Cybersecurity

SURVEY

Open Access

Blockchain abnormal behavior awareness methods: a survey

Chuyi Yan[1,2], Chen Zhang[1,2], Zhigang Lu[1,2], Zehui Wang[1,2], Yuling Liu[1,2*] and Baoxu Liu[1,2]

Abstract

With the wide application and development of blockchain technology in various fields such as finance, government affairs and medical care, security incidents occur frequently on it, which brings great threats to users' assets and information. Many researchers have worked on blockchain abnormal behavior awareness in respond to these threats. We summarize respectively the existing public blockchain and consortium blockchain abnormal behavior awareness methods and ideas in detail as the difference between the two types of blockchain. At the same time, we summarize and analyze the existing data sets related to mainstream blockchain security, and finally discuss possible future research directions. Therefore, this work can provide a reference for blockchain security awareness research.

Keywords: Blockchain, Abnormal behavior, Awareness, Supervision, Security detection

Introduction

With the development of big data, cloud computing, and artificial intelligence, society enters the era of the Value Internet. These technologies have profoundly affected the world's production and lifestyle, making the world have a subversive change. The Value Internet era pays more attention to the value of data, and the characteristics of blockchain technology such as decentralization, anonymity, and non-tamperability fully cater to the era's attention. Since the Ministry of Industry and Information Technology released the "White Paper on China's Blockchain Technology and Application Development (2016)" in 2016, China has paid more and more attention to the development of blockchain, and blockchain projects have also shown a blowout scene in all walks of life. However, the thriving blockchain industry has a "hidden crisis", so it is necessary to aware the security of the blockchain in multiple dimensions to maintain the healthy development of the blockchain industry.

Blockchain security incidents emerge endlessly. According to the incomplete statistics of the China National Vulnerability Database, the number of security incidents in the blockchain field in 2020 alone was as high as 555, causing economic losses of up to 17.9 billion dollars, an increase of 130% from 2019 (CNCERT/CC 2020). Security incidents in the blockchain include both service application security incidents (such as Ponzi schemes, money laundering, etc.) and security incidents caused by technical defects (such as smart contract vulnerabilities, imperfect incentive mechanisms in the consensus, etc.). Among them, the notorious "The DAO" event originated from the vulnerabilities of smart contracts in Etherum Classic (ETC) (Mehar et al. 2019). Hackers carried out a reentry attack on its "transfer first, then reset" model, which led to the hard fork of Ethereum and Ethers worth more than 60 million dollars was stolen. For the blockchain consensus mechanism, ETC encountered 51% attacks in January 2019 (Voell 2020). The attacker double-spent at least 4 transactions with more than 51% of the computing power within 4 h, a total of 54,200 ETC, worth 271 thousand dollars. For the blockchain service security incidents, the OneCoin incident (Attorney 2019) and Rubixi smart contracts (Rubixi 2016) both involve Ponzi schemes, in which the OneCoin incident "evaporates" 1 billion dollars from victims. According to the Massachusetts Institute of Technology (MIT) technology

*Correspondence: liuyuling@iie.ac.cn
[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
Full list of author information is available at the end of the article

report(Orcutt 2020), criminals laundered 2.8 billion dollars through cryptocurrency exchanges just in 2019.

The reasons for security incidents are also multi-dimensional. Firstly, due to the nature of the distributed ledger of blockchain itself, there are more applications with transactional properties; Secondly, the emerging blockchain technology is not that perfect and the threshold of use required by users is relatively high and many users have poor security awareness; Thirdly, transactional projects usually have more funds stored in them, and the attackers are profitable, which leads to frequent security incidents. If we can aware the existence of threats, then we can avoid related property losses (Gong et al. 2017; Xi et al. 2012). At present, blockchain behavior awareness is the use of technologies such as data mining, machine learning and deep learning in the different aspects of blockchians to detect and predict the risks on them. Various methods of behavior awareness can be used to detect threats for the above events, which can respond to attacks in advance and reduce the risk of using blockchains. The following work will describe the awareness technologies in detail of various abnormal events.

Blockchain can be divided into the public blockchain, consortium blockchain, and private blockchain according to different application scenarios. Since the consortium blockchain can be understood as the alliance of multiple private organizations, so consortium blockchain and private blockchain are combined for discussion. The difference between the public blockchain and the consortium blockchain (private blockchain) is shown in Table 1. The public blockchain refers to a blockchain that anyone in the world can read, send transactions which can be effectively confirmed, and can participate in the consensus process (Guegan 2017). Firstly, there is no restriction on the access group, and all nodes can enter and exit freely (Kwon and Buchman 2018). Secondly, its data is disclosed to the entire network, which is completely decentralized. Thirdly, the data that can be collected on it is large in scale and has complex behavioral diversity. In addition, because of the existence of massive data, the transaction speed on it is slow, and the public chain is in a weak trust environment. On the other hand, the consortium blockchain is only for specific groups, whose access requires authorization. Secondly, the semi-centralized method can achieve controllable anonymity. Thirdly, the service on the general consortium blockchain is relatively simple, so the data scale is small and the behavior diversity is simple. In addition, because of the above-mentioned reasons, the transaction speed on the consortium blockchain is faster than the public blockchain, and it is in a strong trust environment.

Because of the differences between the public blockchain and the consortium blockchain (Table 1), the awareness methods are different too. For example, when the data scale is small and the behavior pattern is simple, it is difficult to collect a large amount of data, which will result in insufficient training data set size so the awareness model is more difficult to produce, i.e., it is difficult to use the awareness method that heavily depends on the data set. When the degree of anonymity and centralization are different, the methods of identification are also different. In a completely decentralized environment, it is necessary to establish a reasoning relationship between entities to realize the identification. When the transaction speed is different, the awareness methods for network attacks (such as eclipse attacks) will also be different.

Therefore, according to the application scenarios of the blockchain, we introduce the behavior awareness methods on the public blockchain and the consortium blockchain in "Public Blockchain Abnormal Behavior Awareness" and "Consortium Blockchain Abnormal Behavior Awareness" sections. Combining various existing blockchain behavior awareness methods, we summarize and analyze the existing mainstream blockchain security data sets in "Blockchain Security Data Sets Summary and Analysis" section in order to facilitate blockchain security researchers to conduct experiments and research. Finally, we discuss some possible future research points of blockchain behavior awareness in "Conclusion" section. Figure 1 shows the content structure of this article.

Table 1 The difference between public blockchain and consortium blockchain

	Access characteristics	Centralization	Anonymity	Trust environment	Transaction speed	Data scale	Behavioral diversity	Typical applications
Public blockchain	Unlimited group Free access (Li 2020; Wei 2018)	Decentralization	High anonymity	Week	Slow	Large	Complex	Bitcoin Ethereum
Consortium blockchain (Private blockchain)	Specific group Authorized access (Li 2020; Wei 2018)	Semi-centralization	Controllable anonymity	Strong	Fast	Small	Simple	HyperChain HyperLedger

Public blockchain abnormal behavior awareness

Public blockchain technology can be divided into three dimensions to aware abnormal behaviors. The first dimension is to aware of the risk of public blockchain's *network behaviors*, i.e., focusing on various abnormal behaviors at the network level of blockchain which not has a close connection with the laws and regulations of the physical space. The second dimension is to aware of the risk of the *subject behavior* in the public blockchain, i.e., focusing on the abnormal behavior of subjects on the blockchain. The third dimension is to aware of the risk of the *service behavior* of the public blockchain, i.e., not paying too much attention to the technical mechanism on the blockchain, but on the premise of the correct use of blockchain technology itself, focusing on the behavior risks which touching the relevant laws and regulations of the physical space. The following describes behavior awareness methods in detail and separately describes its common limitations, main ideas and concepts for addressing them or the future directions from these three dimensions in detail.

Public blockchain network behavior risk awareness

Blockchain technology is based on the P2P network. According to the hotspots of security issues in the blockchain's network layer, there are mainly eavesdropping attacks, eclipse attacks, border gateway protocol (BGP) hijacking attacks, segmentation attacks (Guo et al. 2020), selfish mining, 51% attacks, etc. This section mainly introduces the attack procession overview and awareness methods of three classic blockchain network attack behaviors: *eclipse attack* (Radix 2018), *selfish mining* (Frankenfield 2019b), and *51% attack* (Frankenfield 2019a). Table 2 compares the existing work on public blockchain network behavior risk awareness. The following describes the work in the table in detail and summarizes the procession of each attack.

Eclipse attack awareness

Eclipse attack is a common method for distributed network attacks. Attackers use this method to try to isolate and attack specific users instead of launching attacks on the entire network (such as Sybil attacks (Douceur 2002), etc.) (Radix 2018). The process of the eclipse attack was proposed by Heilman et al. (2015) in 2015. As each node in the Bitcoin blockchain network can only own 8 outgoing connections and 117 incoming connections, when a node restarts, it will re-configure the in and out connection, and select the address from the tried table to connect. Because the network layers of various blockchains are similar, not only can Bitcoin suffer from eclipse attacks, other blockchains may also suffer from. Without mastering a large number of zombie nodes, Marcus et al. (2018) can also use a small number of resources to launch the eclipse attack.

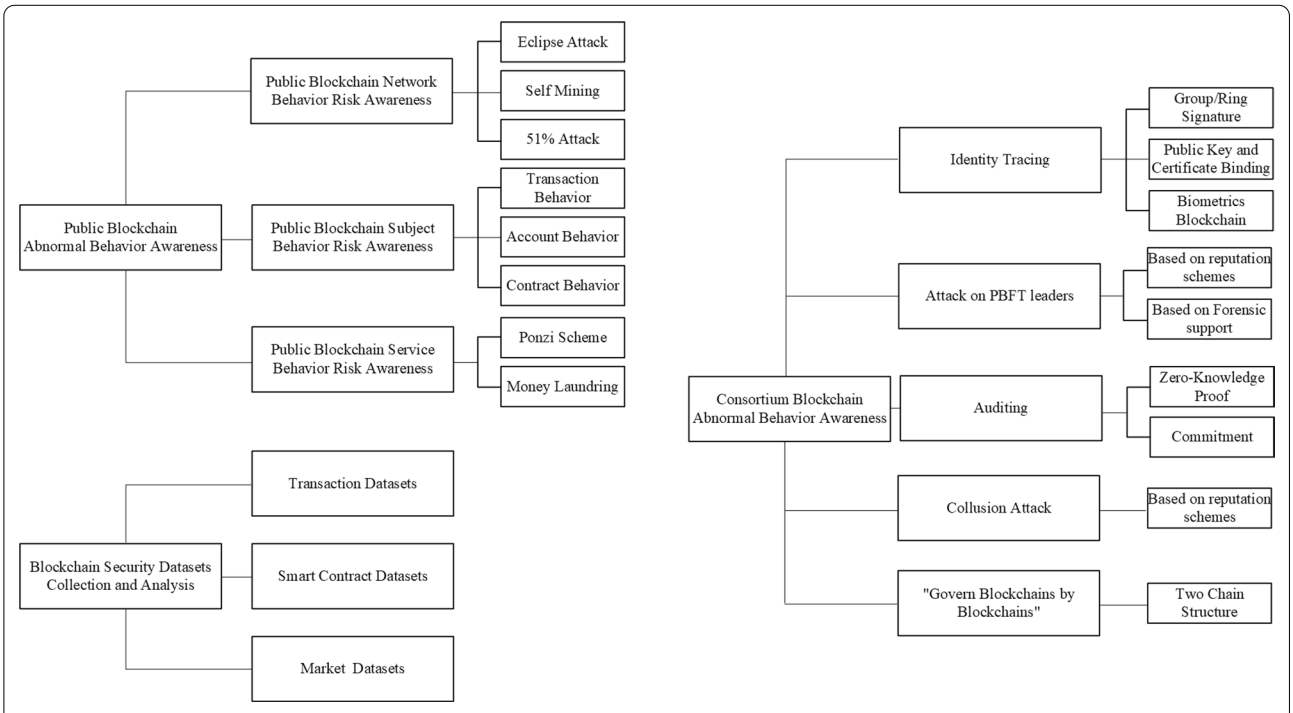

Fig. 1 Article content structure

Table 2 Public blockchain network behavior risk awareness work comparison

Network behavior	Behavior consequences	Awareness difficulty	Awareness methods	Technical details	Awareness accuracy	Awareness speed	Awareness availability	Advantage	Disadvantage	Quantitative performance	Data using
Eclipse attack	Isolate the target node	Hard	Based on game theory	Enhanced search technology based on dynamic voting mechanism (Ismail et al. 2015)	Low	–	Medium	Using fully decentralized	Need to reconfigure the routing table andrelated protocols	Look up success rate: 90–99% Malicious detection rate: 55–60% (depends on parameters)	Self-made data set by OMNeT++ simulator (Pongor 1993) and OverSim (Baumgart et al. 2007)
			Based on supervised learning	Use random forest classification algorithm to detect traffic packets (Xu et al. 2020)	Medium	Fast	High	Fast speed and high robustness	Rely on the label accuracy of the attack data set	Precision: 71% Recall: 95% F1: 0.81%	Access connection packets in Ethereum network with attacks they made by script
			Based on probabilistic model	Suspicious block timestamp warning (Alangot et al. 2020)	High	Slow	High	Easy to deploy, no need to change network configuration and protocols	Too slow	Attack detection time: 3 h Malicious detection rate: 100% Need very long time	Bitcoin block heads information in 2018–2019
			Based on pattern matching	Gossip protocol traffic analysis (Alangot et al. 2020)	High	Fast	Medium	Fast speed, no need to change network configur- ation and protocols	Detection nodes are not easy to deploy	Attack detection time: immediate Malicious detection rate: 97.58%	
Selfish mining	Malicious competition for block rewards	Hard	Based on game theory	Block transaction additional expected confirmation height (Saad et al. 2019)	High	Fast	Low	Does not affect the growth and confirmation rate of the blockchain itself	Loss of a certain transaction fee, need to change the block structure	Success attack needs to reach 50% hash rate	Bitcoin network
			Based on probabilistic model	Fork height simulation statistics (Chicarino et al. 2020)	–	–	–	Analyze the fork probability caused by multiple attackers' hash rates in a simulated environment	Can only be tested in a small-scale simulation network	Success attack needs to reach 30% hash rate	Bitcoin network with simulation attacks by NS3 (Gervais et al. 2016)
51% Attack	Double spend or malicious competition	Easy	–	Monitor large transa- ctions and the time of block production in the mining pool	High	Fast	High	Fast speed and high robustness	–	–	–

Eclipse attack is difficult to aware on the blockchain network layer, but there is still a small amount of related work. In Ismail et al. (2015) proposed a method to deal with the eclipse attack in complex scenarios. This detection method can be applied to various P2P networks, so as blockchain. They use a dynamic voting-based mechanism to enhance the search technology to detect attacks, i.e. it detects whether peer B is attacked, so peer A is used to select a group of nodes that may be closest to B, and they are required to return contact information or forward A's message until B parse it or discarded after the timeout. This method allows fully decentralized use, but because the protocol mechanism is changed, the routing table and related protocols need to be reconfigured, so the availability is moderate and the malicious detection rate is a bit low. Without changing the protocol mechanism, Xu et al. (2020) proposed a random forest classification algorithm in 2019 to detect eclipse attacks on Ethereum through the collection of normal data packets and attack data packets. By analysis, they found that the attack packet contains tags such as packet size, access frequency, and access time which are helpful in detecting eclipse attacks. After training on the data packets collected from the blockchain network, the detection model proposed by Xu et al. can detect the eclipse attack in a higher probability with 71% precision and 95% recall.

Since the use of supervised learning, its accuracy depends on the tags' accuracy of the attack data set, which has certain limitations when the data set is not available. To reduce the degree of data sets dependence, Alangot et al. (2020) proposed two different efficient methods in 2020 to detect whether the Bitcoin client was eclipse attacked. The first method is based on the suspicious block timestamp. If the time between the current block and the previous block is too long, it means that the network has been partitioned. The malicious detection rate can reach 100%, but the awareness speed of this method is slow with 3 h. The second method is to use the natural connection between the client and the Internet to "gossip" outward the view of the client itself. It can detect whether the client is attacked through the analysis of the network traffic, and the awareness efficiency is high as it can detect the eclipse attack immediately and reach 97.58% malicious detection rate. Both methods do not need to change the network configuration and related protocols, but due to the second method of detection nodes are not easy to implement, so the method availability is moderate.

Now much work has focused on designing new protocols to counter eclipse attacks and perceiving the existence of eclipse attacks is also important. In the future, researchers may pay attention to network traffic level, extracting traffic feature of eclipse attacks to directly perceive the existence of attacks in blockchain nodes.

Eclipse attacks are the pre-stage of selfish mining and double-spending attacks. Perceiving the existence of eclipse attacks plays an important role in preventing eclipse attacks and subsequent attacks.

Selfish mining awareness

Selfish mining is an attack that exploits the consensus defect on the blockchain, which can give the attacker opportunities to compete for more benefits for himself in the case of losing the benefits of others (Frankenfield 2019b). The process is: the attacker does not announce the block immediately after finding it, but continues to mine more blocks. According to the situation of the public blockchain, the block is announced to the network strategically and selectively to win in the competition of other miners.

Eyal and Sirer (2014) made a systematic analysis on the attack strategy of selfish mining in 2014. They divided the blockchain into "public chain" and "private chain" during mining. Attackers will keep the blocks they mined on the "private chain", and the blocks mined by the honest will be directly announced to the "public chain". Since the "public chain" is public, after the honest miners announce the block, the attacker can immediately update his "private chain" and continue to mine on his "private chain" until the attacker's "private chain" is as long as the "public chain". Then the attacker chooses to publish the "private chain" and starts to compete with the honest miners for the main chain. Once the "private chain" announced by the attacker becomes the main chain, the purpose of invalidating the honest miners' blocks can be achieved. They increase their own profits at the cost of wasting the computing power of honest miners. When there is half of honest miners mining after the attacker's block, the attacker's computing power must reach 25% to make a profit. When there are no honest miners mining after the attacker's block, the attacker's computing power must greater than 1/3 can make a profit (Han et al. 2018).

To aware of the existence of selfish mining attacks, Saad et al. (2019) proposed an awareness method in 2019, adding the expected confirmation height to each transaction of the block. When the average expected confirmation height is low in the block, it will be considered as a selfish miner. Because selfish miners will announce empty blocks for competition, although they will lose a part of the transaction fee, they can gain a speed advantage in mining. Therefore, when empty blocks appear, the expected confirmation height of the transaction will be lower, and it can be detected at this time and the system will deny the blocks access to the chain. When using this method, attackers need to reach the 50% hash rate to make it successful. But this awareness method needs to change the block structure. Without changing the block

structure, Chicarino et al. (2020) proposed a blockchain network selfish mining awareness method for the proof-of-work (PoW) consensus mechanism blockchain in 2020, simulating the blockchain through the NS3 simulator to detect if the fork height of the block deviates from the standard value. Then it determines whether a selfish mining attack has occurred or not in the current network. When using this method, attackers need to reach the 30% hash rate to make it successful.

There is not a lot of work on perceiving selfish mining. Most of the work is aimed at selfish mining defensive schemes, such as directly changing the consensus mechanism of the blockchain. Shultz and Bayer (2015) proposed a method in 2015 to let honest nodes attach some signatures to prove that the block is accepted by the network and that there is no competing block in the network. Zhang and Preneel (2017) proposed to change the longest chain principle in 2017. They introduced the concept of blockchain weight and proposed a new fork-resolution principle (FRP), which is not based on the longest chain principle to choose the forks but based on the block weight.

The selfish mining attack requires the attacker to reach certain computing power. In the public chain ecology, mining pools have chance to own a large amount of computing power. If attackers do not join the mining pool, they need to control more nodes to increase their computing power, such as implanting mining programs on the node devices. So in our opinion, we can start with the awareness of nodes and detect the pre-attack stage to prevent the implantation of mining programs, thereby detecting selfish mining behavior.

51% Attack awareness

51% attack, as the name suggests, means that the attacker has 51% of the computing power in the blockchain network and can mine new blocks faster than other participants to control the generation of new blocks or recalculate confirmed blocks, in order to tamper with the transaction data on the block and destroy the immutability of the blockchain. The general purpose of the attacker is to achieve double spending (Frankenfield 2019a). Attackers use the cryptocurrency in his hand to trade with others. After they confirm that the transaction is on the chain, the honest one will think that the transaction has been completed. At this time, the attacker has received money or materials. However, the attacker uses the advantage of computing power to start the fork from the block before the payment transaction, and transfers the cryptocurrency back to his address on the new block. The fork will compete with the entire network. Since the attacker has 51% of the computing power, the length of the private chain mined by the attacker will exceed the length of the original main chain, thus the private chain combined with the original main chain becomes the new main chain. Eventually, the block with transactions between the attacker and the honest one becomes a soft fork and is discarded, and will not be accepted by the entire network. The attacker achieves double-spending.

Using behavior awareness technology to detect 51% attacks has less work. The reason is that it is difficult for malicious miners to reach 51% of the computing power even use collusion attack before the mining pool appears. After the mining pool appears, if the mining pools whose computing power is on the top of a collusion attack, their computing power will reach 51%. Due to the assumption that an attacker is a rational person, the purpose of the attacker is to make a profit. If a person with 51% of the computing power performs a double-spending attack, it is equivalent to attack the blockchain trust system so that the corresponding currency is no longer valuable, so it is better to use the advantage of computing power to mine for profits.

Moreover, in Satoshi Nakamoto's article (Nakamoto and Satoshi 2008), it is also clear that if the attacker can catch up with the 6 block gap with the honest, it will take at least 1 h, so large transactions and block producing time from the mining pool can be considered to supervise, and related work can be studied in the future.

Public blockchain subject behavior risk awareness

According to the different subjects on the blockchain, we subdivide the subject behavior risk awareness on the public blockchain into *transaction behavior* awareness, *account behavior* awareness, and *contract behavior* awareness. Table 3 compares the existing work on public blockchain subjects' behavior risk awareness. We describe the work in detail and summarize the awareness process of each kind of subject's behavior.

The three different subject behavior awareness technologies can be roughly divided into two categories: one is based on machine learning, and the other is based on rules. Machine learning based methods can be divided into supervised and unsupervised. Early research on subject behavior awareness technology on the chain mostly used unsupervised machine learning methods. There was less prior knowledge of on-chain behavior and fewer abnormal clues available at the time, therefore, unsupervised learning which does not depends on labels was used. However, the accuracy is not satisfying and it is easy to overlook some potential abnormal behaviors. With the deepening of research work, the increase of abnormal cues and the deepening of on-chain behavior cognition, the subsequent works use of more accurate supervised learning methods and rule-based methods to perceive abnormal behavior on the chain. The supervised

Table 3 Public blockchain subject behavior risk awareness work comparison

Subject behavior	Awareness purpose	Awareness difficulty	Awareness methods	Technical details	Applicable scene	Awareness accuracy	Awareness speed	Advantage	Disadvantage	Precision	Recall	F1	Data using
Transaction behavior	Fraud behavior detection	Easy	Based on unsupervised learning	Multivariate cluster analysis based on trimmed k-means algorithm (Monamo et al. 2016)	More types needs for abnormal behavior classification	Low	Fast	High rubustness	Unsatisfactory precision	Efficient cluster number: 8 malicious detection rate: 16.67%			Bitcoin partial transaction dataset
			Based on unsupervised learning	Cluster analysis based on k-means algorithm after finding outliers (Sayadi et al. 2019)	Less types needs for abnormal behavior classification	High	Fast	High recognition accuracy for a small number of abnormal types	Fewer types of abnormal behavior can be detected	90.00%	99.70%	94.00%	
			Based on rule inference	Anomaly behavior recognition based on transaction motivation (Shen et al. 2021)	Targeted awareness needs for specific abnormal	Medium	Medium	Strong pertinence	Weak robustness	Airdrop candy: 43.62% greedy injection: 54.32%	Airdrop candy: 85.71% greedy injection: 81.35%	Airdrop candy: 57.79% greedy injection: 65.11%	

Table 3 (continued)

Subject behavior	Awareness purpose	Awareness difficulty	Awareness methods	Technical details	Applicable scene	Awareness accuracy	Awareness speed	Advantage	Disadvantage	Precision	Recall	F1	Data using
Account behavior	Suspicious account behavior detection	Medium	Based on unsupervised learning	Cluster analysis based on account behavior graph (Pham and Lee 2016)	More types needs for abnormal behavior classification	Low	Fast	High rubustness	Unsatisfactory precision and recall rate	Efficient cluster number: 7 Significant attack events: 2			Bitcoin partial transaction dataset
			Based on unsupervised learning	Account behavior analysis based on graph mining method combining event information and account high-level interaction information (Ao et al. 2021)	Fine-grained requirements for account behavior description	Medium	Medium	Highly fine-graied awareness	Unsatisfactory speed and accuracy	Average modularity: 0.801			Ethereum on-chain data
	Node behavior pattern classification	Hard	Based on unsupervised learning	Blockchain behavior clustering algorithm BPC analysis and verification (Huang et al. 2017)	Less types needs for abnormal behavior classification	Medium	Fast	High recognition accuracy for a small number of abnormal types	Low fine-grained awareness	Efficient cluster number: 2 Precision: 74.26%			Transaction data of a real blockchain application on stock trading
	Fraud account detection	Easy	Based on supervised learning	Using XGBoost to identify large scale fraudulent accounts and features sensitivity analysis (Ostapowicz and Żbikowski 2020)	Needs for fraud prediction on the account	High	Fast	High accuracy	Low fine- grained awareness	Random forest: 85.71% XGBoost: 78.03%	Random forest: 23.67% XGBoost: 31.32%	Random Forest: 37.09% XGBoost: 44.70%	Ethereum On-chain Data with phishing labels
	Identity inference	Medium	Based on unsupervised learning	Identity analysis based on GNN (Shen et al. 2021)	Deanonymization for abnormal behavior	High	Fast	High recognition accuracy with effectively large-scale graphed computation avoidance	Label dependency	99.17%	99.83%	99.50%	Ethereum phishing transaction network

Table 3 (continued)

Subject behavior	Awareness purpose	Awareness difficulty	Awareness methods	Technical details	Applicable scene	Awareness accuracy	Awareness speed	Advantage	Disadvantage	Precision	Recall	F1	Data using
Contract behavior	Vulnerability detection	Easy	Based on automated exploit	Provide a universal definition of contract vulnerability and exploitation tools (Krupp and Rossow 2018)	All blockchains that support smart contracts	High	Fast	Can be implement on a large scale with a high degree of automation	Limited to contract internal behavior	Successful exploit contract rate: 88.41%			Ethereum on-chain contracts
	Honeypot detection	Medium	Based on symbol execution	Propose honeypot contract classification and construct heuristic honeypot contract search tool (Torres et al. 2019)	All blockchains that support smart contracts	High	Fast	High accuracy and fast perception speed	Limited to contract internal behavior and requires open source smart contract	94.38%	–	–	
	Attack detection and classification	Medium	Based on supervised learning	Use neural network to train the interaction graphs between contracts to automatically discover the stage of new attacks (Su et al. 2021)	Needs for attack stages identification	High	Fast	Can be implement on a large scale with strong robustness	Attacks limited to the use of smart contract transactions	95.07%	94.73%	94.83%	DEFIER extended DApp event dataset

method is more dependent on labels, and it is not easy to obtain high-level information of behavior features. The rule-based awareness method requires a certain degree of cognition of on-chain behaviors, and the design of rules generally only aims at the awareness of specific behaviors, and the robustness needs to be improved.

Transaction behavior awareness

Blockchain transactions are packaged on the chain by miners after broadcasting in the entire network. Due to the transparency of the blockchain, abnormal behavior can be aware through the analysis of transaction records. Monamo et al. (2016) used the trimmed k-means method to perform cluster analysis and abnormal transaction awareness in multivariate settings in 2016. This method can detect more types of abnormal transactions. As using the unsupervised learning method, it does not depend on the accuracy of the data set and the method is robust, but the accuracy is not ideal with the malicious detection rate only 16.67%.

To improve the accuracy, Sayadi et al. (2019) proposed a new model for abnormal transaction detection in 2019. Firstly, they use the One-Class support vector machine (One-Class support vector machine, One-Class SVM) to find transactions outliers. Then they use the k-means algorithm to classify the outliers. In the case of fewer abnormal behaviors' types, it can achieve a higher accuracy rate reached 90% but there are fewer abnormal behaviors' types that can be accurately detected, which is not suitable for blockchain with complex behavior types.

Since transactions have intention, Shen et al. (2021) adopted a rule inference-based fusion method in 2021. They use transaction motivation as the starting point, and design rules for judging two types of abnormal transaction behaviors: airdrop candy and greedy fund injection. They extract the abnormal transaction pattern graph and use the sub-graph matching technology to design the awareness algorithm. They also verify the effectiveness of the method through real cases and the recall reached 85.71% and 81.35% of airdrop candy and greedy fund injection respectively. But this method aware of specific abnormal behaviors on the chain, so when new abnormal behaviors appear, the robustness of this method is weak.

We summarize the general process of transaction behavior awareness as follows: Firstly, obtain a large amount of transaction information from the blockchain (regardless of whether the blockchain supports smart contracts). This information acquisition method is applicable to various blockchains; After obtaining the transaction information, there are two types of processes. One is to use the inter-graph learning technology for model training and prediction after the transaction graph is constructed. The other one is to directly use the features in the original transaction information and then use the clustering method to find abnormal behaviors.

The common limitation of transaction behavior awareness methods is it that multi-classification's accuracy is poor and inadequate understanding of transaction behavior combined with semantics or other features. So researchers may focus on the blockchain transaction modeling combining other features, such as time sequence features, semantics features and etc.

Account behavior awareness

Although the blockchain has a detailed record of the transaction content and the addresses of both participants to the transaction, the account entities still have a certain degree of anonymity, so the awareness of account behavior is also worth exploring. Pham and Lee (2016) proposed a unique address anomaly awareness method for the Bitcoin network in 2017. They use the Bitcoin transaction network to form a graph with transaction addresses as nodes and using the k-means clustering algorithm, Mahalanobis distance, and the unsupervised support vector machine to aware of abnormal address behaviors. This method can find many types of suspicious behaviors, but the accuracy of the three methods needs to be improved.

To improve the classification accuracy, Huang et al. (2017) first proposed an automatic classification method of blockchain account behavior in the same year. By extracting the node behavior sequence and they proposed a clustering algorithm called BPC based on the k-means algorithm. Variety types of behaviors on the blockchain can be classified with good classification accuracy (reached 74.26%). Unfortunately, this method cannot be implemented on a large-scale blockchain. To increase the implementation scale, Ostapowicz and Żbikowski (2020) conducted a large-scale account detection method in 2020, using supervised learning techniques to detect fraudulent accounts on Ethereum, and compared the ability of random forests, support vector machines, and XGBoost classifiers on a data set with more than 300,000 accounts and give the sensitivity analysis of each feature. Though the precision has increased, the recall rate is low.

The account behavior awareness methods generally rely on extracting the statistical features of the account address (or transaction address) and then applying artificial intelligence technologies such as machine learning or deep learning to classify or identify them. But Ao et al. (2021) first combined the account high-level interaction information and time information, from the perspective of graph mining, analyzing the behavior of Ethereum accounts through Temporal High-order Proximity Aware Community Detection (THCD) and verifying the effectiveness on the four real data sets. Moreover, it can be

extended to large-scale transaction data sets with better perceptual fine-grained.

Identity inference aims to make a preliminary inference about account identity. Shen et al. (2021) present a novel approach to analyze user's behavior from the perspective of the transaction sub-graph, which naturally transforms the identity inference task into a graph classification pattern and effectively avoids computation in large-scale graph. They reached 99.17% precision, 99.83% recall and 99.50% F1-score. It performs pretty well. But this method depends on the labels of data sets.

We summarize the general process of account behavior awareness as follows: The general account behavior pattern is more suitable for blockchains that support smart contracts, that is, blockchain 2.0, such as Ethereum and other blockchains. But it can also be used in blockchain 1.0, such as the Bitcoin network. Generally, clustering data mining techniques are used to classify account behaviors, and then abnormal behaviors can be manually detected from the classified categories. In the blockchain 2.0 network, various security reports can be used as the basis, and we can use machine learning technology for classification and identification through statistical features and association methods between accounts.

The balance of precision and recall need to be improved and the new artificial intelligence methods which used in social interaction network can migrate into the on-chain account behavior field.

Contract behavior awareness

The development of smart contracts is fast and changeable. When met certain conditions, the corresponding transaction behaviors will be triggered. A smart contract can be understood as an automated trusted third party. The behaviors within and between contracts have very important influences on on-chain behaviors (Zhao et al. 2020). Since there is a lot of work on smart contracts, the following summarizes the work on the top security conference papers in recent years.

In terms of the internal behavior of the contract, Krupp and Rossow (2018) built a smart contract vulnerability automatic identification and exploit tool TEETHER in 2018. The contract can be exploited when only give the binary bytecode. This method achieves good performance for large-scale implementation on Ethereum and the successful exploit contracts account for 88.41%.

In terms of functional threats, Torres et al. (2019) systematically studied honeypot smart contracts and proposed a honeypot technology taxonomy in 2019, and constructed a heuristic method that uses symbolic execution to find honeypot contracts–HONEYBADGER. After analyzing more than 2 million smart contracts, their method can efficiently and accurately identify contracts with honeypot behavior.

In terms of inter-contract behavior, Su et al. (2021) designed the automatic large-scale attack investigation tool DEFIER for Ethereum in 2021. They use smart contracts with determined Ethereum attack events to form a seed attack set, and then use the Jaccard similarity between contracts to find smart contracts with new attack events, to form the largest attack event data set for smart contracts on Ethereum. Then they use Long Short-Term Memory (LSTM) neural network to train the transaction correlation vectors on the formed data set. After that, they use Multi-Layer Perceptron (MLP) classifier to classify attack stages. In this way, new attacks using smart contracts in Ethereum can be automatically discovered, and the attack stage can be determined. After deploying DEFIER in Ethereum, 476,342 malicious transactions were discovered, including 75 0-day vulnerabilities. In addition, they reached 95.07% precision, 94.73% recall and 94.83% F1-score.

The awareness of smart contracts is the most changeable. It can perceive in the aspect of the vulnerability level, functional threat level, and attack process level. Each level of awareness can use different processes. For example, vulnerability level awareness can establish tools such as fuzzing, symbolic execution, and formal verification. Functional threat level awareness can perceive honeypot contracts and study active traceability techniques, etc. Attack process level awareness is based on the establishment of the database, then it carries out data cleaning and analysis, and finally obtains an accurate attack stage database with awareness model.

Although the discovery of abnormal behavior of the internal code of the contract is very important, the interaction between the contracts also cannot be ignored. At present, there are few works on the interaction between contracts. How to reasonably model the interaction between contracts and adopt a reasonable segmentation method to capture high-level behavior information are points that can be studied in the future.

Public blockchain service behavior risk awareness

The service behavior risks on the blockchain mainly include illegal money laundering, Ponzi schemes, lightning loan utilization, darknet illegal transactions, terrorist financing, and virus extortion (Yang 2020). This section mainly summarizes the work of *Ponzi scheme awareness* and *illegal money laundering behavior awareness* as these two behaviors have more awareness work on the public blockchain. Table 4 compares the existing work on public blockchain service behavior risk awareness. We describe the work in detail and summarize the

awareness process of the above two service behavior risks.

The service behavior awareness technologies mainly based on machine learning as they rely on the abnormal clues. And the technologies limitation is the same as the subject behavior awareness.

Ponzi scheme awareness

Ponzi scheme is a form of financial fraud, which deceives investors and uses the funds of later investors to pay profits to previous investors (Wikipedia 2021b). Because blockchain users do not understand its underlying technology, public blockchains have strong anonymity, lack national supervision and smart contract source codes may be hidden, etc., there are many Ponzi schemes on public blockchains, which bring blockchain investors have caused large economic losses. According to statistics, an average of 7 million U.S. dollars' worth of Bitcoin was obtained by fraudsters within a year (Vasek and Moore 2015).

Chen et al. (2018) proposed a method for Ponzi scam contract identification using XGBoost in 2018. This method combined with security reports, collecting 1382 verified smart contracts from the Ethereum browser. Comparing the transaction flow graph with the non-Ponzi flow graph, and constructing seven statistical features to use XGBoost for classification. They achieved a precision rate of 94% and a recall rate of 81%, which performs efficiently in blockchains with sufficient transaction scale.

Bartolettiet al. (2020) used on-chain transaction records, smart contract source code, and smart contract chain information to analyze the behavioral characteristics and multi-angle impact of the Ponzi scheme. Finally, they find extra 184 ponzi schemes on Ethereum.

In order to further improve the accuracy rate, Zhang and Lou (2021) proposed a Ponzi scheme contract detection method based on the deep neural network in 2021. This method extracts the operation code feature of the smart contract and account feature to form a data set. After deep neural network training, it achieved the 99.6% precision rate and 96.3% recall rate.

This article summarizes the main process of the existing Ponzi scheme awareness methods as follows: Firstly, finding the Ponzi scheme incidents in various security reports and news to extract the malicious contract accounts, transaction associations, and other information in incidents, constructing the Ponzi scheme incidents data sets. Secondly, constructing a transaction flow graph for comparison to complete the feature extraction. Thirdly, using machine learning or deep learning methods to perform model tuning and training. Finally, using the trained model to detect new Ponzi schemes.

The awareness methods of the on-chain Ponzi scheme currently relies on the existing abnormal clues, and in this way the awareness of the Ponzi scheme contract with unknown behavior patterns is difficult. Research may combine behavior patterns defined in the financial field, abstract its network motifs, and cruise on-chain contracts through pattern recognition technology.

Money laundering awareness

Illegal money laundering refers to the illegal process of concealing the source of illegally obtained funds through a series of complex bank transfers or commercial transactions. The overall process is to return "clean" money to the money launderer in an obscure and indirect way (Wikipedia 2021a). The anonymity of cryptocurrency, the convenience of cross-border transactions, and its vulnerability to network attacks all provide a suitable soil for money laundering.

The public blockchain digital currency led by Bitcoin has become a common currency for transactions between certain criminals. In fact, cryptocurrency only provides pseudo-anonymity, but it can achieve true anonymity through certain methods. Cryptocurrency has no restrictions in cross-border transactions, and it is suitable for use as an intermediate currency to transfer between countries and become a medium for money laundering. At the same time, cryptocurrency is traded on the network instead of offline. Due to the instability and vulnerability of the network, a large amount of digital currency can be "evaporated" in a short period. The so-called "evaporation" means that money launderers can claim that the currency has been "evaporated" by cyberspace attacks, but in fact, it may have been withdrawn offline (such as embezzlement of public funds, etc.), which can wash away their own suspicions to a certain extent. A explainable reason was found for the source of the money, and making preparations for the escape after the crime. The above three points are the motivations for money laundering groups to use cryptocurrency for money laundering.

The existing common money laundering methods are shown in Table 5, which can be roughly divided into the following two types: The first one is to use the digital currency mixing strategy with high concealment to achieve anonymity. Bitcoin Fog, DarkLaunder (BitLaunder, CoinMixer are the same type), and Helix are commonly used for mixing. Among them, Bitcoin Fog is used with the "send shared" function of Blockchain.info (blockchain browser's API), to anonymize transactions without linkability and traceability. DarkLaunder (BitLaunder, CoinMixer) is a weak currency mixing strategy. After using it, certain transactions can still have linkability, but the money laundering speed is fast, which only

Table 4 Public blockchain service behavior risk awareness work comparison

Service behavior	Awareness difficulty	Awareness methods	Technical details	Applicable scene	Awareness accuracy	Awareness speed	Advantage	Disadvantage	Precision	Recall	F1	Data using
Ponzi schemes	Easy	Based on supervised learning	Using XGBoost to classify the Ethereum transaction flow graphs (Chen et al. 2018)	Blockchain with sufficient transaction scale	High	Fast	High precision	Unsatisfactory recall rate	94.00%	81.00%	86.00%	Ethereum transaction data
		Based on mixed information model construction	Fusion of transaction records, contract source code and on chain information to build a Ponzi scheme behavior model (Bartolettiet al. 2020)	Blockchain with sufficient transaction volume and support for smart contracts	–	–	Comprehensive analysis dimensions	Multi-dimensional data is weakly correlated	Find extra 184 ponzi schemes on Ethereum			
		Based on unsupervised learning	Smart contract operation code and account feature extraction based on deep neural network (Zhang and Lou 2021)	All blockchains that support smart contracts	High	Fast	High precision and recall rates	High data set dependence	99.60%	96.30%	97.92%	Ethereum transaction data combined with contract code data

Table 4 (continued)

Service behavior	Awareness difficulty	Awareness methods	Technical details	Applicable scene	Awareness accuracy	Awareness speed	Advantage	Disadvantage	Precision	Recall	F1	Data using
Money laundering	Easy	Based on unsupervised learning	Compare the transaction flow graph between the normal transactions and money laundering transactions (Hu et al. 2019)	Blockchain with sufficient transaction scale	High	Medium	High precision	Unable to identify blockchain application with insufficient transaction scale	92.74%	97.37%	95.00%	Bitcoin partial transaction datasets
		Based on supervised learning	Use graph neural network to identify and visualize money laundering transaction sequence diagrams (Weber et al. 2019)	Popular cryptocurrency platform	High	Medium	High robustness and suitable for all kinds of popular crypto-currency platform	Unsatisfactory accuracy	97.10%	67.50%	79.00%	
		Based on active learning	Use less data to achieve recognition (Lorenz et al. 2020)	Blockchain with limited data and manpower	Medium	Medium	Independent of data set size	Robustness needs to be verified	With the number of labeled samples, the best F1: 83.00%			

Table 5 Blockchain money laundering strategies comparison

Type	Strategy characteristic	
	Application	**Application characteristic**
Mixing strategy	Bitcoin fog	Need to cooperate with related functions to achieve complete untraceability
	DarkLaunder (BitLaunder, CoinMixer)	The money laundering speed is fast but the transaction can still be linked
	Helix	It can only be accessed by Tor. Multiple addresses are in the same transaction so the anonymity is weak
Cross-border transfer	Use crypto-currency as an intermediary to complete asset transfer	

takes 1–6 h (de and Hernandez-Castro 2017). Helix can only be accessed using Tor, but in this service, the wallet addresses and withdrawal addresses of multiple users exist in the same transaction, so it is easy to identify these users and does not provide sufficient anonymity.

The second method is to use the efficient and convenient cross-border transfer function of cryptocurrency. The currency of country A can be exchanged for digital currency through the exchange of country A, and then send the digital currency to country B. Now the digital currency can be exchanged to the currency of country B through the exchange of country B, completing the asset transfer. However, lacking supervision, profit-driven exchanges can only rely on consciousness to prohibit money laundering, which is obviously unrealistic.

Money laundering using mixed currency services is not as safe and anonymous as these services claimed. Some very well-known currency mixing services still have major security flaws and privacy restrictions (de and Hernandez-Castro 2017). Hu et al. (2019) collected data for 3 years from 2016 to 2019. By creating transaction flow graphs for normal transactions and money laundering transactions, comparing the differences between the graphs, and then using the node2vec model framework to classify graphs, they achieve better results with high F1-score. But its accuracy needs to be improved on the blockchain with insufficient transaction scale.

In order to enhance the robustness of the method, Weber et al. (2019) adopted a fusion method based on probability statistics in 2019, using Elliptic's data set to construct a sequence diagram containing 200K transaction nodes, 234K transaction edges, and 166 node features and the label of whether the money laundering transaction is or not. Then using the deep learning method of graph neural network (GCN) to identify the money laundering behavior in the graph, and finally visualizing the model. This money laundering recognition method has achieved good results for various popular cryptocurrency platforms but it is highly dependent on manpower and data sets.

In order to reduce dependence on data sets, Lorenz et al. (2020) used active learning methods in 2020, using only 5% of tags to achieve better money laundering identification results. This is the optimized configuration for limited manpower and data scenarios.

The awareness of money laundering behavior is still developing. We summarize the general process of money laundering behavior awareness as follows: Due to the lack of anti-money laundering data sets, the existing work generally conducts the correlation analysis between addresses and transactions to form transaction flow graphs; then use the classification method between graphs or the graph-based deep learning method to classify or cluster the graphs to obtain the feature difference between the normal transaction and the money laundering transaction flow graphs.

However, there is little work on money laundering in cross-border transfers, and it can be prevented from the policy supervision of exchanges.

Consortium blockchain abnormal behavior awareness

The consortium blockchain is a cluster composed of multiple private chains. Generally, multiple industry units form the alliance (Community 2020; Hope-Baillie and Thomas 2016; Skuchain and Forum 2019), and only authorized nodes can access in it. The consortium blockchain will be an important type of chain in the future. For its healthy development, the awareness of abnormal behavior of the consortium blockchain needs more attention. The awareness methods on the public blockchain (generally, the awareness methods based on machine learning, graph neural network and association relationship can be used) can still be used on the consortium blockchain. Since the consortium blockchain is highly controllable and designable, it has some unique awareness methods. Consortium blockchain can conduct risk awareness from the network level, the subject behavior level, and the service behavior level.

In this chapter, we adopt the elaboration idea from point to plane, i.e., from the abnormal behavior perception method for specific nodes to the method for the whole chain. We focus on the awareness of *identity tracking, attacks on PBFT leaders, auditing under privacy*

protection, collusion attacks and "Govern blockchains by blockchains" (Chen 2020), which are the researchers mostly pay attention to because they are different abnormal behaviors from the public chain. Although there are few related works, these works can be used as reserve methods.

The following describes the work in detail and separately describes its common limitations, the main ideas and concepts for addressing them or the future directions.

Table 6 compares the existing work on the risk awareness of abnormal behaviors of the consortium blockchain under the premise of highly controllable and designable.

Identity tracing

Due to the autonomy of the consortium blockchain, it can learn from the work of identity tracing on some public blockchains (Li and Xu 2020; Lu and Xu 2017). For the identity tracing mechanism of the consortium blockchains, the following related work can be referred to as reserve algorithms.

Group and ring signature

Generally, linkable, traceable group, and ring signatures are used for identity tracing with the premise of privacy protection, which can realize user identity tracing and transaction auditing under certain conditions (Zheng et al. 2018; Fujisaki and Suzuki 2007; Liu et al. 2004). The signature can be verified, but the identity of the signer is kept secret. The administrator can open the signer's identity under the set conditions, i.e., obtaining the signer's public key, so as to realize the tracing and complete the identity association (Li et al. 2021). However, its application scenarios are limited, which is applicable to consortium blockchains with a small number of users. Though in Zheng's work (Zheng et al. 2018), the operations of multiplication and exponentiation reached 20 and 27 in signature generation phase respectively, it still take a long preparation time before signing.

The common limitation of group and ring signature methods is it that the speed in preparation phase, generation phase and verification phase is not that satisfying using in the large scale consortium blockchain. Now the researchers aim to propose more efficient group and ring signature by using small groups with the balance of security but there is little breakthrough. So, in the follow-up, some works proposed schemes for specific ledger identity tracing, trying to address the speed issues.

Public key and certificate binding

Ateniese et al. (2014) proposed a certificate-based bitcoin authentication system in the public blockchain, in which the user registers with a trusted third party (TTP), and the TTP issues certificates to the users so that it improves the credibility of Bitcoin addresses. This centralized method is convenient for the TTP to trace user identities and because the script is used the same as Bitcoin, the efficiency is also as good as Bitcoin scripts. Using this certificate binding method can trace the identity of the consortium blockchain. But this centralized approach is prone to abuse of the supervision power. El Defrawy and Lampkins (2014) proposed a multi-server collaborative storage and supervision scheme based on multi-party secure computing. The bottom layer uses secret sharing technology. When identity tracing is required, more than the threshold number of servers must participate in order to reveal user identity, which effectively prevents the abuse of supervisory power in the consortium blockchain.

However, in previous schemes, users need to register again to change its public keys for trading. It increase the burden of supervision center (no matter it is centralized or not). Li et al. proposed a tracing scheme which seperate the user's public key and the certificate for tracing, exploiting the transparency and integrity of the blockchain to ensure traceability. Unfortunately, it still abused of supervisory power as the single supervision center.

Most identity tracing methods are to bind the user's public key to the certificate. The supervision center uses the certificate to trace the public key correlation with the user's identity. The common limitation of public key and certificate binding methods is it that the heavy burden of supervision center as the updates and changes of users' public keys and the abuse of supervisory power with the single supervision center (common method using in the real scene). The main concept to address the limitation is to change the binding way and distribute the supervision centers by cryptographic primitives, respectively. However, there is still lack of the scheme that aim to address both of light burden and the distributed supervision centers. Moreover, although some of the existing technologies can complete identity tracing and content auditing, their robustness is weak and cannot be universally applied to various blockchains.

Biometrics information mergence

Alharthi et al. (2021) tried to balance the burden and privacy so they proposed a biometrics blockchain (BBC) to track malicious accounts on-chain using biometric information to lable message senders for privacy and ensure the credibility. It is applicable to the internet of vehicles with consortium blockchain as they tested their scheme demonstrating that the packet loss rate is less than 5% and the user computational cost is between 0.1 and 0.3 ms.

Table 6 Consortium blockchain abnormal behavior risk awareness work comparison

Awareness purpose	Technical methods	Technical details	Applicable scene	Advantage	Disadvantage	Quantitative performance	Data using
Identity tracing	Group signature and ring signature	The identity of the signer is confidential but verifiable, and the administrator can open the identity of the signer (Zheng et al. 2018; Fujisaki and Suzuki 2007; Liu et al. 2004)	Consortium blockchain acquiring anonymity while can tracing identity in the abnormal situation	Satisfy identity anonymity and linkability	Long preparation time and slow speed before signing	Operations of Multiplication: 20 Operations of Exponentiation: 27 Low computational complexity	–
	Public key and certificate binding	Register the public key with a trusted third party to in- crease the credibility of the address (Ateniese et al. 2014)	Consortium blockchain with single institution of the supervision center	Fast and accurate identity tracing	Abuse of supervisory power	Script efficiency is as good as Bitcoin	–
		Cooperative registration of public keys based on multi-party secure computing (El Defrawy and Lampkins 2014)	Consortium blockchain with multi-institutions of supervision centers	Can prevent abuse of supervisory power	Low tracking efficiency and slower speed	–	–
		With the change of the user's public key, only a single registration to the supervision center (Li et al. 2021)	Consortium blockchain with single institution of the supervision center	Reduce the burden on users and the supervision center	Abuse of supervisory power	Efficiency is as good as Groth–Sahai proof system (Groth and Sahai 2008)	–
	Biometrics blockchain (BBC)	Using biometric information to track malicious accounts on-chain by BBC architecture (Alharthi et al. 2021)	Internet of vehicles with consortium blockchain	Use biometric information to label message senders for privacy and ensure message credibility	Latency time of chain updated	Packet Loss Rate: Less than 5% Computational Cost: 0.1–0.3 ms	Self-made chain with simulation attacks by OMNeT++ (Pongor 1993)

Table 6 (continued)

Awareness purpose	Technical methods	Technical details	Applicable scene	Advantage	Disadvantage	Quantitative performance	Data using
Attack on PBFT leaders	Based on reputation schemes	Use reputation model to evaluate leaders' scores and perceive leader nodes (Lei et al. 2018)	Consortium blockchain using PBFT consensus mechanism	Identify malicious leader leader nodes in time	The effects in non-experimental environments need to be further tested	Feasibility: Delay time with in 22.0 s Reliability: Linear	Self-made chain prototype
	Forensic support based	Use cryptographic primitives such as aggregate signatures and commitments to take BFT forensic support (Sheng et al. 2021)	Consortium blockchain using BFT consensus mechanism	Forensic support can visualize	Large scale test need to be further tested	–	–
Auditing under privacy protection	Zero-knowledge proof and commitment	Additive homomorphism commitment to hide sensitive data and complete the audit with zero-knowledge proof guaranteeing audit reliability (Narula et al. 2018)	Consortium blockchain with audit content kept confidential	High privacy and audit reliability	Limited auditing operations	Computational Cost: Linear (Validate for 20 nodes in less than 200 ms) Auditing Time with More Nodes: Linear	Self-made chain prototype
Collusion attack	Based on reputation schemes	Use Bayesian inference model to evaluate reputation scores and perceive collusion attacks (Yang et al. 2018)	Internet of vehicles with consortium blockchain	High feasibility in consortium ranges	Latency time of chain updated	Feasibility: Less than 1 s Reliability: Exponent	Users in vehicular and blockchain simulation platform
		Use the reputation chain to improve the performance of the transaction chain and perceive collusion attacks (Huang et al. 2020)	E-commerce environment with consortium blockchain	Sharding improves chain's throughput	The effects in non-experimental environments need to be further tested	Feasibility: 20.4 s delay time at least Reliability: Linear	Self-made chain prototype with simulated users
		Use smart contracts to evaluate reputation scores and resist collusion attacks (Zhou et al. 2021)	E-commerce environment with consortium blockchain	High feasibility and reliability	The robust effect on other attacks remains to be verified	Feasibility: Less than 0.8 s Reliability: Constant	Partial Ethereum users participant in

Table 6 (continued)

Awareness purpose	Technical methods	Technical details	Applicable scene	Advantage	Disadvantage	Quantitative performance	Data using
"Govern blockchains by blockchains"	Double-chain architecture	The double-chain consists of a detection chain and a data public chain. The detection chain deploys multi-feature models to detect malicious behaviors on the data public chain (Gu et al. 2018)	Consortium blockchain with multi-institutions of supervision centers	High accuracy and small scale high detection speed	Large scale testing is inefficient	Accuracy: 92.5% Recall: 94.6% F1: 93.5%	Drebin Dataset (Arp et al. 2014)
		The double-chain consists of the transaction chain and the custody chain. The custody chain uses a neural network to identify illegal transactions, and the double-chain anchors the public blockchain (Wu et al. 2020a)	Consortium blockchain with multi-institutions of supervision centers	Fast transaction speed, high scalability, and strong credibility	Practicability needs to be verified	Accuracy: 90.1% Recall: 18.5% F1: 30.8%	Elliptic Dataset

Biometrics technology seems satisfying but the biometrics information collection and the latency time of update are still unsatisfactory.

The methods for identity tracing are mainly based on the cryptographic primitives. Secure and reliable tracing in one side, unsatisfying speed in the other side. In our point of view, researchers may construct identity knowledge base and using artificial intelligence methods to trace the malicious identity.

Attack on PBFT leaders

Consensus mechanism in consortium chain mainly uses Practical Byzantine Fault Tolerance(PBFT)-based protocol, which is different from PoW or PoS consensus protocol used in the public chain. PBFT usually selects the leader nodes that complete consensus in the current epoch with assumption that more than 2/3 nodes are honest. For this reason, the leader node of each epoch becomes a vulnerable object and therefore needs to be aware of abnormal behavior aiming at the leaders. The related abnormal behavior awareness work is not that sufficient so this chapter only lists the representative research works.

Lei et al. (2018) proposed a reputation model to evaluate leaders' scores. The lower score nodes given, the lower probability can the nodes be selected as leaders. All the participants in this consortium chain can contribute to the reputation model so the scores can reflect the abnormal behavior on the leaders. As the scores are updated in real time, it can identify and dispose malicious leader nodes in time with high reliability (Linear). But their scheme were only tested in the experimental environment. Sheng et al. (2021) used cryptographic primitives such as aggregate signatures and commitments to take BFT forensic support. They mathematically formalize the study of forensic support of BFT protocols, aiming to identify as many of the malicious replicas as possible and in as distributed manner as possible.

The attack on PBFT leaders may suffer a heavy lost. However, the related works focus much on proposing a new consensus mechanism to avoid the risk of PBFT, while the deployment of leader node abnormal behavior awareness mechanism embedding in PBFT-based consortium chain is not sufficient. Such work is very important, because existing consortium blockchain projects have already used PBFT consensus protocol, which has many users. Due to the immutability of blockchain, updating the underlying consensus algorithm is costly, so the future research can give attention on the automatically abnormal behavior awareness of leader node and timely disposal.

Auditing under privacy protection

In terms of awareness methods for audit under privacy protection purposes, we can still use the method of group and ring signature and in this sub-section, we focus on the application perspective.

Narula et al. (2018) proposed a technology zkledger in 2018, that integrates privacy and auditing. It uses Pederson promises to hide sensitive data, and uses the additive homomorphism promised by Pederson to perform audit operations on the hidden data. It perceives the correctness of functions on the blockchain through a simple ciphertext summation method, and at the same time using zero-knowledge proof to ensure the reliability of the audit. In terms of audit integrity, they construct a unique accounting method and propose a virtual token called the audit token to make the transaction public and verifiable while ensuring the privacy of participants. In terms of system efficiency, zkledger adds commitment caches to each entity to improve the efficiency of transaction creation and auditing and uses the map/reduce parallel computing method to improve the efficiency of auditing in the case of multiple entities. The computational cost and auditing time grows linearly as the number of nodes increases.(Validate for 20 nodes in less than 200 ms) At the same time, in order to reduce communication costs, they use the Fiat Shamir non-interactive zero-knowledge proof instead of Schnorr's interactive zero-knowledge proof. This is a blockchain supervision architecture, which can be considered in integrating the authentication and authorization of members on the chain, and then upgrade zkledger to a consortium blockchain platform with a wider range of available scenarios.

The common limitation of auditing under privacy protection is it that the types of auditing operations are limited (only sum operations in Narula) as the cost of multiplication operations with privacy on-chain is costly and the auditing relies on the auditors, which cannot automated. So in the future, research can focus on the various types of auditing operations with low cost and the automated auditing.

Collusion attack

Collusion attack is a type of security attack or threat in which a node intentionally makes a secret agreement with an adversary. If fewer nodes participate in the consortium chain, the success rate of collusion attacks will be higher. Though using double-chain architecture can perceive collusion attack to some extent, the main technical methods are based on reputation schemes.

Yang et al. (2018) used Bayesian inference model to evaluate reputation scores which can detect the compromised nodes and perceive the collusion attacks in time.

Timeliness of the scheme make it suitable for the internet of vehicles with consortium blockchain. It only takes less than 1 s to evaluate the reputation scores with high feasibility but the reliability is exponent, i.e., unfair ratings increase exponentially as the number of malicious nodes increase. It means that when there are too many malicious nodes, the system may cannot detect the collusion attack precisely.

Huang et al. (2020) took a next step. They proposed a scheme using the reputation chain to improve the performance of the transaction chain which increases the reliability, i.e., unfair ratings increase linearly as the number of malicious nodes increase. As they use another reputation chain, more than 20.4 s delay time is needed for the chain updated, though they use sharding to improve chain's throughput.

To balance the feasibility and the reliability, Zhou et al. (2021) used smart contracts to evaluate reputation scores for the collusion attack awareness and control. The scheme only uses less than 0.8 s to evaluate the reputation scores with the reliability is constant, i.e., the reputation scores tend to be a constant as the rating number increases.

The common limitation of collusion attack awareness based on reputation schemes is it that the participants in the consortium chain need to be sufficient or the reputation score may not reflect the collusion attacks precisely. To address this limitation, history data can be added in the initialize phase.

The methods for collusion attack awareness mainly aims on the participants in the consortium chain who have the write-permission. Reputation schemes detect collusion attack with the assumption that most participants are honest. In our opinion, study can focus on merging the reputation schemes and history data into one architecture and balance the cost, feasibility and reliability to aware collusion attacks and other unknown attacks in the future.

"Govern blockchains by blockchains"

The academician of the Chinese Academy of Engineering Chun Chen proposed the supervision technology model of "govern blockchains by blockchains", i.e., using blockchain technology to govern the blockchain and its applications (Chen 2020). This model is also applicable to the abnormal awareness of consortium blockchains. By establishing a supervision blockchain and connecting to the supernodes of the company blockchain projects, it can detect the malicious behavior of each blockchain project through the supervision blockchain in time, and finally deposit evidence on the supervision blockchain permanently.

Gu et al. (2018) proposed a double-chain architecture in 2018, which consists of a detection chain, deploying multi-feature models to detect malicious behaviors on the data public chain, and a data public chain, which stores data with transparency and integrity. This scheme achieve high detection accuracy (F1 score is 93.5%) and high speed in small scale with multi-institutions of supervision centers.

Wu et al. (2020a) proposed a double-chain architecture in 2020, which consists of a transaction chain and a custody chain. As the core part of the system, the consortium blockchain is responsible for processing transaction collection, verification and packaging. As a participant in the consensus process, the regulator directly participates in the operation of the consortium blockchain. Users' complete transaction data encrypted is stored in the blocks of the consortium blockchain for traceability and privacy preservation. There are only a few participants in the consortium blockchain. In order to improve the credibility of the system, an anchor with the public chain can be considered. Store user status changing information and the block hash of the consortium blockchain in the public blockchain, which can prevent the members of the consortium blockchain from launching collusion attacks. Unfortunately, though their scheme achieve a satisfying accuracy, the recall rate is not practicable.

The technology of "govern blockchains by blockchains" needs further development, and its development includes data collaboration on and off the chain. However, it is difficult to achieve such a huge amount of data in a blockchain through traditional manual supervision and awareness methods. It requires code to implement rules and software for internal supervision. It is also a "game" process, which is a complex technological realization process (Hong et al. 2020).

The common limitation of double-chain architecture is it that the supervision chain only perceive one type of abnormal behavior and lacking of one supervision chain integrates more types of abnormal behaviors. Maybe we can use supervision chain with multi-classification abnormal behaviors detection model deployed. There is still a long way to go to develop "govern blockchains by blockchains" into systemizing.

Blockchain security data sets summary and analysis

In order to facilitate blockchain security researchers to explore and make experiments on the blockchain security issues, we summarize 13 data sets (a total of 29 subtable information) on the mainstream blockchain.

The current abnormal behavior awareness method of the consortium blockchain is booming. Various research objects are used while works on consortium chain mainly

Table 7 Blockchain security data sets summary and analysis

Dataset type	Labeled or not	Dataset name	Dateset description	Data size	Table name	Table description	Columns number	Columns description	Applicable scene	Source
Transaction	Yes	Elliptic datasets (Elliptic 2019; Weber et al. 2019)	Transaction graph classifying the illicit and licit nodes collected from the Bitcoin blockchain	200,000 bitcoin transactions	Elliptic_ txs_classes	Licit transactions or not	2	Transaction id and its class	Money laundring detection; Ponzi schemes detection	Ellipic
					Elliptic_txs _edgelist	Nodes and edges	2	Source and destination transaction ids		
					Elliptic_ txs_features	Transactions features	167	Transaction features		
		Transaction network of phishing nodes (Wu et al. 2020c; Yuan et al. 2020)	Phishing account information from Etherscan	1262 phishing accounts	Address	The node list for phishing detection	7	Part of block head information, transactions flow and contract address	Phishing account detection	Xblock
					Ethereum-network	Transaction subgraph	4	Transactions flow with time		
		The label of phishing account on ethereum (Chen et al. 2020)	The different attack tags of phishing account	2881 phishing addresses	Phishing_label	The different attack tags of phishing account	4	Account classes and its balance with transaction counts		
		Ethereum phishing transaction network (Chen et al. 2020)	A huge Ethereum transaction network extending from phishing nodes reported in Etherscan	2,973,489 nodes, 13,551,303 edges and 1165 labeled nodes	MulDiGraph	The network attributes	6	Transaction parties are phishing account or not and the transaction flow		
		Bitcoin partial transaction datasets (Wu et al. 2020b)	Snapshots containing partial transaction records of Bitcoin transaction data from 2014 to 2016	22,500,000 Bitcoin transactions	Blockhash	Information of block	4	Block head information	Money laundering detection; Mixing service detection	
					txhash	Transaction ID and hash pairs	2	Transactions id with its hash		
					Addresses	Bitcoin address ID and address pairs	2	Address with its ID		
					tx	Information of transaction	5	Transaction flow with time and its block id		
					txin	List of all transaction inputs	3	Input flow		
					txout	List of all transaction outputs	3	Output flow		
					Label	Addresses of mixing services	1	Addresses of mixing services		

Table 7 (continued)

Dataset type	Labeled or not	Dataset name	Dateset description	Data size	Table name	Table description	Columns number	Columns description	Applicable scene	Source
	No	Ethereum on-chain data (Zheng et al. 2020)	Ethereum on-chain data getting from the Ehtereum full node	10,999,999 blocks information	Block	Ethereum block information	14	Complete block head information	Extracting and exploring Ethereum	
					Normal transaction	Normal transactions information	10	Transaction flow with block head and gas information		
					Internal EtherTransaction	Smart contract execution transactions information	8	Contracts execution flow with its re-lated transactions and block information		
					ContractInfo	Contract information	11	Contracts information in its whole lifetime		
					ContractCall	Contract calling	11	Contracts calling flow with calling function, type and status		
					ERC20 transaction	ERC20 token transaction information	7	Transaction flow in ERC20		
					ERC721 transaction	ERC721 token transaction information	7	Transaction flow in ERC721		
Contract	Yes	Smart Ponzi scheme labels (Chen et al. 2018)	Labels of smart Ponzi contracts by manually check	3794 contracts labels	Ponzi_label	The labels of whether a contract is a smart Ponzi scheme	2	Contract type	Ponzi schemes detection	Xblock
		DEFIER extended DApp event Dataset (Su et al. 2021)	Labels of attack stages on contracts calling flow of DApps	92,644 Labels of attack stages on contracts calling flow	Stage_labels	The labels of which attack stages (using the kill chain model) the contract calling flow at	4	DApp's attack stages with its transaction lists	DApp events detection	Institute of Information Engineering, CAS
	No	Smart contract attribute dataset (Huang et al. 2019)	All open source contracts in Ethereum	14,000 contracts opening source in Ethereum	Open source contract info	All open source contracts in Ethereum	9	Contract basic information with contract code and transactions in it	Ponzi schemes detection	Xblock

Table 7 (continued)

Dataset type	Labeled or not	Dataset name	Dateset description	Data size	Table name	Table description	Columns number	Columns description	Applicable scene	Source
Market	No	Ether price and volume dataset (Han et al. 2020)	Price and Volume from 2015 to 2019 of Ether	7892 market data about Ether	4 h_data_eth	Market data about Ether as the exchange rate is ETH/USDT	8	High and low price and volume of Ether	Ether market causality analysis	Xblock
		Bitcoin price and volume dataset (Han et al. 2020)	Price and Volume from 2015 to 2019 of Ether	7892 market data about Bitcoin	4 h_data_btc	Market data about Bitcoin as the exchange rate is BTC/USDT	8	High and low price and volume of Bitcoin	Bitcoin market causality analysis	
		Mt.Gox leaked transaction (Chen et al. 2019)	Transaction data leaked by Mt.Gox exchange	Mt.Gox leaked transactions from 2012 to 2013	Complete_edge_v2	Transactions of bitcoin market	8	Transaction flow with users' type	Bitcoin market user behavior analysis	
		Activity information of DApps (Zheng et al. 2017)	Information about DApps'activity	1,400,000 DApp's activity information	Radar	Activity information of DApps on DappRadar	15	DApp's basic information, transactions and contracts in it and users information	DApp activity analysis	
					State_of_the_dapp	Activity information of DApps on State of the Dapps	28	DApp's state information, transactions and contracts in it and users information		

use data such as self-built prototypes. There is no standard public data set on consortium chain, so this chapter mainly introduces the data set on the public chain.

According to the type of data sets, it is divided into *transaction data sets*, *contract data sets* and *market data sets*. Each type of data sets is divided into labeled and unlabeled. The specific data sets' names, descriptions, scales, attributes, sources and etc. are shown in the following Table 7. At the same time, we also propose possible application scenarios for each data set for the reference of security researchers.

Conclusion

We discuss the behavior awareness work of public blockchains and consortium blockchains. Researchers have already done related work in this field, but blockchain behavior awareness field is still in its infancy. Therefore, this research direction still needs more work. Combining the above-mentioned limitation of awareness methods, we give some possible directions for future research on blockchain behavior awareness.

- *Blockchain interactive behavior modeling and processing*: The existing abnormal behavior awareness work on the blockchain usually starts with a single node of information, and may ignore the interaction behavior information between nodes or even between chains. However, the blockchain is a social network. Transaction behavior is essentially an interaction behavior. New interaction behaviors may be the pre-stage of a certain attack. Therefore, it is important to perceive potential threats brought by interaction behavior in time. In the future, works can further explore how to model the interaction behavior between various data on the chain, analyze the dynamic features of the interaction behavior, and design a recognition technology that can self-adapt to behavior changes on the chain.
- *Blockchain abnormal behavior fine-grained awareness*: The awareness methods of abnormal behavior on the existing blockchain is usually aiming at a single abnormal behavior, and the awareness of multiple behaviors results are unsatisfactory. Therefore, future work can further analyze various abnormal behaviors on the chain, deepen their understanding and characterization, integrate the features of different behaviors, and combine machine learning and deep learning methods to design fine-grained awareness methods and models of various abnormal behaviors on the chain.
- *Blockchain cross-space identity association technology*: For the healthy and long-term development of the blockchain, the contradiction between its anonymity and the requirement for real-name supervision needs to be resolved. It is difficult to aggregate a large number of anonymous accounts on the blockchain, and it is necessary to consider the association of the on-chain identities, especially with the help of the access mechanism of the consortium blockchain. We can study the correspondence between the identity of the physical space and the identity on the blockchain, in order to open up the "physical-virtual" space.
- *Blockchain dynamic monitoring and analysis technologies*: The existing research on the supervision and governance technology of the blockchain still needs to be further promoted. Therefore, future works can face the security risks existing in the blockchain ecology, study the regulatory technical framework and refine in-depth analysis, dynamic monitoring and identification technology, and achieve the accurate awareness of abnormal behaviors on the blockchain ecology and timely disposal of malicious behaviors.

Acknowledgements
Not applicable.

Authors' contributions
All authors have contributed to this manuscript and approve of this submission. CY participated in all the work and drafting the article. ZW and CZ did some basic collection work. Prof. YL, ZL and BL made a decisive contribution to the content of research and revising the article critically. All authors read and approved the final manuscript.

Funding
This research is supported by National Key Research and Development Program of China (Nos. 2021YFF0307203 and 2019QY1300), Youth Innovation Promotion Association CAS (No. 2021156), the Strategic Priority Research Program of Chinese Academy of Sciences (No. XDC02040100) and National Natural Science Foundation of China (No. 61802404). This work is also supported by the Program of Key Laboratory of Network Assessment Technology, the Chinese Academy of Sciences, Program of Beijing Key Laboratory of Network Security and Protection Technology.

Availability of data and materials
Not applicable.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China. [2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China.

Received: 7 September 2021 Accepted: 21 December 2021

References

Alangot B, Reijsbergen D, Venugopalan S, Szalachowski P (2020) Decentralized lightweight detection of eclipse attacks on bitcoin clients. In: 2020 IEEE international conference on blockchain (blockchain). IEEE, pp 337–342

Alharthi A, Ni Q, Jiang R (2021) A privacy-preservation framework based on biometrics blockchain (BBC) to prevent attacks in VANET. IEEE Access 9:87299–87309

Ao X, Liu Y, Qin Z, Sun Y, He Q (2021) Temporal high-order proximity aware behavior analysis on Ethereum. World Wide Web, pp 1–21

Arp D, Spreitzenbarth M, Hubner M, Gascon H, Rieck K, Siemens C (2014) Drebin: Effective and explainable detection of android malware in your pocket. Ndss 14:23–26

Ateniese G, Faonio A, Magri B, De Medeiros B (2014) Certified bitcoins. In: International conference on applied cryptography and network security. Springer, pp 80–96

Attorney O (2019) Manhattan U.S. Attorney announces charges against leaders of "OneCoin," a multibillion-dollar pyramid scheme involving the sale of a fraudulent cryptocurrency (2019). https://www.justice.gov/usao-sdny/pr/manhattan-us-attorney-announces-charges-against-leaders-onecoin-multibillion-dollar Accessed 8 March

Bartoletti M, Carta S, Cimoli T, Saia R (2020) Dissecting Ponzi schemes on Ethereum: identification, analysis, and impact. Future Gener Comput Syst 102:259–277

Baumgart I, Heep B, Krause S (2007) Oversim: a flexible overlay network simulation framework. In: 2007 IEEE global internet symposium. IEEE, pp 79–84

Chen C (2020) The key technologies of consortium blockchain and the supervision challenges of blockchain. China Ind Inf Technol 2020(11):54–58

Chen L, Peng J, Liu Y, Li J, Xie F, Zheng Z (2020) Phishing scams detection in Ethereum transaction network. ACM Trans Internet Technol 21(1):1–6

Chen W, Zheng Z, Cui J, Ngai E, Zheng P, Zhou Y (2018) Detecting Ponzi schemes on Ethereum: towards healthier blockchain technology. In: Proceedings of the 2018 world wide web conference. pp 1409–1418

Chen W, Wu J, Zheng Z, Chen C, Zhou Y (2019) Market manipulation of bitcoin: Evidence from mining the Mt. Gox transaction network. In: IEEE conference on computer communications. pp 964–972

Chen W, Guo X, Chen Z, Zheng Z, Lu Y (2020) Phishing scam detection on Ethereum: towards financial security for blockchain ecosystem. In: International joint conferences on artificial intelligence organization. pp 4506–4512

Chicarino V, Albuquerque C, Jesus E, Rocha A (2020) On the detection of selfish mining and stalker attacks in blockchain networks. Ann Telecommun 75:143–152

CNCERT/CC (2020) 2020 Blockchain security situation perception report. https://bc.cnvd.org.cn/notice_info?num=0c4088bbb6f7346000c3ac1ce13f0347 Accessed 5 Mar 2021

Community B (2020) Beam: the scalable confidential cryptocurrency. 2 Feb 2020. https://docs.beam.mw/BEAM_Position_Paper_0.3.pdf

de BT, Hernandez-Castro J (2017) An analysis of bitcoin laundry services. In: Springer (ed.) Nordic conference on secure IT systems. pp 297–312

Douceur JR (2002) The sybil attack. In: International workshop on peer-to-peer systems. Springer, pp 251–260

El Defrawy K, Lampkins J (2014) Founding digital currency on secure computation. In: Proceedings of the 2014 ACM SIGSAC conference on computer and communications security, pp 1–14

Elliptic: Elliptic data set (2019) https://www.elliptic.co

Eyal I, Sirer EG (2014) Majority is not enough: bitcoin mining is vulnerable. In: International conference on financial cryptography and data security. Springer, pp 436–454

Frankenfield J (2019a) 51% Attack. 6 May 2019. https://www.investopedia.com/terms/1/51-attack.asp

Frankenfield J. (2019b) Selfish mining. 1 Apr 2021. https://www.investopedia.com/terms/s/selfish-mining.asp

Fujisaki E, Suzuki K (2007) Traceable ring signature. In: International workshop on public key cryptography. Springer, pp 181–200

Gervais A, Karame GO, Wüst K, Glykantzis V, Ritzdorf H, Capkun S (2016) On the security and performance of proof of work blockchains. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. pp 3–16

Gong J, Zang X, Su Q, Hu X, Xu J (2017) Survey of network security situation awareness. J Softw 28(4):1010–1026

Groth J, Sahai A (2008) Efficient non-interactive proof systems for bilinear groups. In: Springer (ed.) Annual international conference on the theory and applications of cryptographic techniques, pp 415–432

Gu J, Sun B, Du X, Wang J, Zhuang Y, Wang Z (2018) Consortium blockchain-based malware detection in mobile devices. IEEE Access 6:12118–12128

Guegan, D.: Public blockchain versus private blockhain (2017)

Guo, Z., Guo, S., Zhang, S., Song, L., Wang, H.: Analysis of cross-chain technology of blockchain. Chin J Internet Things 35–48 (2020)

Han J, Zou J, Jiang H, Xu Q (2018) Research on mining attacks in bitcoin. J Cryptol Res 5(5):470–483

Han Q, Wu J, Zheng Z (2020) Long-range dependence, multi-fractality and volume-return causality of ether market. Chaos Interdiscip J Nonlinear Sci 30(1):011101

Heilman E, Kendler A, Zohar A, Goldberg S (2015) Eclipse attacks on bitcoin's peer-to-peer network. In: 24th {USENIX} security symposium ({USENIX} security 15), pp 129–144

Hong X, Wang Y, Liao F (2020) Review on the technology research of blockchain security supervision. Bulletin of National Natural Science Foundation of China 34(01):18–24

Hope-Bailie A, Thomas S (2016) Interledger: Creating a standard for payments. In: Proceedings of the 25th international conference companion on world wide web. pp 281–282

Hu Y, Seneviratne S, Thilakarathna K, Fukuda K, Seneviratne A (2019) Characterizing and detecting money laundering activities on the bitcoin network. arXiv:1912.12060

Huang B, Liu Z, Chen J, Liu A, Liu Q, He Q (2017) Behavior pattern clustering in blockchain networks. Multimed Tools Appl 76(19):20099–20110

Huang C, Wang Z, Chen H, Hu Q, Zhang Q, Wang W, Guan X (2020) Repchain: a reputation-based secure, fast, and high incentive blockchain system via sharding. IEEE Internet Things J 8(6):4291–4304

Huang Y, Kong Q, Jia N, Chen X, Zheng Z (2019) Recommending differentiated code to support smart contract update. In: Proceedings of the 27th international conference on program comprehension, pp 260–270

Ismail H, Germanus D, Suri N (2015) Detecting and mitigating p2p eclipse attacks. In: 2015 IEEE 21st international conference on parallel and distributed systems (ICPADS). IEEE, pp 224–231

Krupp J, Rossow C. (2018) teether: Gnawing at Ethereum to automatically exploit smart contracts. In: 27th {USENIX} security symposium ({USENIX} security 18), pp 1317–1333

Kwon J, Buchman E. (2018) A network of distributed ledgers. Cosmos Dated 1–41

Lei K, Zhang Q, Xu L, Qi Z (2018) Reputation-based byzantine fault-tolerance for consortium blockchain. In: IEEE (ed.) 2018 IEEE 24th international conference on parallel and distributed systems (ICPADS), pp 604–611

Li D (2020) Discussion on block chain ecological construction based on china's independent and controllable basic public block chain. Inf Sec Technol 9(9):6–9

Li P, Xu H (2020) Blockchain user anonymity and traceability technology. J Electron Inf Technol 42(5):1061–1067

Li P, Xu H, Ma T (2021) An efficient identity tracing scheme for blockchain-based systems. Inf Sci 561:130–140

Li P, Xu H, Ma T (2021) Research progress of blockchain privacy protection and supervision technology. J Cyber Sec 6(3):159–168

Liu JK, Wei VK, Wong DS (2004) Linkable spontaneous anonymous group signature for ad hoc groups. In: Australasian conference on information security and privacy. Springer, pp 325–335

Lorenz J, Silva MI, Aparício D, Ascensão JT, Bizarro P (2020) Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity. arXiv:2005.14635

Lu Q, Xu X (2017) Adaptable blockchain-based systems: a case study for product traceability. IEEE Softw 34(6):21–27

Marcus Y, Heilman E, Goldberg S (2018) Low-resource eclipse attacks on Ethereum's peer-to-peer network. IACR Cryptol ePrint Arch 2018:236

Mehar MI, Shier CL, Giambattista A, Gong E, Fletcher G, Sanayhie R, Kim HM, Laskowski M (2019) Understanding a revolutionary and flawed grand

experiment in blockchain: the Dao attack. J Cases Inf Technol (JCIT) 21(1):19–32
Monamo P, Marivate V, Twala B (2016) Unsupervised learning for robust bitcoin fraud detection. In: 2016 information security for South Africa (ISSA). IEEE, pp 129–134
Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system. Decentralized Bus Rev 21260
Narula N, Vasquez W, Virza M (2018) zkledger: privacy-preserving auditing for distributed ledgers. In: 15th {USENIX} symposium on networked systems design and implementation ({NSDI} 18), pp 65–80
Orcutt M (2020) Criminals laundered \$2.8 billion in 2019 using crypto exchanges, finds a new analysis (2020). https://www.technologyreview.com/2020/01/16/130843/cryptocurrency-money-laundering-exchanges/ Accessed 16 January
Ostapowicz M, Żbikowski K (2020) Detecting fraudulent accounts on blockchain: a supervised approach. In: International conference on web information systems engineering. Springer, pp 18–31
Pham T, Lee S (2016) Anomaly detection in bitcoin network using unsupervised learning methods. arXiv:1611.03941
Pongor G (1993) Omnet: objective modular network testbed. In: Proceedings of the international workshop on modeling, analysis, and simulation on computer and telecommunication systems, MASCOTS'93, San Diego, CA, USA. Society for Computer Simulation International, pp 323–326
Radix (2018) What is an eclipse attack? 7 June 2018. https://www.radixdlt.com/post/what-is-an-eclipse-attack
Rubixi (2016) Rubixi smart contract. https://bitcoindtalk.org/index.php?topic=1400536.0 Accessed 14 Mar 2016
Saad M, Njilla L, Kamhoua C, Mohaisen A (2019) Countering selfish mining in blockchains. In: 2019 international conference on computing, networking and communications (ICNC). IEEE, pp 360–364
Sayadi S, Rejeb SB, Choukair Z (2019) Anomaly detection model over blockchain electronic transactions. In: 2019 15th international wireless communications and mobile computing conference (IWCMC). IEEE, pp 895–900
Shen J, Zhou J, Xie Y, Yu S, Xuan Q (2021) Identity inference on blockchain using graph neural. Network 2104:06559
Shen M, Sang A, Zhu L, Sun R, Zhang C (2021) Abnormal transaction behavior recognition based on motivation analysis in blockchain digital currency. Chin J Comput 1:193–208
Sheng P, Wang G, Nayak K, Kannan S, Viswanath P (2021) BFT protocol forensics. In: Proceedings of the 2021 ACM SIGSAC conference on computer and communications security. pp 1722–1743
Shultz BL, Bayer D (2015) Certification of witness: mitigating blockchain fork attacks. Undergraduate Thesis in Mathematics, Columbia University in the City of New York (2015)
Skuchain Forum, W.E.: Inclusive Deployment of Blockchain for Supply Chains (2019). https://www.weforum.org/whitepapers/inclusive-deployment-of-blockchain-for-supply-chains-part-protecting-your-data Accessed 5 June 2019
Su L, Shen X, Du X, Liao X, Wang X, Xing L, Liu B (2021) Evil under the sun: understanding and discovering attacks on Ethereum decentralized applications. In: 30th {USENIX} security symposium ({USENIX} security 21)
Torres CF, Steichen M et al (2019) The art of the scam: demystifying honeypots in Ethereum smart contracts. In: 28th {USENIX} security symposium ({USENIX} Security 19), pp 1591–1607
Vasek M, Moore T (2015) There's no free lunch, even using bitcoin: Tracking the popularity and profits of virtual currency scams. In: Springer (ed.) International conference on financial cryptography and data security, pp 44–61
Voell Z (2020) Ethereum classic hit by third 51% attack in a month. https://www.coindesk.com/ethereum-classic-blockchain-subject-to-yet-another-51-attack Accessed 30 Aug 2020
Weber M, Domeniconi G, Chen J, Weidele DKI, Bellei C, Robinson T, Leiserson CE (2019) Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. arXiv:1908.02591
Wei A (2018) Public blockchain technology and its application value. Internet Econ 7:26–31
Wikipedia (2021a) Money laundering. 8 July 2021. https://en.wikipedia.org/w/index.php?title=Money_laundering&oldid=1032228344
Wikipedia (2021b) Ponzi scheme. https://en.wikipedia.org/w/index.php?title=Ponzi_scheme&oldid=1030419781. Accessed 8 July 2021
Wu G, Yu P, Wang K (2020) Transaction regulatory research on double-chain blockchain. Comput Eng Appl 56:116–123
Wu J, Liu J, Chen W, Huang H, Zheng Z, Zhang Y (2020) Detecting mixing services via mining bitcoin transaction network with hybrid motifs. arXiv:2001.05233
Wu J, Yuan Q, Lin D, You W, Chen W, Chen C, Zheng Z (2020) Who are the phishers? Phishing scam detection on Ethereum via network embedding. IEEE Tran Syst Man Cybern Syst
Xi R, Yun X, Jin S, Zhang Y (2012) Research survey of network security situation awareness. J Comput Appl 32(01):1–4
Xu G, Guo B, Su C, Zheng X, Liang K, Wong DS, Wang H (2020) Am i eclipsed? a smart detector of eclipse attacks for Ethereum. Comput Secur 88:101604
Yang X (2020) Research of blockchain ecology security challenges and solutions. Inf Secur Technol 11(3):50–55
Yang Z, Yang K, Lei L, Zheng K, Leung VC (2018) Blockchain-based decentralized trust management in vehicular networks. IEEE Internet Things J 6(2):1495–1505
Yuan Z, Yuan Q, Wu J (2020)Phishing detection on Ethereum via learning representation of transaction subgraphs. In: Blockchain and trustworthy systems, pp 178–191
Zhang R, Preneel B (2017) Publish or perish: a backward-compatible defense against selfish mining in bitcoin. In: Cryptographers' track at the RSA conference. Springer, pp 277–292
Zhang Y, Lou Y (2021) Deep neural network based Ponzi scheme contract detection method. Comput Sci 48(1):273–279
Zhao G, Xie Z, Wang X, He J, Zhang C, Lin C, Zhou Z, Chen B, Rong C (2020) Contractguard: defend Ethereum smart contract with embedded intrusion detection. Chin J Netw Inf Secur 6(2):35–55
Zheng H, Wu Q, Qin B, Zhong L, He S, Liu J (2018) Linkable group signature for auditing anonymous communication. In: Australasian conference on information security and privacy. Springer, pp 304–321
Zheng P, Zheng Z, Wu J, Dai H-n (2020) Xblock-eth: Extracting and exploring blockchain data from Ethereum. IEEE Open J Comput Soc 1:95–106
Zheng Z, Xie S, Dai H, Wang H (2017) An overview of blockchain technology: architecture, consensus, and future trends. In: 2017 IEEE international congress on big data (BigData Congress). pp 557–564
Zhou Z, Wang M, Yang C-N, Fu Z, Xin S, Wu QJ (2021) Blockchain-based decentralized reputation system in e-commerce environment. Future Gener Comput Syst 124:155–167

Publisher's Note