rest of this paper is organized as follows. The second section reviews related work. The third section defines the network security behaviour and describes the construction process of the network security behaviour knowledge base. The fourth section proposes design process of blockchain empowered federated learning for distributed network security behaviour knowledge base. In the fifth part, we deployed the prototype and conducted experiments to verify the performance of the proposed distributed knowledge base. Finally, the sixth section summarizes the paper.

2. Related Work

Since the concept of federated learning (FL) was proposed, it has been applied by many researchers in the field of network security to provide privacy protection [12, 13, 16–18]. Mothukuri et al. [13] provided a comprehensive study concerning FL's security and privacy aspects that can help bridge the gap between the current state of federated AI and a future in which mass adoption is possible. They believed that the most specific security threats currently are communication bottlenecks, poisoning, and backdoor attacks,-while inference-based attacks are the most critical to the privacy of FL. Lu et al. [16] presented a new privacy-preserving federated learning mechanism and designed a two-phase mitigating scheme consisting of intelligent data transformation and collaborative data leakage detection. Some researchers [17, 18] also used FL for local storage and maintenance on different devices in 5G network, while providing privacy protection for participants.

However, the above work lacks security considerations in the process of federated learning data exchange. We hope to construct a distributed knowledge base based on the distributed and privacy protection characteristics of federated learning. But once the knowledge provided by the participants is maliciously modified, it will have a huge impact on the performance of the global model.

Blockchain, as a distributed ledger technology [19–21], provided a decentralized solution that was a feasible measure to protect the security of federal learning data, thereby helping FL get rid of the central server and improve security. There had been some work dedicated to the combination of blockchain and FL to deal with the security risks of FL [22–28]. We summarized in Table 1 a review of related work that combines blockchain with federated learning and analysed whether they meet the needs of 6G network security.

However, in the face of the rapidly increasing number of device connections in the 6G network, it is difficult to effectively analyse changeable malicious behaviours only by relying on the distributed framework of blockchain and FL [29]. The biggest difference from the traditional blockchain-enabled FL is that we introduce a knowledge base to form a closed-loop feedback for the whole process, so that the entire system can accumulate knowledge and iteratively update it autonomously. Therefore, we use blockchain empowered FL to build a distributed network security behaviour knowledge base framework to meet the new security requirements of 6G networks. In this paper, we addressed the aforementioned research gaps in the literature.

3. Network Security Behaviour Knowledge Base

In this section, we define the network security behaviour and analyse the features it contains. After that, we describe the construction process of the network security behaviour knowledge base.

Table 1: Comparison with related works.

Ref.	Year	Security requirements of 6G network					
		Access authentication	Access control	Privacy protection	Trusted device	Safe life-cycle	Loop feedback
[22]	2020	×	×	√	√	×	×
[23]	2020	×	×	√	√	×	×
[24]	2021	×	×	√	×	×	×
[25]	2021	√	×	√	×	×	×
[26]	2021	×	√	√	√	×	×
[27]	2021	×	×	√	×	×	×
[28]	2021	×	×	×	√	√	×
Ours	2022	√	√	√	√	√	√

3.1. Network Security Behaviour Description. The attack surface of cyberspace continues to expand under the environment of massive access devices on 6G networks, and malicious attacks continue to show scale and organization. Existing intrusion detection schemes based on statistical methods or machine learning have problems such as complex security knowledge structure, difficulty in feature extraction, and lack of adaptive adjustment detection models. Therefore, it is difficult to adapt to complex and massive access scenarios. In view of the above problems, it is urgent to build a universal and complete network security knowledge base to store and analyse complex network attack knowledge.

Behaviour is the basic characteristic shown by different individuals or groups under certain conditions, or the active response to the stimulus of internal and external environmental factors. The essence of the network is the communication between users. Therefore, we believe that network behaviour is the effective connection between user behaviour and communication behaviour. Then, the problem of network security is that user behaviour or communication behaviour deviates from the benchmark. The purpose of building a network security behaviour knowledge base is to accumulate data of various structures and huge scales and transform them into knowledge that the network can understand and use.

We define network behaviour (NB) as

$$NB \triangleq (\text{User behavior}, \text{Flowbehavior}, \text{Attacker behavior}). \tag{1}$$

The trigger condition of malicious behaviour is the user behaviour because users are usually the initiators of malicious behaviours, and the sources of monitoring user behaviour data mainly include system logs, browsing records, packet load, and operating habits. Malicious behaviours are manifested as flow behaviours in the network because flow is usually the executor of malicious behaviours. Among them, the flow behaviour characterizes the change of the data amplitude of the data stream in real-time communication and reflects the characteristics of flooding attacks for the purpose of denial of service, such as DDoS attacks. The main sources of data that can monitor flow behaviour are packet capture and flow analysis. By analysing user behaviour and traffic behaviour, attackers and their behaviours can be judged. Therefore, the attacker behaviour can also be defined as

$$\text{Attacker behavior} \triangleq \sigma\begin{pmatrix}\text{User behavior}\\ \text{Flow behavior}\end{pmatrix}, \tag{2}$$

where σ is the selected knowledge reasoning method.

Based on the above data sources, we extract the data into features that the network can understand. User behaviour includes Request information (RI), URL information (UI), Browser information (BI), Source port (SP), Destination port (DP), Source IP (SI), Destination IP (DI), and Transmission protocol (TP). Flow behaviour includes characteristics such as Flow duration (FD), Flow speed (FS), Flow number (FN), Flow type (FT), and Flow flag (FF). Attacker behaviour includes all the features mentioned above. Therefore, the network security behaviour (NSB) is finally expressed as

$$NSB \triangleq \begin{cases} RI, UI, BI, SP, DP, SI, DI, TP, etal, \\ FD, FS, FN, FT, FF\ etal, \\ \text{AttackerFeature}. \end{cases} \tag{3}$$

3.2. Knowledge Base Construction. Figure 1 shows the construction process of network security behaviour knowledge base. The knowledge base is based on multi-source heterogeneous data to accumulate knowledge of user behaviour, flow behaviour, and attacker behaviour in the form of a knowledge graph and uses AI technology such as federated learning or blockchain to make comprehensive judgments and real-time feedback. The knowledge base provides reliable knowledge basis for various offensive and defensive scenarios, thereby turning passive defense into active interception, and curbing the occurrence of malicious behaviours from the source.

Figure 2 shows the storage structure of the network security behaviour knowledge base based on the knowledge graph, including the global network behaviour graph, user behaviour graph, flow behaviour graph, and attacker behaviour graph. The knowledge base generates a global network behaviour graph based on environmental knowledge, including the relationship between user entities and flow entities. Suppose the set of user entities $U = \{u_1, u_2, \cdots, u_n\}$ and the set of flow entities $F = \{f_1, f_2, \cdots, f_{m,}\}$, n and m are the number of users and flow in the graph.

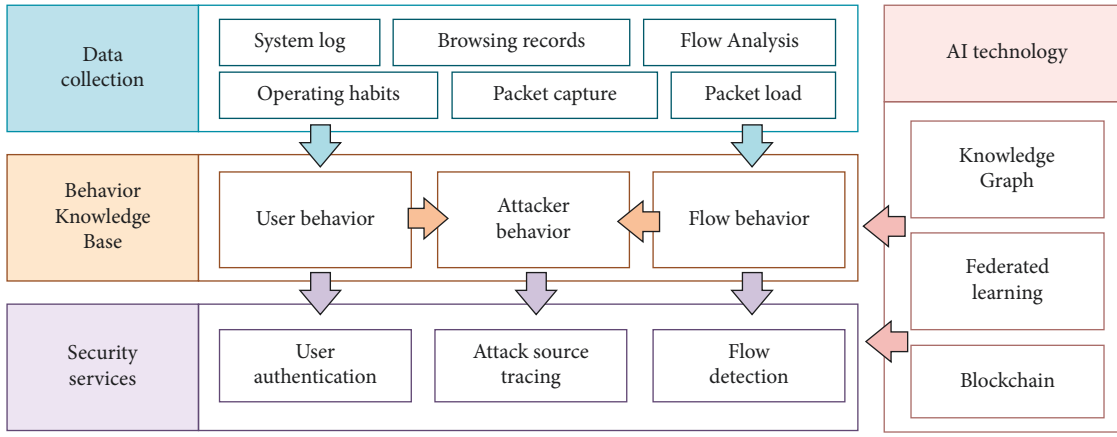

Figure 1: Construction process of network security behaviour knowledge base.

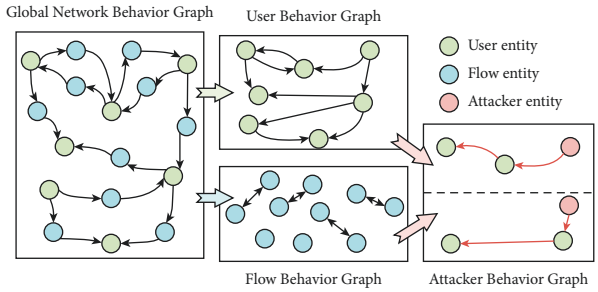

Figure 2: Storage structure of network security behaviour knowledge base based on knowledge graph.

$$f: u_i \longrightarrow f_k \longrightarrow u_j \overset{\theta}{\Rightarrow} u_i \longrightarrow u_j, \tag{4}$$

$$\begin{aligned} f:&\ u_i \longrightarrow f_k, \\ &u_i \longrightarrow f_q \overset{\theta}{\Rightarrow} f_k \leftrightarrow f_q. \end{aligned} \tag{5}$$

We extract the relationship between users in U according to the mapping rule of formula (4), extract the relationship between flow in F according to formula (5), and finally form user behaviour graph and flow behaviour graph. Among them, θ represents that the relationship between the entities on the left needs to be mapped to the relationship between the entities on the right. We obtain user behaviour features and flow behaviour features, combined with AI technology to establish a detection model to infer malicious users or attack flow, and then form attacker behaviour graph to help security research such as attack source tracing.

3.3. Application Scenario. Based on the characteristics of the network security behaviour knowledge base and recent AI security technology research, we introduced the application scenarios of the network security malicious behaviour knowledge base. We cover the three key processes of network security protection: the network entrance, the communication process, and the security issues of reaching the equipment and provide corresponding security services: user authentication, flow detection, and attack source tracing. The application issues of the knowledge base are discussed separately.

3.3.1. Abnormal User Access Behaviour Authentication. Focusing on user access behaviours at the entrance of the network, we can build user access behaviour profiles based on user behaviour in the knowledge base, and then use technologies such as graph convolutional networks to characterize user normal behaviour baselines. When the hacker's behaviour deviates from the baseline of normal behaviour, a judgment is made and an early warning is made, and abnormal behaviour is discovered in time from the source.

3.3.2. Malicious Attack Flow Detection. The network malicious behaviour detection map is established based on the flow characteristics in the knowledge base to guide the construction and rule design of the intrusion detection system. The purpose is to analyse and block malicious flow during the communication process. Based on knowledge reasoning to optimize the flow characteristics in the knowledge base, we can extract more discriminative flow characteristics and improve the detection efficiency and capability.

3.3.3. Alarm Triage and Attack Source Tracing. As the depth and breadth of device detection capabilities increase, the scale of alerts for passive access to the knowledge base continues to expand. The accumulation of a large number of continuous, homogeneous, and low-information warning information prevents the knowledge base from finding truly threatening malicious behaviours. Based on reinforcement learning, we can design automated alarm classification technology to evaluate and classify the alarm information of arriving equipment, thereby alleviating the dependence on the experience of offensive and defensive experts. Eventually, it can improve the accuracy of attack source tracing and improve the input-output ratio of the network security behaviour knowledge base.

3.4. Blockchain Empowered Federated Learning for Distributed Knowledge Base. With the widespread deployment of smart devices and the increasingly powerful

computing capabilities of terminals, the use of distributed deployment of knowledge base design solutions in terminals can effectively utilize low-latency and high-bandwidth communication technologies to form intelligent knowledge base groups. Distributed solutions can well solve the problems of single point of failure, poor scalability, and high development costs in single-machine deployment. Federated learning can make full use of the data of each distributed node for model training on the basis of ensuring the privacy of local data, so it provides a solution for the knowledge privacy of the distributed knowledge base.

However, the current federated learning technology generally believes that all participants are honest, and therefore trust all data providers unconditionally. However, once the attacker gains control of the aggregation server or part of the participants without being identified, she/he can manipulate the local model parameters on the compromised client device during the learning process, resulting in global model update errors [14]. The starting point of the distributed knowledge base is to provide secure data support for the massive access devices in the 6G network, so a high degree of security is required.

With its unique decentralization and unalterable characteristics, the blockchain has shown great potential in the security of federated learning. Specifically, the decentralization of the blockchain can well solve the problem of relying on the aggregation server in the federated learning process, and its immutability allows each participant to upload the model to the chain to receive the evaluation of other participants.

This article introduces a general blockchain empowered federated learning method for the construction of a distributed network security behaviour knowledge base, as shown in Figure 3. We summarized the process of blockchain empowered federated learning through the following steps:

(i) Request distributed knowledge. A group of terminal devices accumulate user data and flow data to maintain their own network security behaviour knowledge base. Any participant can initiate a federated learning request to obtain transaction data of all parties and perform mining tasks when requested to record the requested information on the chain.

(ii) Upload global model. The device that initiates the training request also serves as an aggregation server to replace the original concept of a central server. Each participant uploads the local training weights to the requesting device, which collects the gradients of the participants and updates the model uniformly, thereby using all the knowledge of network security behaviours to train the global security defense model.

(iii) Mining to reach consensus. The device exports and stores the trained global model in the block. In this way, all participants conduct mining to verify the newly created block and reach a consensus (for example, Proof of Work).

(iv) Security evaluation. The participant device that meets the security evaluation qualification downloads the updated global model and uses local network security knowledge to verify the performance of the model. Obviously, the global model is trained by aggregating the network security knowledge provided by all participating devices, so it should show good recognition accuracy when applied to each local network security knowledge. If the test accuracy exceeds the threshold T_{acc}, it is considered that the initiator of the training is not controlled by the attacker. Once it is discovered that the trainer may be manipulated by the attacker, a device with security assessment qualifications will create a block to store the abnormal device information and broadcast it. Other devices will disconnect from the attacked host and seek help from a third-party cleaning device to deal with the attack. We introduce the security factor $\mathrm{Sec}-F$ to determine whether the equipment is qualified for safety evaluation every day, namely:

$$\mathrm{Sec}-F = U_{\mathrm{sec}} \times F_{\mathrm{sec}}\ Sec-F, \quad U_{\mathrm{sec}}, F_{\mathrm{sec}} \text{ is either 0 or 1.} \tag{6}$$

Among them, U_{sec} is the user security factor and F_{sec} is the flow security factor. If the device does not detect abnormal user behaviour based on the local knowledge base and anomaly detection model within a day, U_{sec} is 1, otherwise it is 0. Similarly, F_{sec} is determined by whether abnormal flow behaviour is detected within a day. Only when $\mathrm{Sec}-F = 1$, we consider the device to be trustworthy and give it the qualification for security assessment of the day.

Relying on the decentralization and security of the blockchain, we propose the above-mentioned distributed framework to realize an intelligent security assurance system that combines a security-guaranteed distributed knowledge base and federated learning. Combined with the local network security behaviour knowledge base, it can realize the change from passive defense to active interception for malicious network behaviours, thereby forming a closed-loop feedback strategy in distributed scenarios.

4. Prototype Deployment and Experiment

In this section, based on the proposed distributed network security behaviour knowledge base framework, we deployed a prototype system based on the framework in a real environment and conducted experiments to verify the performance of the proposed distributed knowledge base.

4.1. Prototype System Construction. Prototype deployment: as shown in Figure 4, we implement the prototype in the server. We used DELL PowerEdge R720 rack-mount servers with Intel Xeon E5-2609 CPU, 32G memory, and 1 TB hard-disk storage. We have emulated Ethernet scenarios, mobile access scenarios, IoT scenarios, and satellite networks as the

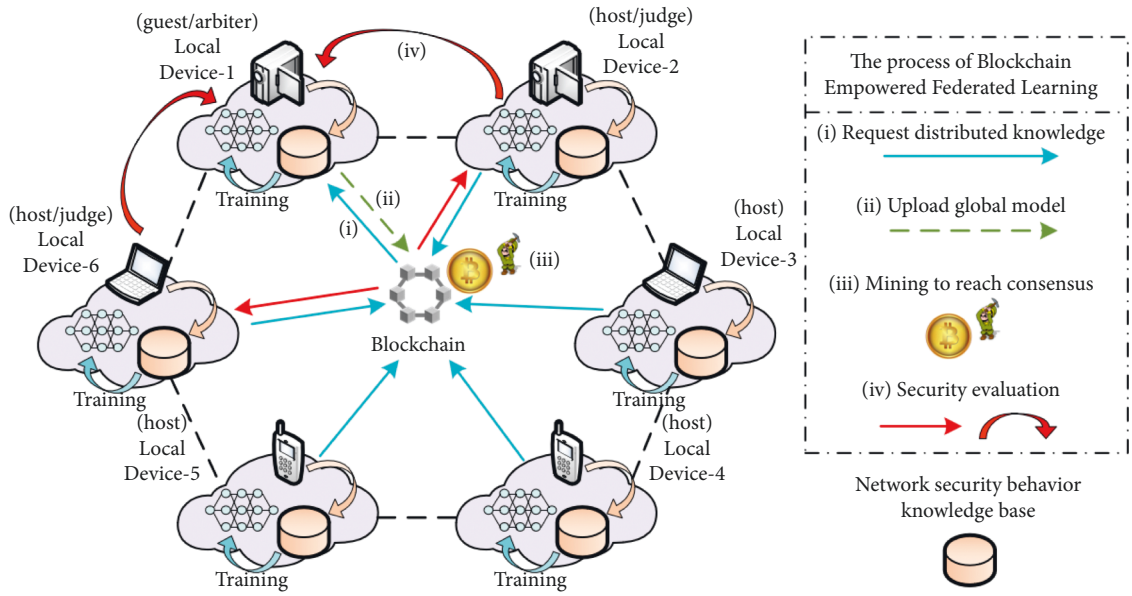

Figure 3: Design process of blockchain empowered federated learning for distributed network security behaviour knowledge base.

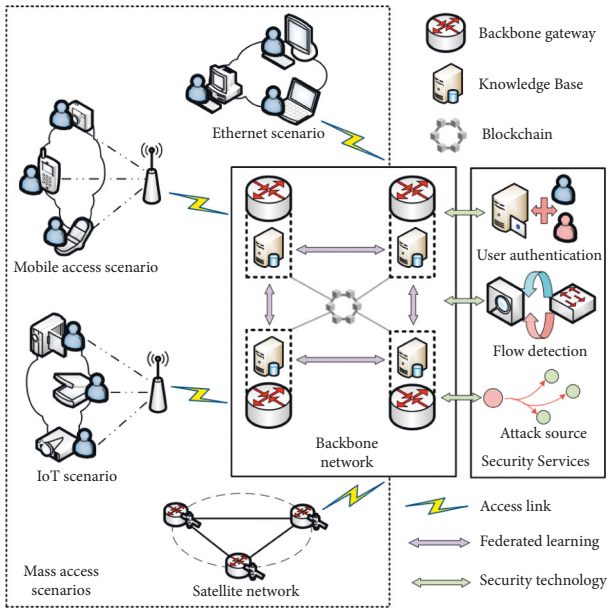

Figure 4: Prototype deployment.

6G mass access device scenarios. Then, we deployed four high-performance access gateways in the backbone network and installed Neo4j [30], FATE [31], and Ethereum [32] in each gateway to implement the blockchain empowered Federated Learning for distributed network security behaviour knowledge base. Combined with AI technology, the framework we proposed can provide multiple security services such as user authentication, flow detection, and attack source tracing.

Datasets, models, and parameter settings: in order to prove the effectiveness of the proposed framework, we generated and captured the most destructive and influential DDoS attack data set in the prototype system. The data set contains 22 common DDoS attacks and normal flow. Among them, the scale of malicious flow data exceeds 1.5 million, and the normal flow data are about 900,000. We assign 5 to 6 different types of DDoS attacks to each gateway as their respective local data sets and use 80% of them as the training set, and 20% of the data sets containing all DDoS attack types as the testing set. For the knowledge base, we use the Neo4j graph database to store user behaviour and flow behaviour in the prototype system. For the federated learning, we use the open-source framework FATE to implement a complete federated learning process among the four access gateways. The learning model uses SecureBoost, the purpose of which is that all participants jointly learn a shared gradient boosting tree model through a privacy protection protocol. Learning rate is set to 0.01, and the max_depth is set to 7. For the blockchain, we also use Ethereum in the four access gateways, and the consensus mechanism uses the default Proof of Work (PoW).

Performance parameters: in order to evaluate the performance of the proposed framework, we introduce four evaluation indicators: accuracy, loss function, packet loss, and number of attack sources tracing. Among them, accuracy is the proportion of users or traffic types that are correctly predicted to the total sample, and the value range is [0,1]. The larger the value, the better the predictive ability of the model. The loss function is used to estimate the degree of inconsistency between the predicted value of the model and the true value. The smaller the loss function, the better the robustness of the model. The packet loss rate refers to the ratio of the number of data packets lost in the test to the transmitted data group. The smaller the packet loss rate, the more complete the data transmission between distributed devices. The number of attack sources tracing is used to evaluate the number of attack source nodes captured during the attack source tracing process. The larger the number, the better the attack source tracing effect of the model.

4.2. Performance Evaluation. We show in Figure 5 the distributed network security behaviour knowledge base deployed in the access gateway based on Neo4j. Each access gateway only saves access to its own user or flow knowledge, referring to the structure and mapping relationship in Figure 2 to form the global network behaviour graph, user behaviour graph, and flow behaviour graph, respectively. The global network behaviour graph covers 15 users and 24 flow between access users and maintains a total of 48 directed relationships. According to the mapping relationship of formulas (4) and (5), we then obtain the user behaviour graph and flow behaviour graph. Next, we will conduct experiments to verify the security services provided by our proposed framework such as user authentication, flow detection, and attack source tracing.

We propose two baseline methods and the proposed BeFL method for comparative experiments in our prototype of the distributed cybersecurity behaviour knowledge base:

Standard federated learning (FL) [12]: use the default settings and aggregation algorithm.

Federated learning based on DDoS Open Threat Signal [33] (DOTS-FL): DOTS is a protocol for responding to coordination and mitigation of DDoS attacks, which can ensure that data can still be transmitted through data channel when subjected to DDoS attacks. Therefore, we use DOTS data channel transmission to replace FL original socket communication method to improve data security.

We run the above three methods in four gateways and use three typical DDoS attacks, network layer DDoS, distributed reflection DoS (DRDoS), and low-rate DDoS, on one of the access gateways, to compare the performance of the above three methods on user authentication, flow detection, and attack source tracing.

4.2.1. User Authentication. Assuming the worst-case scenario, we believe that an access gateway has been controlled by an attacker and is trying to contaminate the global model with a completely wrong dataset. We use three distributed training methods to compare the performance of the tainted global model in user authentication. Figure 6 show the training accuracy and loss function curves of the three methods for user authentication within 30 s. It can be seen that due to the misleading data set provided by the malicious user, the model aggregation effect of the two baseline methods is greatly affected, and the training accuracy is less than 50%. The BeFL method we proposed ensures the credibility of access users or devices by evaluating security assessment qualifications. It also provides closed-loop feedback for the entire system, ensuring that the user authentication accuracy of the aggregation model is around 90%.

4.2.2. Flow Detection. Figure 7 shows the detection accuracy of three methods for different real-time DDoS attacks and testing set. In the face of each type of attack, the three methods showed far lower detection effects on the test set. This is because when the gateway is attacked by DDoS, its resource occupancy rate is close to 100%, which affects the execution of the model. Compared with the other two baseline methods, our proposed BeFL broadcasts the attacked information through the blockchain to obtain timely feedback of the detection model. It shows a higher detection accuracy in various attack scenarios, which are 86.22%, 75.27%, and 64.36%, respectively. DOTS-FL opens data channels to forward requests to cleaning equipment in response to flooding attacks. Therefore, it performs well under the first two DDoS, but it cannot effectively deal with low-rate DDoS attacks. The standard FL does not have protection measures for data forwarding, and it is difficult to perform effective detection in the case of high resource occupancy, and therefore exhibits the worst performance.

In order to verify whether the shared weights provided by each distributed device to aggregate the global model of federated learning are lost due to high resource consumption, we monitor and capture packets on a random data exchange interface used for federated learning that accesses the gateway. We verified the impact of packet loss in the communication process on the accuracy of attack detection. Figure 8 shows the packet loss of three methods under different DDoS attacks. Although FL reached a packet loss close to 90% as we imagined, the packet loss of the proposed BeFL is always higher than that of DOTS-FL under flooding attacks. This is because although packet loss occurs in the weight transmission process, the blockchain-based scheme requires each participant to re-mine until the shared parameters are obtained. Therefore, a higher packet loss does not mean that the gateway has not received the weight parameters. In addition, BeFL shows a low packet loss under each attack, which verifies the stability of our proposed method. In general, the packet loss of each method under attack is inversely proportional to the detection accuracy.

4.2.3. Attack Source Tracing. We limit the number of IP addresses that initiate DDoS attacks per minute within the range of 1 to 10 and continuously launch random attack tests for 10 minutes to verify the performance of the above three methods in attack source tracing. Figure 9 shows the number of attack sources correctly captured by the three methods in each time period. Thanks to the security evaluation qualification mechanism we put forward in Section 4, BeFL correctly captures more attack source nodes in each time period. DOTS-FL and FL lacked credible authentication of the equipment, which led to a certain degree of misjudgement of the attack source.

Finally, Figure 10 shows the captured attacker and malicious traffic based on the user authentication scenario and the flow detection scenario at the first minute. Compared with Figure 5, we record malicious entities in the user behaviour graph and the popular behaviour graph, respectively. Referring to the first minute of Figure 9, we completely capture 6 DDoS attack sources and 7 groups of malicious flows from them based on BeFL and save their attack paths in the attacker behaviour graph. Further, we extract 14 relationships between attackers and malicious flow behaviours based on formulas (4) and (5) at the first minute. The attack source tracing accuracy rate of the

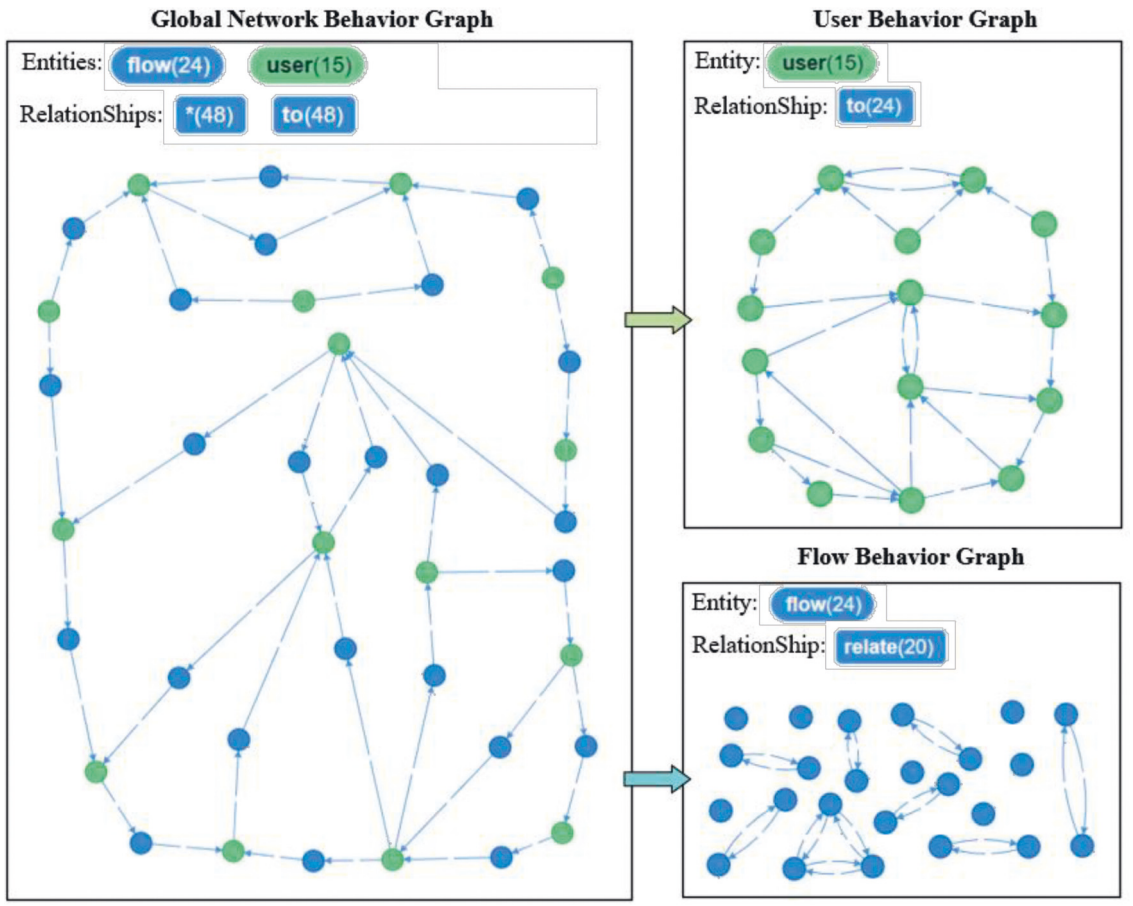

Figure 5: Distributed network security behaviour knowledge base in one of the gateways.

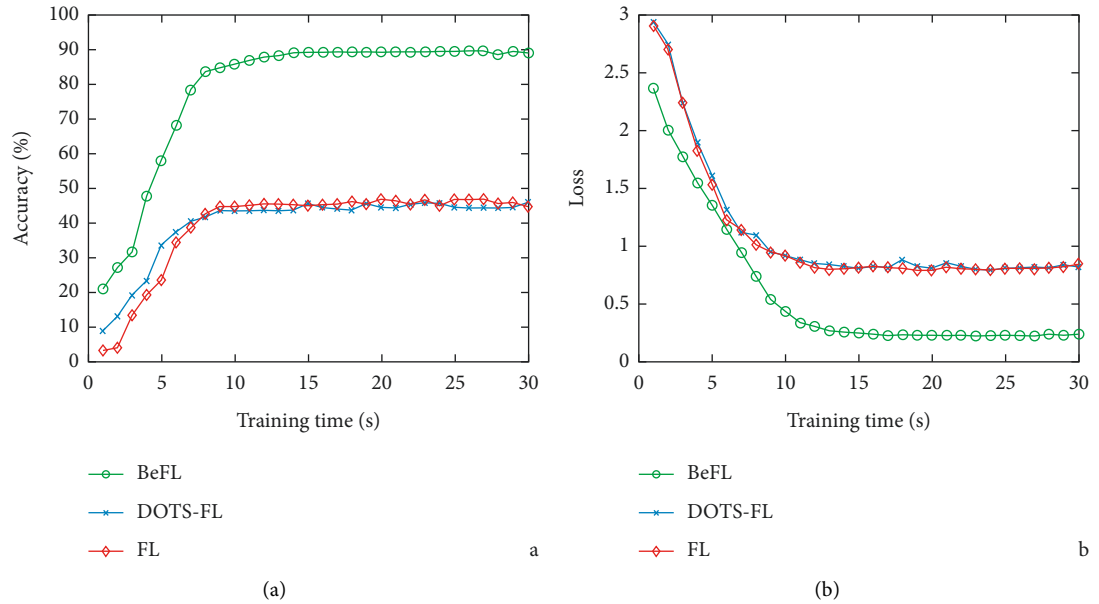

Figure 6: (a) Accuracy changes with training time. (b) Loss changes with training time.

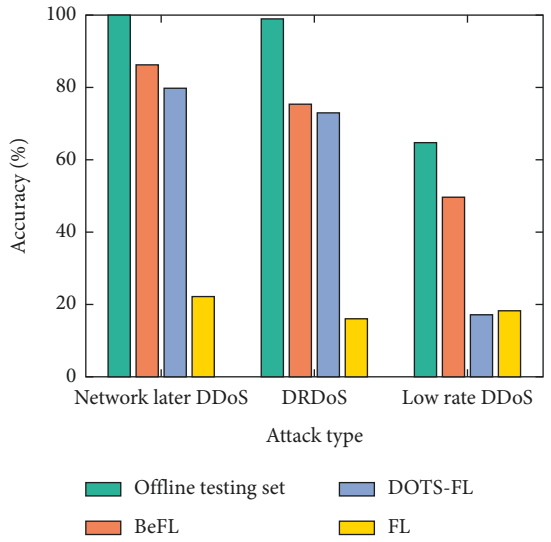

Figure 7: Detection accuracy of three methods and testing set.

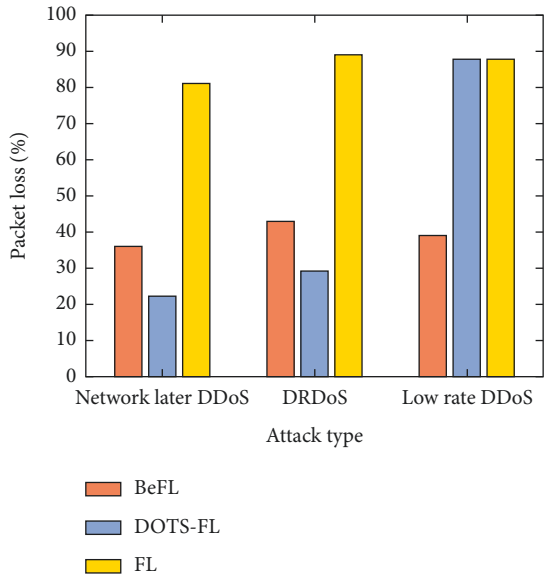

Figure 8: Packet loss of three methods.

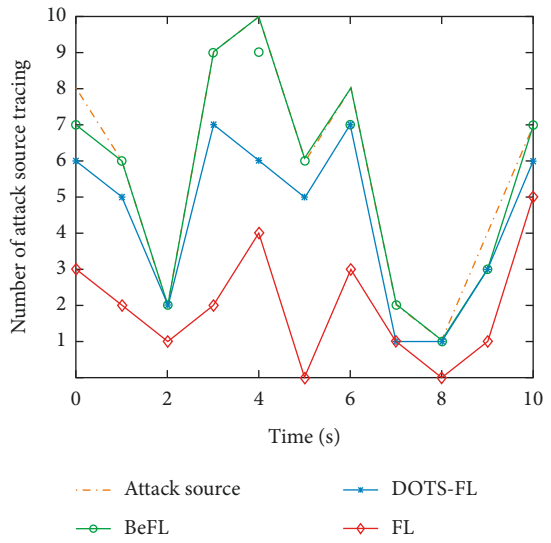

Figure 9: Number of attack source tracing with time.

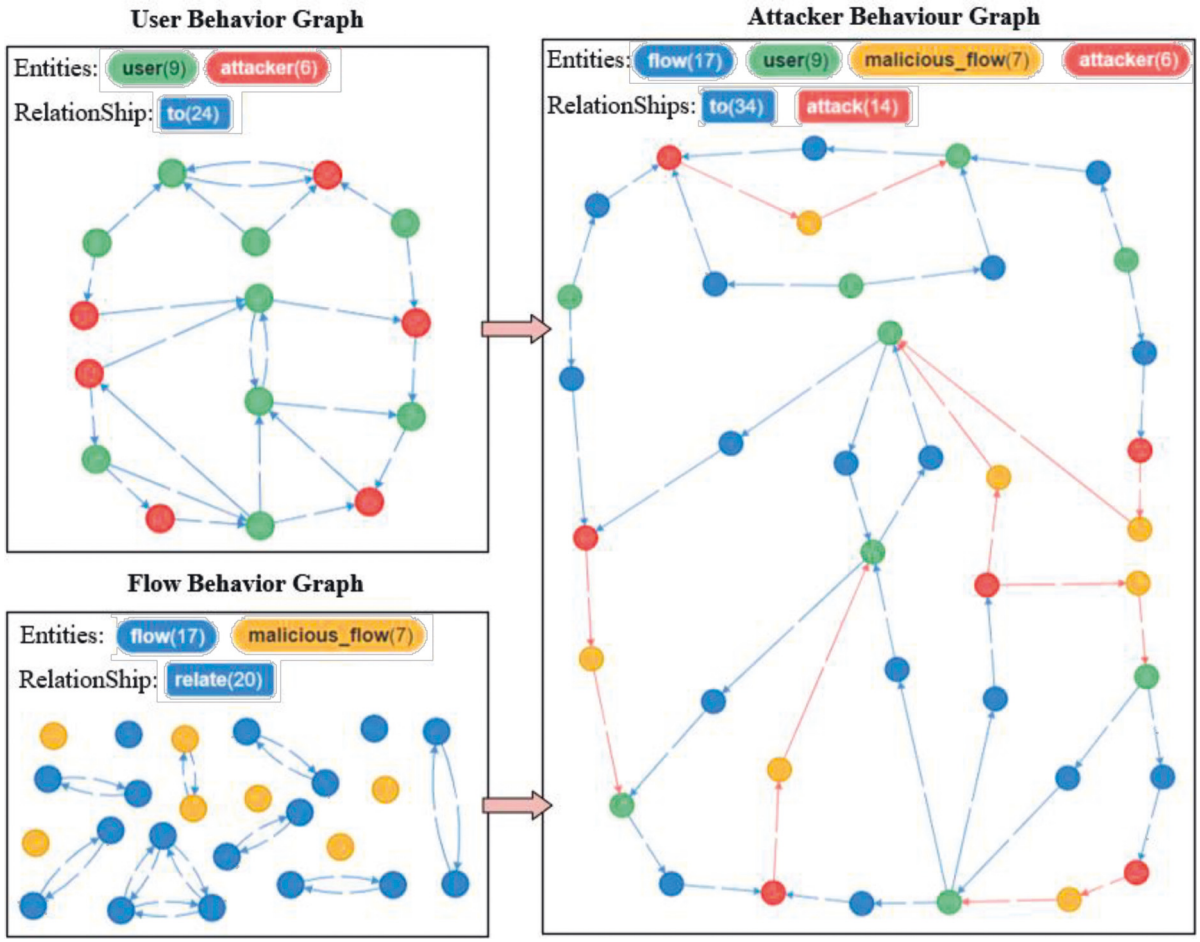

Figure 10: Updated distributed network security behaviour knowledge base at the first minute.

attacker behaviour graph reaches 100%. Therefore, the network can effectively use the knowledge of security behaviour to guide network security services and continuously improve the security capability of the network.

5. Conclusions

We use blockchain-enabled federated learning to build a distributed network security behaviour knowledge base, aiming to establish a new framework for collaboration between security monitoring and malicious behaviour control in 6G networks. We deploy the above framework in the prototype and used different DDoS attacks to verify the security and stability of the framework. The results prove that compared with the other two baseline methods, our proposed framework has the highest accuracy of attack recognition and the best survivability. In the next step, we will conduct more in-depth and specific research on the security issues faced in 6G networks such as user identification, flow detection, and attack source tracing.

Data Availability

Part of the data set the authors used can be found at https://github.com/kun9717/STIN-data-set/.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This paper was supported by the National Key R&D Program of China under Grant no. 2018YFA0701604 and in part by the Fundamental Research Funds for the Central Universities under Grant nos. 2021YJS008 and 2021YJS012.

References

[1] K. B. Letaief, W. Chen, and Y. Shi, "The roadmap to 6G: AI empowered wireless networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[2] H. Yang, A. Alphones, and Z. Xiong, "Artificial-intelligence-enabled intelligent 6G networks," *IEEE Network*, vol. 34, no. 6, pp. 272–280, 2020.

[3] F. Tang, Y. Kawamoto, and N. Kato, "Future intelligent and secure vehicular network toward 6G: machine-learning approaches," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 292–307, 2020.

[4] M. Wang, T. Zhu, and T. Zhang, "Security and privacy in 6G networks: new areas and new challenges," *Digital Communications and Networks*, vol. 6, no. 3, pp. 281–291, 2020.

[5] G. Gui, M. Liu, and F. Tang, "6G: opening new horizons for integration of comfort, security, and intelligence," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 126–132, 2020.
[6] M. Ylianttila, R. Kantola, and A. Gurtov, "6g white paper: research challenges for trust, security and privacy," 2020, https://arxiv.org/abs/2004.11665.
[7] X. You, C. Wang, and J. Huang, "Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences*, vol. 64, no. 1, Article ID 110301, 2021.
[8] K. Zhang, X. Song, and C. Zhang, "Challenges and future directions of secure federated learning: a survey," *Frontiers of Computer Science*, vol. 16, no. 5, pp. 1–8, 2022.
[9] J. Zhang, B. Chen, and X. Cheng, "Poisongan: generative poisoning attacks against federated learning in edge computing systems," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3310–3322, 2021.
[10] H. Yang, K. Zhan, and M. Kadoch, "BLCS: brain-like distributed control security in cyber physical systems," *IEEE Network*, vol. 34, no. 3, pp. 8–15, 2020.
[11] Q. Guo, F. Zhuang, and C. Qin, "A survey on knowledge graph-based recommender systems," *IEEE Transactions on Knowledge and Data Engineering*, 2020.
[12] B. McMahan and E. Moore, "Communication-efficient learning of deep networks from decentralized data," *Artificial intelligence and statistics*, PMLR, vol. 54 , pp. 1273–1282, 2017.
[13] V. Mothukuri, R. M. Parizi, and S. Pouriyeh, "A survey on security and privacy of federated learning," *Future Generation Computer Systems*, vol. 115, pp. 619–640, 2021.
[14] M. Aledhari, R. Razzak, and R. M. Parizi, "Federated learning: a survey on enabling technologies, protocols, and applications," *IEEE Access*, vol. 8, Article ID 140699, 2020.
[15] U. Bodkhe, S. Tanwar, and K. Parekh, "Blockchain for industry 4.0: a comprehensive review," *IEEE Access*, vol. 8, Article ID 79764, 2020.
[16] Y. Lu, X. Huang, and Y. Dai, "Federated learning for data privacy preservation in vehicular cyber-physical systems," *IEEE Network*, vol. 34, no. 3, pp. 50–56, 2020.
[17] Y. Liu, J. Peng, and J. Kang, "A secure federated learning framework for 5G networks," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 24–31, 2020.
[18] L. U. Khan, S. R. Pandey, and N. H. Tran, "Federated learning for edge networks: resource optimization and incentive mechanism," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 88–93, 2020.
[19] X. Li, P. Jiang, and T. Chen, "A survey on the security of blockchain systems," *Future Generation Computer Systems*, vol. 107, pp. 841–853, 2020.
[20] K. Gai, J. Guo, and L. Zhu, "Blockchain meets cloud computing: a survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 2009–2030, 2020.
[21] M. R. Nosouhi, S. Yu, and W. Zhou, "Blockchain for secure location verification," *Journal of Parallel and Distributed Computing*, vol. 136, pp. 40–51, 2020.
[22] M. A. Rahman, M. S. Hossain, and M. S. Islam, "Secure and provenance enhanced internet of health things framework: a blockchain managed federated learning approach," *IEEE Access*, vol. 8, Article ID 205071, 2020.
[23] Y. Lu, X. Huang, K. Zhang, and S. Maharjan, "Blockchain empowered asynchronous federated learning for secure data sharing in internet of vehicles," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4298–4311, 2020.
[24] D. C. Nguyen, M. Ding, and Q. Pham, "Federated learning meets blockchain in edge computing: opportunities and challenges," *IEEE Internet of Things Journal*, vol. 8, no. 16, Article ID 12806, 2021.
[25] Y. Qi, M. S. Hossain, and J. Nie, "Privacy-preserving blockchain-based federated learning for traffic flow prediction," *Future Generation Computer Systems*, vol. 117, pp. 328–337, 2021.
[26] Y. Li, C. Chen, and N. Liu, "A blockchain-based decentralized federated learning framework with committee consensus," *IEEE Network*, vol. 35, no. 1, pp. 234–241, 2021.
[27] Y. Lu, X. Huang, and K. Zhang, "Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 5098–5107, 2021.
[28] W. Zhang, Q. Lu, and Q. Yu, "Blockchain-based federated learning for device failure detection in industrial IoT," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5926–5937, 2021.
[29] Y. Qu and L. Gao, "Decentralized privacy using blockchain-enabled federated learning in fog computing," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5171–5183, 2020.
[30] Neo4j, "Neo4j," 2021, https://www.snort.org/.
[31] Fate, "Fate," 2021, https://www.fedai.org/.
[32] Ethereum, "Ethereum," 2021, https://ethereum.org/.
[33] M. Boucadair and T. Reddy, "Distributed denial-of-service open threat signaling (DOTS) data channel specification," *RFC*, vol. 8783, 2020.