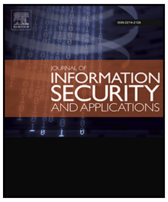
Journal of Information Security and Applications 60 (2021) 102859

Contents lists available at ScienceDirect

Journal of Information Security and Applications

journal homepage: www.elsevier.com/locate/jisa



Survey of security supervision on blockchain from the perspective of technology

Yu Wang, Gaopeng Gou, Chang Liu, Mingxin Cui, Zhen Li, Gang Xiong *

Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Keywords:
Blockchain technology
Security supervision
Peer-to-peer (P2P) network
Transaction
User identity
Heuristics
Graph theory

ABSTRACT

Due to decentralization, immutability, circulation, anonymity, blockchain technology has become a hot topic but also a hotbed of various cyber-crimes. Many perpetrators attack blockchain to steal cryptocurrencies or use anonymous addresses to conduct illicit financial transactions or receive ransoms while hiding their identities. Since blockchain technology has not developed for a long time, the security issues in this area cannot be well resolved through its own mechanisms, which brings great challenges to protect the security of blockchain and users. Although there are some protective measures to prevent attackers from attacking, most of them are proposed after attacks, and it is impossible to find the masterminds behind modern cyber-crimes, so it is necessary to continuously monitor suspicious nodes or users. In this paper, we first present a systematic overview of blockchain technology and security issues according to the four-layer structure, and explain the problem of security supervision of blockchain. Subsequently, we divide the key technologies for security supervision of blockchain into three aspects: node discovery technology on the network layer, data analysis technology of transaction records on the transaction layer, and network traffic analysis technology on the application layer. In terms of each aspect, we summarize the studies from various angles according to its characteristics. In the end, we discuss the relationship between blockchain and traditional law. Moreover, we present the challenges of security supervision and possible future research directions in this field.

1. Introduction

With the rapid development of the e-commerce and online finance industry, billions of network transactions are executed on the Internet every moment [1]. How to ensure the security of these transactions and how to achieve security storage, exchange, sharing of massive transaction data has become urgent problems. Blockchain technology is a practical technology to solve these problems. Because of this fact, blockchain technology is developing rapidly. There are more than 4900 digital cryptocurrencies and more than 1500 tokens based on blockchain technology. Market capitalization of blockchain reaches 266 billion dollars [2]. Gartner's forecast pointed out that transactions based on blockchain will be more than 10 billion dollars by 2022 [3]. According to Digital Marketing Ramblings, the scale of blockchain industry is expected to reach 20 billion dollars by 2024 [4]. And 10% of global GDP will be stored through blockchain technology by 2025. Blockchain is usually regarded as the next generation of the Internet, and it is also a new milestone in the history of human credit development [5,6].

Due to blockchain characteristics, especially anonymity, it attracts many perpetrators, leading to frequent security incidents in recent years. In January 2019, the attackers carried out at least 15 double-spending attacks on Ethereum, causing more than two hundred thousand ETH (i.e., one digital token of Ethereum) to be implicated, which is approximately 1.1 million dollars, while the cost of the attack was only twenty thousand dollars. In March, the DragonEx exchange was hacked and stolen more than 6 million dollars through software containing 0day. In June, the Plustoken wallet was suspected to stop serving users, involving more than 1 million people, and the digital currency in these addresses is worth 400 million dollars [7]. In April 2020, hackers used the compatibility issues between Uniswap and ERC777(Etherum Request for Comments) to attack Lendf.Me(i.e., a decentralized finance loan protocol). They obtained imBTC(i.e., tokenization of Bitcoin assets) from Uniswap through reentrancy attack when conducting ETH-imBTC transactions. Lendf.Me lost a total of twenty four million dollars in digital assets. In July 2020, many influential accounts on Twitter were stolen, and posted phishing information related to Bitcoin, including Bill Gates, Jeff Bezos, Obama, Biden, etc. On July 16th, 366 transactions were transferred to hacker's blockchain addresses [8]. From January 2018 to December 2020, criminal activities have caused about 18 billion dollars in economic losses [9].

* Corresponding author at: Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China.
E-mail address: xionggang@iie.ac.cn (G. Xiong).

https://doi.org/10.1016/j.jisa.2021.102859

Available online 25 May 2021

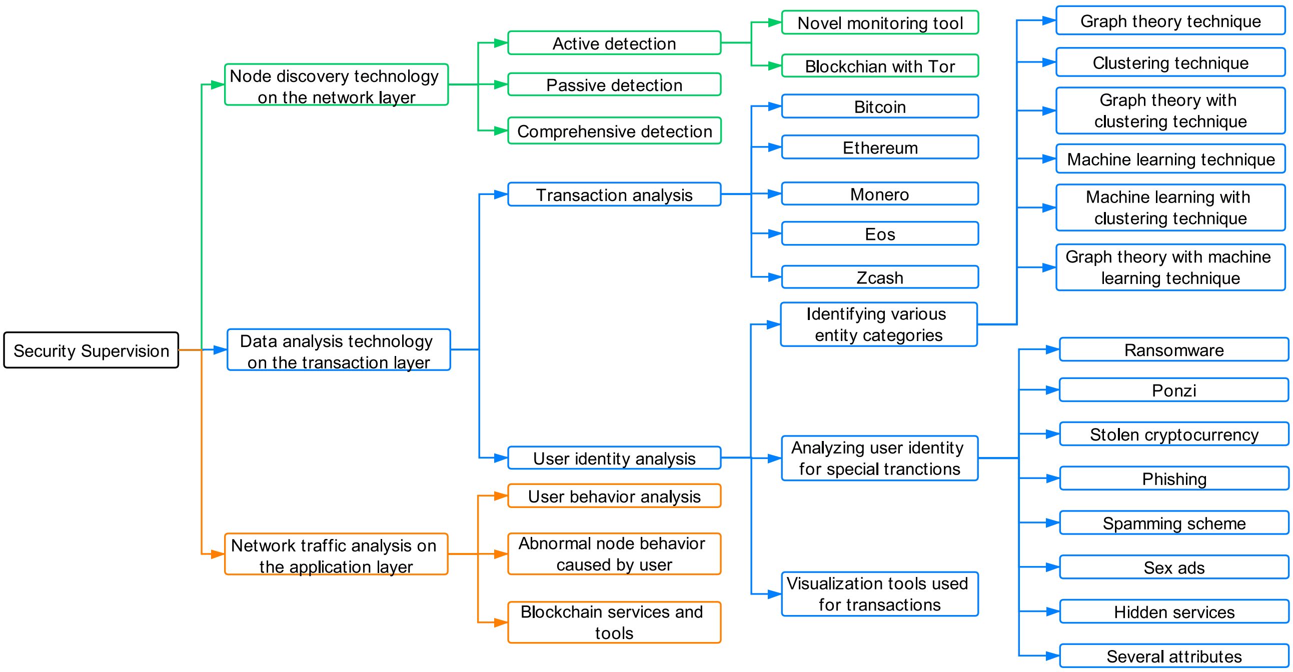

Security Supervision
Node discovery technology on the network layer
Data analysis technology on the transaction layer
Network traffic analysis on the application layer
Active detection
Passive detection
Comprehensive detection
Transaction analysis
User identity analysis
User behavior analysis
Abnormal node behavior caused by user
Blockchain services and tools
Novel monitoring tool
Blockchian with Tor
Bitcoin
Ethereum
Monero
Eos
Zcash
Identifying various entity categories
Analyzing user identity for special tranctions
Visualization tools used for transactions
Graph theory technique
Clustering technique
Graph theory with clustering technique
Machine learning technique
Machine learning with clustering technique
Graph theory with machine learning technique
Ransomware
Ponzi
Stolen cryptocurrency
Phishing
Spamming scheme
Sex ads
Hidden services
Several attributes

Fig. 1. Structure of security supervision on Blockchain from the Perspective of Technology.

Although there are some protection methods for these attacks, it is unable to find the masterminds behind modern cyber-crimes. Many criminals use blockchain as a tool for receiving the ransom, decentrally transfer and store funds, then launder money in batches, so protection methods do not work. And blockchain uses special consensus mechanisms and communication mechanisms, which makes it impossible to obtain identities and information of real users directly. Therefore, blockchain technology has attracted attention from researchers in the area of supervision and governance. Researchers have two attitudes towards blockchain supervision. On the one hand, the stream of research studies that involve money laundering and digital crime argues for establishing stringent supervision and guidance frameworks [10–12]. On the other hand, some research argues for more open-minded supervision in order to ensure the rapid development of blockchain [13–15]. We will not detail these studies, because we focus on the technical level.

The technical aspect of blockchain security supervision is mainly to detect attacks or abnormal behaviors on blockchain in time and find criminals in the anonymous network of blockchain. As Fig. 1 shows, the security supervision of blockchain refers to three aspects: the first is nodes detection for the network layer, the second is data analysis of transaction records for the transaction layer, and the last is network traffic analysis for the application layer. Combining three parts of technologies may help to find and trace malicious nodes, dubious transactions and abnormal users, thereby improving blockchain security.

In terms of nodes detection for the network layer, the regulator can obtain various information of abnormal nodes or the whole network topology, such as, the geographic location [16–18], online regular [19], behaviors [20–23], etc. There are three technologies: active detection, passive monitoring and comprehensive monitoring. Active detection methods are defined as constructing a series of specific probe packets sent to neighbor nodes and iterating the process to infer the relevant information about abnormal nodes, which are characterized by strong targeting. Passive monitoring does not require interaction with other nodes, only capture all the traffic sent and received. Another passive method refers to capture traffic packets by extracting and utilizing features in the network packets [24]. Using passive methods can avoid the extra network burden and detect nodes behind network. Comprehensive monitoring method combines active detection and passive monitoring. Some malicious nodes may be discovered through analysis in the network layer, but the other nodes belonging to the same malicious entity are hiding in the network, and they are secretly trading. Finding all addresses of abnormal nodes under malicious entities cannot be achieved only through detection in the network layer, it needs to be combined with analysis in the transaction layer.

In terms of data analysis of transaction records for the transaction layer, since all historical transactions are recorded on blockchain, including illegal transactions, ransom, stolen currencies, etc., many researchers utilized graph-theoretical methods to construct transaction graph and entity graphs, so that they can analyze currency flow of transactions [25,26], the details of mining pool behaviors [27]. By extracting features from the graphs and using clustering algorithms, abnormal behaviors of users and transactions can be detected [28]. Due to the emergence of smart contract and internal transactions in Ethereum, contract creation graph (CCG) and contract invocation graph (CIG) are proposed to analyze the relationships between contracts (i.e., smart contract creation, smart contract invocation) [29,30]. In order to capture more details of the graphs, the authors use time window for dynamic analysis [31]. As for Monero, it aims to solve the linkability and traceability issues in Bitcoin through three central methods: Stealth addresses, Ring Signatures, and Confidential transactions, which bring great challenges for tracing malicious transactions. For these unique problems, some new heuristics for Monero are used to reduce the size of anonymity sets of malicious transactions [32–35]. Since EOSIO(i.e., an open-source blockchain platform) has high throughput by using DPoS(Delegated Proof-of-Stake, up to 8000 transactions per second), the transaction volume increases rapidly. The graphs constructed by transactions and addresses have fewer details, so time information is added to the graph [36,37]. As for Zcash, its anonymity relies on the shielded pool, in which users can spend ZEC(i.e., one digital token of Zcash) without revealing the details of the transactions. Some researchers focus on applicable clustering methods for Zcash according to the mechanism of transaction [38–41] and block reward [42]. The above studies analyze all historical transactions, and some research aims to track the currency flow and extend addresses of illegal events appeared in the history of blockchain, such as CryptoLocker [43], Silk Road [44], Mt.Cox [45], Ponzi scheme [46], phishing scams [47]. In order to have a more intuitive view of malicious events, [48] and [49] design tools for dynamically visualizing.

However, it is possible to obtain the capital flow of an illegal transaction and related blockchain addresses of a suspicious entity, but it cannot associate abnormal addresses or entities with real users. So the identities of blockchain addresses on public occasions (e.g., forums, donation sites, social networks, tor market, dark web, exchanges, etc.) are utilized to analyze user identities. Only a small portion of tagged blockchain addresses can be crawled from these websites, it is necessary to use clustering algorithms for all transaction addresses to expand the address set of malicious or illegal entities, thus making the experiment more convincing. Users' habits of using blockchain can be obtained by analyzing the blockchain community [50]. By simulating the usage of Bitcoin in universities, and clustering through user behaviors, it is proven that 80% of user identities can be associated with Bitcoin addresses [51]. In [52], the authors use thirteen classifiers trained by 12 types of observations to classify 10000 unknown observations. Some researchers concentrate on analyzing illegal events by new heuristics or novel detection methods, such as, pseudo-spam transactions [53], sex ads [54], high yield investment programs [55,56], tor hidden services [57], the dark web [58,59], ransomware [60], fraudulent accounts [61]. The yet-unidentified entities are predicted by supervised machine learning methods [62,63]. Jourdan et al. utilized a discrete-time graph to analyze leakage of information about entities, which is obtained from the entity graph through temporal aggregation [64]. In [30,65], the concept of motifs is introduced into the analysis of blockchain anonymity. But most of research in the transaction layer is to crawl the relevant malicious addresses for the illegal incidents that have occurred, then expand the size of the address set. The abnormal accounts and addresses obtained by crawling are always limited. Therefore, it is necessary to combine the analysis methods of the application layer to detect abnormal users and user behaviors.

In terms of network traffic analysis technology for the application layer, abnormal users' IP(Internet Protocol) addresses can be obtained more intuitively. Sequence data is extracted from the activities of the peers, and using deep learning to distinguish normal peers and abnormal peers [66]. But there are millions of peers in public blockchain, which request a tremendous amount of effort. A novel algorithm termed Behavior Pattern Clustering (BPC) can automatically cluster the behaviors of all peers into categories, thereby reducing resource consumption [67]. By using the message propagation mechanics of blockchain, transactions issued from a device can be clustered through network traffic analysis [68]. Blockchain technology can construct programmable cryptocurrency and decentralized applications through smart contracts, thus forming a fully functional decentralized world, which has a profound impact on the Internet, finance and other fields. But due to the vulnerability of smart contracts, used by attackers, blockchain has a large economic loss every year. Classification on DApps(Decentralized Applications) and DApps user behaviors may help to detect suspicious behaviors and assist in the hunt of attackers by monitoring the potential DApps users [69,70]. Token is also an important tool in blockchain, inconsistent token behaviors lead to user confusion and financial loss, different recognition methods of token behaviors are proposed [71,72].

Due to the importance of blockchain, a few surveys have been published in related topics, such as blockchain architecture and technological challenges [73,74], consensus algorithms [74–76], the body of knowledge of smart contracts [77], privacy and anonymity [78–80], and applications of blockchain [73,81,82]. With more blockchain security issues appearing, there also exist a number of surveys that cover the blockchain security threats along with their cause and the proposed countermeasures [83–86], including smart contracts security [87], data security [88], network security [89]. The survey in [84] mainly attributed attack viability in the attack surface to three aspects and outlined several effective defense measures. In [87], Atzei et al. studied the security vulnerabilities of smart contracts and summarized a series of attacks which exploit these vulnerabilities. Zhu et al. [88] presented a comprehensive classification and summary of the security of blockchain data from some aspects, including privacy, availability, integrity and controllability. Similarly, Anita et al. [90] presented taxonomy of security threats and the methodologies to mitigate them in detail. Zaghloul et al. [89] also investigated the major security threats and countermeasures of Bitcoin, and analyzed network security risks. Wang et al. [85] systematically analyzed the security of six layers of blockchain and possible cyber attacks. These survey articles do not pay attention to supervise the masterminds behind cyber-crimes. In this work, we aim to present a comprehensive review of state-of-the art studies on the security supervision on blockchain from the perspective of technology. As shown in Fig. 1, we categorize the existing techniques and results into three main parts: (*i*)node discovery, (*ii*)data analysis and (*iii*)network traffic analysis. Specifically, the key contributions of this paper are listed as follows:

(1) To the best of our knowledge, we are the first to conduct a comprehensive survey on the security supervision of blockchain from the perspective of technology, including nodes detection technology on the network layer, data analysis technology of transaction records on the transaction layer and network traffic analysis technology on the application layer.
(2) According to the technology type of node detection, we divide the studies into three categories: active detection, passive detection and comprehensive detection.
(3) As for the transaction layer, we provide an in-depth review on the data analysis technology with a proposed three-layer structure. The first layer is based on different targets of data analysis, the related studies are divided into two categories: transaction analysis and user identify analysis. Furthermore, the second layer is summarized according to the type of blockchain platform and the scope of the entity involved. The third layer is divided into various aspects based on the types of techniques (e.g., graph theory, clustering, machine learning) and entities (e.g., ransomware, sex ads, ponzi).
(4) We review the studies on the network traffic analysis, including user behavior analysis, abnormal node behavior caused by the user, blockchain services and tools, which makes up for deficiencies in the lack of the association between user and blockchain addresses.

The rest of the paper is organized as follows. Section 2 gives a systematic review of blockchain technology and security issues, and presents the security supervision of blockchain. Section 3 details the node discovery technology on the network layer. Section 4 provides an in-depth survey of the approaches of data analysis on the transaction layer. Section 5 reviews the studies of network traffic analysis on the application layer. In Section 6, we discuss the relationship between traditional law and blockchain, challenges of blockchain supervision and suggest future research directions. Lastly, Section 7 concludes the paper.

2. Background of security supervision of blockchain

2.1. Overview of blockchain technology

In 2008, the concept of blockchain was first proposed in a paper published by Satoshi Nakamoto [91], and blockchain is called “a chain of block” in the paper, which highlights the essence of blockchain. In 2009, Bitcoin, the first open-source blockchain project [92], became the first decentralized cryptocurrency. The relevant concepts (e.g., proof-of-work, distributed ledger) have been verified their effectiveness in real-world practice.

The significant interest in blockchain technology can be attributed to two reasons: the first is its characteristics such as decentralization, immutability, liquidity, openness, anonymity, transparency, security; The second is the massive user-base of cryptocurrencies. Blockchain

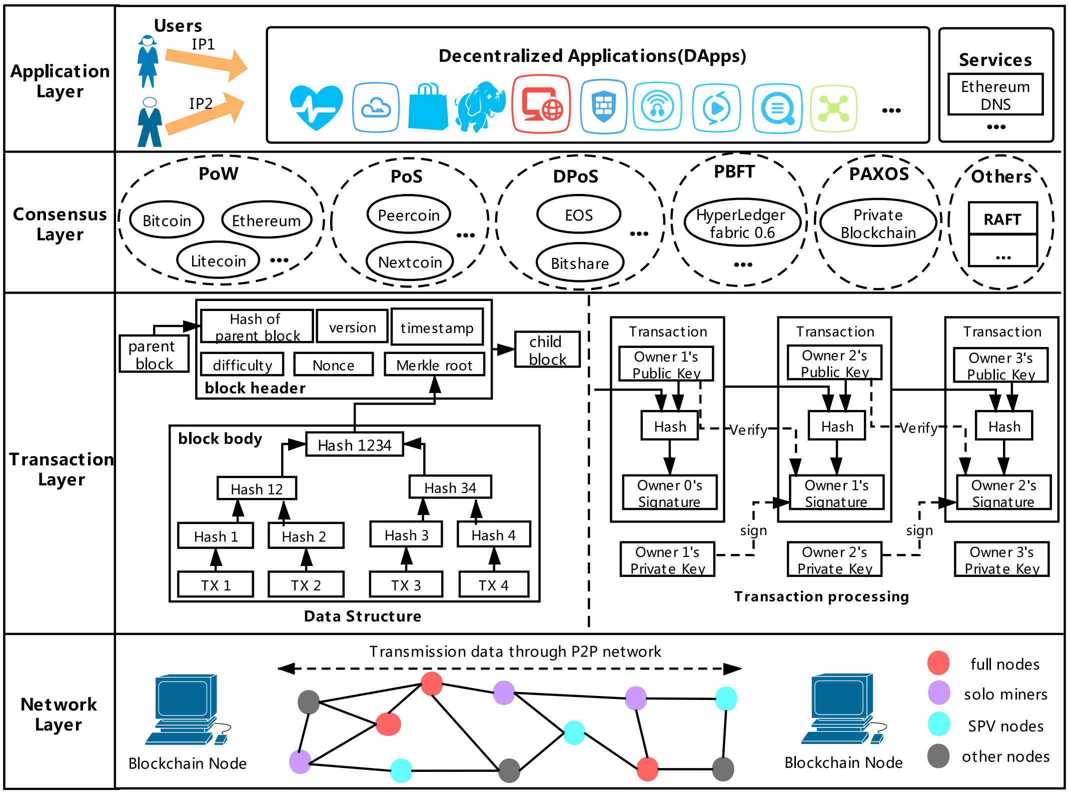


Fig. 2. The architecture of blockchain technology.

is also gradually increasing its functions, including cryptocurrencies, smart contracts [93], decentralized applications [94]. Under the premise of mutual distrust between nodes, blockchain technology guarantees the security, integrity, and non-tampering of transaction data through digital signature, consensus mechanism, timestamp, and block reward mechanism. Although these technologies are not completely secure or unassailable, they demonstrate resilience and functionality. Due to the long-term autonomous operation of blockchain and supporting global transactions, which are instant, efficient, and reliable, so the application scenarios are extensive, covering the fields of finance, education, network, privacy, medical treatment, and trade management. As of March 2020, there are more than three thousand decentralized applications [95], which are mainly deployed on blockchain platforms such as Ethereum [96], EOS [97], and Steem [98]. Most of people needs can be met in the world of DApps.

Bitcoin is one of the most typical architecture in all blockchains. Blockchain is mainly composed of the network layer, transaction layer, consensus layer, incentive layer, contract layer and application layer [99]. However, with the development of blockchain technology, many modules are weakened or no longer used. Since the private blockchain and consortium blockchain are both internal chains, the incentive layer is weakened. Therefore, considering the nature of blockchain and the future trend of blockchain development, we subdivide the blockchain architecture into four layers: the network layer, transaction layer, consensus layer, and application layer. The structure diagram is shown in Fig. 2.

The problem of network security is always the main theme of information society. With the development and application of blockchain technology, the number of security incidents is increasing. In 2018, the economic losses caused by security incidents peak at approximately 2.2 billion dollars [100]. In the following subsections, we will introduce blockchain technology and summarize the security issues of blockchain according to the four-layer structure mentioned above. Attack methods are shown in Fig. 3.

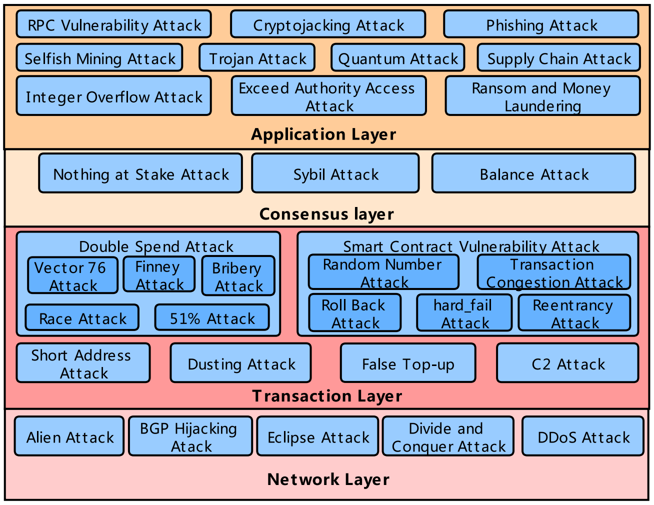


Fig. 3. Attack methods on different layers of blockchain.

2.2. Network layer

The network layer implements distributed networking mechanisms through P2P technologies such as P2P networking and node communication. Blockchain is essentially a P2P network, so there is no need to utilize centralized servers to store data and manipulate the system. Blockchain is jointly maintained by all nodes, the users both use and provide the foundation of the network, which is entirely voluntary. All resources and services in the network are evenly distributed to each node. Data can be directly transmitted between nodes without the intervention of third parties [101].

2.2.1. P2P network topology

Blockchain uses a P2P network, which has different structures. According to whether networks are decentralized and whether node addresses are structured, P2P networks can be divided into the following four types:

2.2.1.1. Centralized topology. The central server of the network structure is used to store address information of the access nodes. Local nodes obtain address information of other nodes by connecting the central server, thus realizing data transmission between nodes [102].

2.2.1.2. Decentralized unstructured topology. This structure has no central server. Nodes can join and leave the network freely. There is no unified standard of structure for node addresses [102,103]. The structure of the network is a random graph. Bitcoin uses this network structure and finds new nodes through the flooding mechanism for networking, which is high consumption.

2.2.1.3. Decentralized structured topology. This network structure uses a distributed hash table (DHT). By using hash functions, addresses are standardized to the uniform length. Each node stores the hashes of neighbor nodes in local files [102,104]. Each node in Ethereum has a unique identifier, called node ID, which is a 512-bit cryptographic ECDSA(Elliptic Curve Digital Signature Algorithm) public key, and stores neighbor nodes' information such as NodeID and IP address through a DHT which is a modified implementation of Kademlia [105]. Kademlia is a UDP-based (UDP, User Datagram Protocol) protocol for distributed nodes to store and retrieve content. DHT is a K-bucket structure, and the K-bucket is refreshed regularly [23]. The logical distance between two nodes based on the XOR (exclusive OR) of Keccak-256 hashes of two nodes IDs. Now, Ethereum uses Node Discovery Protocol version 4 and version 5 to networking, and we will introduce them in the following. RLPx (i.e., the network protocol of Ethereum) is used to encrypt the communication among nodes, which makes Ethereum more secure.

2.2.1.4. Partially decentralized topology. This structure combines the advantages of the above structures, and divides nodes into ordinary nodes and super nodes. Super nodes maintain part of network nodes addresses, indexes, etc. Their functions are similar to the central server. EOS(Enterprise Operation System) uses the structure. There are 21 super nodes (i.e., block producers) in EOS blockchain [36]. When they cannot perform their duties in time, they will be removed and replaced by new super nodes.

2.2.2. Networking process

When a new node joins the network, different blockchains have different ways to obtain the addresses of neighbor nodes. The networking processes of popular blockchain are as follows:

2.2.2.1. Bitcoin. The seed nodes are hard-coded in the source code to provide the information of the initial access nodes. After connecting the network, the new node broadcasts its own information to other neighbor nodes, or actively obtains addresses of neighbor nodes through Net.GetAddresses() method [106]. The node repeats this process until the number of outgoing connections reaches eight, the default maximum, and the maximum number of incoming connections is 117. In order to avoid the limitation of network bandwidth and the number of seed node connections, new nodes obtain address information of other nodes by broadcasting, and they save information in a local file and periodically send ping-pong messages to update the local file.

2.2.2.2. Ethereum. Similar to Bitcoin, new nodes join Ethereum network through hard-coded seed nodes. Users also can download files containing stable nodes and import them into the local file, such as Mars nodes of EthFans [107]. Findnode and neighbors messages are the request and reply of node discovery. The new node sends Findnode message for requiring node information, and the recipient replies neighbors messages, which consist of the closest 16 nodes to the target, and the information is found in its local routing table. Ping and pong messages are used to verify that the neighbor node is alive. Ethereum nodes use two structures to store the information of neighbor nodes: one is called DB which stores the information into the disk, the other is called table (also known as routing table) which stores it into memory. Due to the Kademlia protocol, users can query addresses to find random nodes or specific nodes. Besides, the information in each table includes node ID, IP, TCP(Transmission Control Protocol) port and the time last seen.

2.2.2.3. EOS. A new node connects EOS through seed nodes, which are contained in the configuration file. Because of the existence of super nodes, there is no need to save historical neighbor nodes. The information of neighbor nodes can be obtained from super nodes. The new node randomly selects some neighbor nodes, and requests their node lists. Then it synchronizes itself nodes list. Finally, the new node broadcasts its own node information to the entire network.

Because blockchain nodes are distributed all over the world, they can join and exit freely, so blockchain has good liquidity. Until now, Bitcoin's network is the largest in all blockchains, about 10300 available nodes per day [108]; Ethereum is the second, about 7486 available nodes per day [109]. Due to the large block data, some nodes are still in the stage of synchronizing blockchain data.

2.2.3. Communication process

The process of nodes communication can be divided into three parts: maintaining connection, synchronizing data, and forwarding transaction information:

2.2.3.1. Maintain connection. When the number of neighbor nodes reaches the maximum, they periodically confirm the neighbor nodes whether they are reachable through the ping and pong messages [23]. Then the unreachable nodes are replaced by new reachable nodes.

2.2.3.2. Synchronize data. After the new node joins the network, it needs to synchronize the block header if it is a light node. If it is a full node, it needs to synchronize the data of all historical block, from the genesis block to the highest block. Different blockchain platforms use diverse block synchronization methods. In Bitcoin, the nodes use Inv, GetBlocks, GetData, SendHeaders and other messages. In Ethereum, the nodes use GetBlockHeaders, MaxHeaderFetch messages, etc [83,105].

2.2.3.3. Forward transaction information. The initiating node of a transaction needs to broadcast the transaction to other nodes for verification. First, the node compares the block height with neighbor nodes, and requests or reverse requests blocks to reach the same height. Then the initiating node forwards the transaction to the neighbor node. The neighbor node also compares the block height with its neighbors, forwards the transaction, then neighbors repeat the process. Finally, the transaction is spread throughout the blockchain by recursive flooding. All nodes are constantly exchanging block information with neighbor nodes to ensure that the block information of each node is the same.

2.2.4. Security issues

2.2.4.1. Alien attack. This attack is also called pollution of the address pool. Because the same type of blockchain system does not recognize the identity of nodes, which induces nodes of the same type blockchain to invade and pollute each other. Attackers collect addresses of Ethereum nodes and handshake in a malicious way. They achieve the purpose of polluting the address pool through node handshake. Nodes of different chains interact with each other, and push their address pool to other blockchain nodes. The victim node eventually broadcasts its address pool to the entire network, causing node congestion, abnormal behaviors of the main network and other phenomena.

Table 1
Characteristics of the two accounting models.

Accounting model	Advantage	Disadvantage
UTXO	1. High security. Each transaction put the change [110] to a new address, and all inputs of one transaction belong to a user. So it is easy to verify transaction sequence and connection relationship. 2. More privacy. Users can use new addresses for transactions, the relation between new addresses and previous addresses is hard to track. 3. More efficiency. It can be processed concurrently.	1. The transaction involves multiple UTXO, which need to be signed separately, resulting in inefficiency. 2. This model cannot implement complex logic, and has poor programmability.
Account model	1. Smart contracts are programmable scripts, which is conducive to develop. 2. Batched transaction is convenient, and reduces the consumption of resource. 3. The verification of light node is simple.	1. There is no dependency between accounts, so the replay problem needs to be solved by nonce. 2. It is not conducive for forensics of Internet.

2.2.4.2. BGP hijacking attack. Border Gateway Protocol (BGP) is a routing protocol. BGP is used to exchange network reachability information with other BGP systems and regulates the route of IP packets to the destination. The attacker may leverage BGP routing to intercept the blockchain network traffic, thereby splitting the blockchain network. Maria et al. analyzed the impact of routing attacks from two aspects, node-level and network-level attacks [111,112]. They found that the distribution of mining power is an important factor in how many internet prefixes can be hijacked. In 2014, the attacker stole about 83 thousand dollars of cryptocurrency by rerouting traffic to a mining pool controlled by the attacker [113]. Network operators need to rely on monitoring systems to receive the rogue announcements [114].

2.2.4.3. Eclipse attack [115]. The attacker may monopolize all of the victim's connections, including incoming and outgoing, thereby isolating the victim from other nodes [116]. In other words, they utilize many malicious nodes to send requests for connection to the victim's node so that reach the node's maximum links, and keep the node in an isolated network. This attack causes the victim node unable to connect with other nodes and accept requests from other normal nodes. A type of node denial of service attack.

2.2.4.4. Divide and conquer attack. The blockchain network is divided into several communities with similar computing power. Each community mines blocks under different subtrees. The attacker hides the block mined by all computing power of the attacker, and make it become the heaviest branch. Repeating this process, the attacker's hidden chain will be longer, and the heaviest attacker's hidden chain will become the main chain.

2.2.4.5. DDoS attack. The DDoS (Distributed Denial of Service) attack in blockchain cause miners and nodes to spend more time waiting to process blocks. For instance, attackers may initiate low gas price opcode (e.g., reading state information stored on a disk) many times per block, which lead to the network speed slowing down drastically. Creating a large number of empty accounts causes issue—waste of hard drive space, thus increasing sync time. According to [117], 74% of Bitcoin-related sites have suffered DDoS attacks.

2.3. Transaction layer

All blocks are generated to store transactions. The transaction layer is the core of blockchain. It guarantees the stable and reliable transmission of valuable data between two nodes. The block mainly includes previous block hash, version, merkle root hash, timestamp, difficulty, nonce and transactions [74]. The first transaction in a block is called coinbase transaction [118], which has exactly one output. Other transactions mainly include the number of inputs, transaction inputs, the number of outputs, transaction outputs, timestamp, block number and etc.

Miners link the latest verified block to the current blockchain. A successful miner obtains transaction fees and mining rewards [119]. Each node has a complete historical record of blockchain transactions, so it can provide positioning and traceability of blockchain data. The writing time of the block is recorded in the block header, thus this timestamp can be used to correlate the timestamp of blockchain network traffic, and combined with the historical transactions to reproduce the history, which provides possibilities for the supervision of blockchain.

2.3.1. Accounting model

With the improvement of blockchain technology, Bitcoin uses a UTXO-based (UTXO, Unspent Transaction Outputs) accounting model, while Ethereum uses an account-based accounting model. The characteristics of the two accounting models are shown in Table 1.

2.3.1.1. UTXO. The status of ledger is composed of a series of "valid outputs" [120]. Each output is owned by a corresponding private key and has its own value. The "effective output" must meet three conditions: each referenced input must be valid and has not been used; The transaction signature should match the signature of each input owner; Due to the transaction fee, the sum of input values should not less than the sum of output values [81]. Bitcoin uses scripts to implement complex transactions, such as delayed transactions and conditional transactions.

2.3.1.2. Account model. It is very similar to the account model of bank. Each Ethereum account has its own balance. As long as the balance of the transaction initiator is more than the transaction fee, the transaction is valid. Ethereum allows users to deploy a piece of code in Ethereum Virtual Machine (EVM) on their nodes, the code is called smart contract [93], which is Turing-complete. Smart contract adds a layer of logic and computation to the trust infrastructure [77]. When conditions of the smart contract are triggered, transactions are automatically executed. Thereby, smart contract provides more services. There are two types of accounts in Ethereum, Externally Owned Account (EOA) and Contract Account (CA). EOA is owned by blockchain users. Only CA has executable code and they can be created by an EOA [81,121].

Regardless of whether it is Bitcoin, Ethereum or EOS, all transactions will be recorded into the global ledger, which consists of blocks. Contracts (e.g., scripts, smart contracts) and parameter information are also stored in the ledger. They are connected according to the written time and jointly maintained by all nodes.

2.3.2. Security issues

2.3.2.1. Double spend attack. Double spend attack, also known as double payment, uses digital characteristics of currency to complete two or more different transactions by using the same input amount [122,123]. No new token is generated, but the used money can be recovered again, which can also be understood as the token is transferred to another address, then the attacker uses attack method to roll back the transaction. Double attacks can be divided into the following types.

- *Race Attack:* By controlling miner's fee to achieve double spending [124]. The attacker sends two transactions, which are sent to himself and the merchant. The first transaction contains a higher miner's fee, so this transaction has a higher probability of being packaged into the new block. Therefore, this transaction takes priority over the transaction sent to the merchant, and the transaction sent to the merchant will be rolled back [89].

- *Finney Attack:* By controlling the broadcast time of the block to achieve double spending [124]. The attack target is the merchant which accepts zero-confirmation transactions. When the attacker mines the newest block, which contains the transaction from A to B. Address A and B both belong to the attacker, but the attacker does not broadcast this block, and immediately finds a merchant that accepts zero-confirmation transactions. The attacker buys an item to send a transaction to the merchant. When the transaction is broadcast, the attacker broadcasts the previous block. Since the time of this block is prior to the transaction sent to the merchant, double spending of the same token is achieved [89].
- *Vector 76 Attack:* This attack is also known as "one confirmation attack" [125], combining race attack and finney attack. The attacker creates two nodes, node 1 only connects the merchant node, and node 2 connects other nodes in the network. The attacker uses the same token to initiate two transactions. Transaction 1 is sent to the merchant, transaction 2 is sent to the address of his wallet, and a higher miner fee is added to transaction 2. The attacker starts mining transaction 1. When the block is mined, it is not broadcast. Instead, node 1 sends transaction 1 , and node 2 sends transaction 2. However, due to different connected nodes, node 2 connects more nodes and is more likely to be recognized by the network. When transaction 2 is considered valid, the mined block is broadcast immediately. The merchant thinks that the transaction is successful because it pays after receiving a confirmation. Thereby, the attacker cashes out and transfers assets. Because the branch chain of transaction 2 is longer, transaction 1 is rolled back, and the transfer has already been withdrawn. Double spending is achieved [89].
- *51% Attack:* The attacker occupies more than 50% computing power of the entire blockchain network [83,126]. During the period of controlling computing power, the attacker can create a new chain with a height higher than the original chain, so old transactions are rolled back. Most of them realize 51% attack by leasing a large amount of computing power.
  Yang et al. [127] presented a protocol to alleviate 51% attack, combining history weighted information of miners with the total calculation difficulty, named as historical weighted difficulty based Proof-of-Work (HWD-PoW), which increases the cost of the attack by two orders of magnitude. However, there is a assumption in this work that in an honest blockchain branch, miners of new blocks would most likely be the miners who mined the previous blocks. But the new protocol gives old miners a lot of power, which may easily lead to super nodes and reduce the efficiency of transactions mined to blockchain.
- *Bribery Attack [128]:* Even with low hash power, attackers can still bribe other miners to let them work on the existing blockchain, then generating a fork and continue to work on the shorter chain, thereby achieving the purpose of "reverse" transactions. The goal of this attack is to increase the probability of double-spending.

2.3.2.2. Smart contract vulnerability attack.

- *Roll Back Attack [129]:* This attack is commonly used in games or gambling on blockchain. When there is an inline transaction (e.g., user's betting action and contract lottery are in one transaction), the attacker uses the principle of detecting the status of smart contracts during the transaction to get the lottery information. Then the attacker decides whether to roll back the betting transaction according to the lottery information. At present, 12 DApps have suffered this attack.
- *Transaction Congestion Attack:* The attacker can send a large number of deferred transactions before the deferred lottery transaction of smart contract through some methods [36,130], maliciously invading the CPU(Central Processing Unit) resources in the block, so that the deferred lottery transaction cannot be executed due to insufficient resources and needs to wait for the next block [131]. Although the execution results of the same deferred lottery transaction in different blocks are different. The attacker can know whether to win the prize, and he can send a large number of deferred transactions to make it re-draw. In 2019, 22 guessing DApps lost a lot of money because of the attack.
- *Random Number Attack [87]:* The attack targets the nonce generation algorithm of smart contract because many projects use the information on blockchain as the nonce seed of smart contract [132]. When the attacker obtains the nonce generation algorithm through prediction, reverse, etc., he can predict the upcoming nonce of the project according to the nonce generation algorithm. In November 2018, the attacker launched a continuous random number attack on EOS.WIN and obtained 20,000 EOS [133].
- *Hard_fail Attack:* Hard_fail means that the system has an error but does not use an error handler to handle the error. In EOS, the status of the transaction are divided into five types: executed, soft_fail, hard_fail, delayed and expired [134]. Most of the status of transactions observed by users are executed and delayed. Users think that there is no failed transaction because they are not familiar with the transaction status. Therefore, when designing a smart contract, they do not check the transaction status. The attacker uses this bug to construct a hard_fail transaction, deceiving the on-chain game or exchanges for fake recharge [132]. In March 2019, after attacking Vegas town DApp by this method, the attacker attacked more DApps.
- *Reentrancy Attack [87]:* The most classic reentrancy attack is the DAO(Decentralized Autonomous Organization) attack on Ethereum, which led to the original Ethereum fork. The original Ethereum was divided into Ethereum Classic (ETC) and Ethereum (ETH). The attacker also stole about 60 million dollars. The transfer model adopted by the developer first sends the transaction to the user, then it modifies the user balance. The smart contract designed by the malicious designer calls the transfer function again while accepting the transaction, so that the user's balance status does not change, but the attacker keeps withdrawing developer currency until developer currency is none [88,135].

2.3.2.3. Short address attack. Short address attack uses the automatic completion mechanism of input bytecode in EVM to attack Ethereum. When calling the transfer function of ERC20 to transfer ETH, if there is one or more zero after the address provided by the attacker, the attacker can omit the zero behind the address and provide the missing address. If the address is a missing address, the encoded data is 134 bytes, which is 2 bytes less than the normal data. EVM will add zero at the end of encoded data, and the missing zero in the address segment is filled with zero in the data segment, then the missing zero in the data segment is filled by EVM. So the amount of transaction becomes the nth power of the original amount * 16, where n is the number of missing zero. As a result, the attacker steals assets of exchanges and wallets.

2.3.2.4. Dusting attack. Dust means that the transaction amount is very small compared to normal transactions, and many currency holders tend to ignore these balances [136]. Malicious users use the fact that the account model in Bitcoin is the UTXO model. They send dusty transactions to a large number of accounts to make the environment dusty. Through these transactions, the attacker associates other addresses of the account and analyzes the user in real-world, thereby destroying Blockchain anonymity [41]. Due to the increase in transactions and handling fees, the system efficiency decreases.

Table 2
Comparison of different consensus mechanisms.

Property	PoW	PoS	DPoS	PBFT
Consensus confirmation time	Long	Long	Very short	Short
Decentralization degree	High	High	Medium	Medium
Energy consumption	High	low	very low	very low
Security	High	Medium	Medium	Medium
Transactions per second	<20	<20	<500	<1000
Scalability	Strong	Strong	Strong	Weak
Forking	Low probably	Almost impossible	If less than one third nodes are byzantine	Probably
Efficiency	Low	High	High	High
Throughput (TPS) [Confirmation latency]	Sub-ten to tens [6m–60m]	Tens [10m–60m]	Thousands [<1s]	Thousands [1s–1m]
Fault tolerance	50% computing power	50% deposited stake value	33% delegates	33% servers
Information propagation	Gossiping (some via secure channels)	Gossiping	Broadcast among delegates	Broadcast
Representative blockchain	Bitcoin, Ethereum	Peercoin, Nextcoin	Eos, Tron	Hyperledger Fabric [137]

2.3.2.5. False top-up attack. False Top-up attack can be divided into two types: against exchanges and against smart contracts [138,139]. During the process of the attack, no real token is received, but the user gets a real recharge record. Therefore, attackers can steal real money from exchanges or smart contracts through non-existent money without real recharge. If exchanges or smart contracts do not strictly verify whether the source of currencies come from the official, the false top-up attack will occur, such as EOS tokens [134]. If the token-id is not verified, anyone can issue a TRC10(Tron Request for Comments) token with 1024 TRX(i.e., one digital token of Tron) to fake recharge.

2.3.2.6. C2 attack. C2 refers to command execution and control. Because blockchain provides a natural and non-tamperable database, attackers write attacking payload into transactions, such as politically sensitive topics, virus signatures, etc., and permanently save them in the blockchain database by sending commands. Even if they are found by other people, no one has the ability to modify them, such as “Peking University Open Letter Incident” and “Vaccine Incident”.

2.4. Consensus layer

The purpose of this layer is to allow highly decentralized nodes to reach consensus efficiently on the validity of block data in the decentralized network. This layer is the cornerstone of blockchain. Because different nodes have different capabilities of processing, consensus algorithms should have good performance of fault-tolerance. When a new block is added to the blockchain, only one node has the function of accounting, and other nodes need to verify and broadcast the block. Different consensus algorithms are used in different blockchains. For example, Bitcoin and Ethereum use PoW, Peercoin [140] and Nextcoin [141] use PoS(Proof-of-Stake), EOS and Tron [142] use DPoS, Paxos is more suitable for the Private Blockchain [75]. A comparison of various consensus mechanisms is shown in Table 2.

2.4.1. Consensus protocols

2.4.1.1. PoW [143]. It decides the node that will obtain the right to add the new block to the current blockchain by providing the sufficient proof of its workload, in other words, the node determines a proper nonce that results in a block hash starting with a certain number of zeros. In a period of time, the difficulty of calculating the nonce is constant and cannot be arbitrarily small. If the node pays the largest workload and the node is the first to find the nonce, the transaction is packaged into the newest block by this node and recorded in the ledger. This process is called mining, and the nodes that participate in this process are called miners [81]. At present, the accounting nodes are almost monopolized by nodes with strong computing power, such as mining pools, therefore PoW brings many security issues (e.g., 51% attacks).

2.4.1.2. PoS [78]. The system allocates accounting rights based on the stake of miners. The difficulty of the puzzle changes with their stake, the puzzle becomes easier if the stake increases, so the node have more probabilities to obtain the accounting right. In other words, it relies on having an adequate stake in blockchain to participate in block creation [144]. This consensus mechanism saves computer resources. PoS assumes that the owner of equity is more willing to maintain the consistency and security of the system [145]. In the upcoming version of Ethereum, the consensus mechanism is changed from PoW to PoS.

2.4.1.3. DPoS [146]. It is based on voting and election. On the basis of PoS, all nodes are divided into leaders and followers. Followers will be notified when leaders reach consensus. Leaders can also be called as block producers which is more specialized because they are selected by votes from token holders [75]. Due to the fewer participants for block validation, DPoS facilitates faster block generation [76]. But the consensus mechanism improves performance by sacrificing some decentralization characteristics.

2.4.1.4. PBFT (Practical Byzantine Fault Tolerance) [147]. Practical Byzantine fault tolerance is an improved version of Byzantine fault tolerance (BFT). It reduces the operational complexity of BFT. PBFT assumes that some nodes in the system are faulty or dishonest, and it is proposed to be a high-performance consensus mechanism that relies on a set of trust nodes [81]. The nodes are ordered in a sequential manner, one node is primary and other nodes are backups [148]. The primary is called a view of the system. If the primary occurs problems, the primary is replaced by another node. Each round has a primary chosen in round-robin, so that the nodes are chosen with the frequency in proportion to the voting power [149].

2.4.2. Security issues

2.4.2.1. Sybil attack [150]. Sybil attack control a malicious node that creates a large number of pseudonymous identifiers and disguises them as normal nodes, thereby obtaining a disproportionate level of control over the network and destroying the reputation system of the blockchain network, such as affecting the voting results.

2.4.2.2. Nothing at Stake (NAS) attack [124]. When a fork occurs in a PoS system, the block-producing node can generate blocks at the same time without any loss. No matter which fork is used as the main chain in the future, the node can obtain all the benefits without any cost. So some nodes support or initiate illegal transactions to generate new forks, other nodes queue up blocks on multiple chains, and there will be more and more forks, and block-producing nodes will be unable to reach consensus.

2.4.2.3. Balance attack. The balance attack essentially violates the persistence of the master branch prefix and allows double spending, aiming to against POW-based blockchain [151]. It allows attackers with low mining capabilities to immediately interrupt the communication between subgroups with similar mining capabilities.

2.5. Application layer

This layer runs at the top of the blockchain. After blockchain operation for several years, the functions of blockchain is not limited to cryptocurrency, it broadens the application scope and establishes a Turing-complete system. Some blockchains use smart contracts to build DApps, domain name services, data storage applications, etc.

2.5.1. Decentralized applications

DApps are operated through blockchain nodes and do not rely on any centralized server. They exist on the Internet in a way that is not controlled by any single entity and uses SSL/TLS(Secure Sockets Layer/Transport Layer Security) encrypted protocols to transmit data [69]. With the rapid expansion of the scale of DApps, the necessity of illegal content supervision on DApps or blockchain become increasingly prominent. DApps are different from traditional applications. Because there is no centralized server in DApps, all nodes can be understood as servers of DApps. The change brings challenge to security supervision on blockchain.

2.5.2. Blockchain services

At present, each blockchain has its own cryptocurrency. Miners can earn cryptocurrency through mining blocks [119]. Users can purchase various blockchain services and trade other users' products through cryptocurrency [81]. Due to the immutability of blockchain data, important but non-secret files can be stored on blockchain.

Blockchain technology has been applied to various application scenarios [73,152–155], such as financial services (e.g., securities trading, banking financial management, crowdfunding management), credit reporting and ownership management, distributed cloud storage, trade management, resource sharing (e.g., short-term rental sharing, community sharing), IoT(Internet of Things), advertising, medical treatment, social communication, voting, traceability, forecasting, gaming, copyright protection and other scenarios. These scenarios show the strong market potential of blockchain.

2.5.3. Security issues

2.5.3.1. RPC (Remote Procedure Call) vulnerability attack. In Ethereum, we can use RPC service to implement remoting of Ethereum interfaces, and call the API(Application Programming Interface) through different methods: HTTP(HyperText Transfer Protocol), IPC(Inter-Process Communication), WebSocket, etc [156]. Port 8545 is the default port to provide RPC service. RPC service uses JSON for data transmission. Attackers can use most of API commands to obtain information about the connected nodes.

Because the attacker can directly use API commands of the victim's node though RPC service without passwords or authentication. If the attacker successfully finds nodes that open the RPC service through network scanning, some sensitive information of the victim's node may be leaked (e.g., balance, coinbase address, account address, status of the node, connecting nodes, etc.). The attacker can also use $e$th_sendTransaction() method to continuously send transactions to the account controlled by the attacker. When the victim needs to trade with other nodes, his account will be unlocked, and the transaction of transferring the balance from the victim's account to the attacker will be successful. The attack lasted for two years. The amount of stolen ETH reached 20 million dollars. This vulnerability not only brings economic losses, but also leaks sensitive information of nodes, thereby leaking user privacy.

2.5.3.2. Cryptojacking attack. Cryptojacking is a malicious behavior on blockchain. The attacker induces users to load mining code, infects the computers of victims, and uses the bandwidth and calculation capacity of devices for mining without authorization [157]. The intrusion methods include malicious connections or emails, mining scripts on websites, etc. Because mining requires a large amount of computing power, the attacker attacks multiple computers to have sufficient computing power for mining blocks.

2.5.3.3. Selfish mining attack. Selfish mining attack is introduced as "Block discarding attack" in [158] and also as "Block withholding attack" in [159]. The attacker utilizes "block concealing". Selfish miners or selfish mining pools reveal the block at a special time. The goal of selfish mining attack by selfish miners or mining pools is achieving more rewards in comparison with its hashing power [160]. Eyal et al. [161] proposed that selfish mining attack invalidates the work of honest miners if the attacker achieves the attack through selfish mining strategies.

2.5.3.4. Phishing attack. The attacker pretends to be a trusted person and obtains the victim's account address, password, private key, and other information through links, aggressive malicious spam, communication software and domain hijacking [162]. The spam messages follow a variety of formats, including missed fax messages, overdue invoices, and financial statements. CTB-Locker was delivered via spear-phishing emails weaponized with a ZIP file [163]. In November 2017, Bitcoin Gold team stated on Twitter twice times that mybtgwallet.com is a secure wallet website. They required users to upload private keys and recovery the seed to generate a Bitcoin Gold wallet. But a few days later, the assets of some users had transferred away, users lost a lot of ETH, BTC, LTC(Litecoin), etc; In February 2018, a Ukrainian hacking organization purchased keyword advertisements related to cryptocurrencies in Google search engine and disguised advertisements as legitimate websites. This organization stole more than 50 million dollars from Blockchain.info; In June 2019, attackers sent extortion messages to some exchanges and obtained more than 400,000 dollars through phishing emails.

2.5.3.5. Trojan attack. The attacker hides a piece of malicious code with special functions in normal code and uses it to hide the control program in the victim's system to steal user personal information or remotely control the computer. In June 2011, Trojan files disguised as wallet.dat in the victim's computer and searched sensitive files such as Bitcoin wallets. Trojan sent the important files to the attacker through smtp.wp.pl email service, then victim's data is stolen; In 2014, Mt.Gox key file was stored in local, which is a plain text, and the wallet.dat file was leaked due to Trojan horse infection, resulting in a large amount of economic loss. Finally, Mt.Gox went bankrupt.

2.5.3.6. Quantum attack. As many cryptocurrencies use the SHA256 encryption algorithm, with the development of quantum computing technology and the realization of quantum hegemony(i.e., quantum computing has computing power that surpasses all traditional computers), the encrypted algorithm would be cracked, and it is no problem to destroy the cryptocurrency. This attack can use Grover's algorithm to accelerate brute force attacks by a quadratic factor. The algorithm can be used to accelerate mining in blockchain(i.e., speeding up the generation of nonce), which would allow for recreating entire blockchain fast, thereby undermining their integrity [164].

Table 3
Corresponding relations between the problems of central structure and characteristics of blockchain.

	Authenticity verification	Server problems	Incentive mechanism	Unequal distribution	Privacy issues	Ownership of intellectual property	Quality of network
decentralization		✓					
Immutability of Data	✓					✓	
Information Sharing		✓					
Currency liquidity			✓	✓			✓
anonymity					✓		
Traceability of Transaction						✓	

2.5.3.7. Supply chain attack. Supply chain attack uses dependencies on packages and modules. If the system was infected at any link will cause problems in all links [165,166]. In November 2018, the attacker attacked Copay which belongs to Bitpay. The malicious behavior was hidden for two months. In this period, the attacker polluted EvenStream and left the value of Copay-related variables in the backdoor. Finally, the user's private key was stolen through targeted attacks on Copay. In addition, there is an another attack that user cannot detect the presence of counterfeit products in supply chain, thereby reducing the manufacturer's influence in the market [167].

2.5.3.8. Integer overflow attack [168]. Each type of blockchain data has a boundary. The uint8 variable of Ethereum can store 0–255 size data. When exceeding the boundary, the data overflows and has errors. When a malicious user manipulates his account to send tokens to other accounts more than his own balance through underflow, the malicious user's balance will underflow into an oversized value if it is not checked in the smart contract. The attacker sells a large number of tokens, which will crash the entire system. In August 2010, block 74638 contained a Bitcoin transaction of more than 180 billion dollars. When the client added the transaction to the block, it did not verify whether the amount of transaction was too large. In 2018, the attacker utilized vulnerability in the smart contract to achieve this attack, such as BatchOverflow [169], ProxyOverflow [170], MultiOverflow [171].

2.5.3.9. Exceed authority access attack. Exceeding authority refers to accessing or executing permissions beyond the current account, such as the famous BetDice game on EOS. Due to the router in the smart contract (i.e., a custom event forwarder), the source account is not strictly verified, and ordinary users can use push action to access transfer which is the key function in the smart contract [172], bypassing the transfer process and betting directly. The vulnerability allows the attacker to obtain nearly 50,000 EOS in the BetDice prize pool with no cost; In July 2017, Parity wallet was stolen. The attacker used the Enhanced-wallet.sol bug in the Multi-Sig library file written by the Parity founder to reset the owner of wallet and take over the wallet. Finally, Parity lost 150,000 ETH. In the same year, Parity deployed a new WalletLibrary contract, which has a similar vulnerability. Anyone can use the init_wallet function to tamper the ownership of wallets. In November, devops199, a Github user, used the vulnerability of the contract. He set the owner of the contract as himself and called the suicide function. Finally, it led to freeze all assets, about 278 million dollars.

2.5.3.10. Ransom and money laundering [173]. Ransomware is a common attack method. Attackers use phishing websites, links, or directly send ransomware to encrypt host files to extort currency. At present, many ransomware attacks start to use cryptocurrencies (i.e., Bitcoin, Monero, etc.) as a means of payment. For example, GandCrab virus obtained 2 billion dollars in a year through cryptocurrencies; WannaCry ransomware infected 230 thousand victims across 150 countries in two days [174]. The attackers firstly used Bitcoin to extort ransom. Then attackers used money laundering to converted Bitcoin to Monero, which is more anonymous, so that ransom cannot be traced. Thus achieving the purpose of escaping supervision. Compared to other digital currencies, Bitcoin has the lowest risk of being utilized for money laundering [175]. Cody et al. designed Dark Wallet which can make Bitcoin transactions completely private and stealth. It used stealth addresses and coin mixing to provide services [176].

2.6. Security supervision of blockchain

2.6.1. Supervision issues

Since blockchain technology has a wide range of application scenarios, it has attracted extensive attention. According to the security issues mentioned above, although blockchain technology brings convenience and economic benefits, its unregulated nature and inherent anonymity have attracted many attackers to conduct illicit activities and attacks, which seriously harm the interests of normal users and countries. It is possible to prevent attackers from attacking through some protective measures, such as modifying the vulnerability code [133,169–172]. In other words, the network administrator can only protect their users through some methods after the attacker exploits the vulnerability to attack the victim, but the loss has already occurred, and the inability to track the attacker would continue to threaten normal users and countries. Therefore, while not affecting the development of blockchain technology, the regulator needs to supervise suspicious users, prevent risks, and maintain normal order on blockchain. The security supervision of blockchain will play an important role in the development and realization of blockchain technology.

The emergence of blockchain technology brings shock and solutions to traditional centralized technology, which is high cost and low efficiency [6]. Table 3 shows which characteristics of blockchain can be used to solve the problems of the centralized structure. However, blockchain is significantly different from traditional centralized architecture. As for blockchain, the performance and security capabilities of nodes are different, because nodes are deployed on the computers with different hardware configurations and there is no unified manager. Traditional supervision methods are no longer suitable for blockchain. For example, the traditional identity traceability mechanism based on KYC(Know Your Customer) policy is difficult to accurately track the identity of the attacker.

2.6.2. Blockchain supervision

According to the characteristics of blockchain, we can divide blockchain supervision into the following four categories:

2.6.2.1. Transaction supervision. It refers to detect suspicious transactions, and obtain the currency flow of transactions, thereby finding the source address of transactions.

Many attacks can be understood as attackers stealing cryptocurrencies from normal user accounts or threatening users to transfer cryptocurrencies to their accounts. The attackers will quickly spend illegal currencies through other illegal transactions or make them become legal income through money laundering rather than store them for a long time. A blockchain address is a string of letters and numbers, which is automatically generated by the users. Users use them as pseudonyms (i.e., input transaction address, output transaction address) in blockchain platform [18,51]. Historical transactions of blockchain are completely public. So the currency flow of a specific transaction can be obtained through the association of relevant blockchain addresses.

2.6.2.2. Entity identity supervision. It mainly refers to find the relationship between blockchain addresses and user IPs/IDs.

Although we can obtain the currency flow of illegal transactions and find the source address through some methods, we cannot determine the real masters behind attacks, so associating blockchain addresses with real user identities is more necessary.

2.6.2.3. Threat information supervision. It refers to the supervision of additional information in blockchain transactions and harmful information of decentralized applications.

In Aachen University of Technology and University of Frankfurt, a group of researchers conducted a quantitative analysis of Bitcoin's additional information and found that 59 documents contained politically sensitive content, the link of privacy violations, illegal images [177]. Although only a few Bitcoin transactions contain illegal information, it is sufficient to make victims at risk. With the emergence of smart contracts, the scale of decentralized applications is also rapidly expanding. It is necessary to supervise illegal content of DApps which have the function of spreading information. When there is harmful information in parsed additional information, the regulator can supervise this address for a long time [70,178]. When the owner of the address participates in other blockchain activities, the identity of the user may be associated, thereby supervising the user who spreads harmful information.

2.6.2.4. Abnormal behavior supervision. Due to the characteristics, such as self-organization, decentralization, changeable application scenarios, protocol encryption, etc., blockchain nodes are given different functions. By controlling nodes, attackers take advantage of anonymization to steal cryptocurrencies, isolate target nodes [115,116], continuously send small amounts of transactions [136], etc. Regulators can supervise nodes that have abnormal behaviors by analyzing different node behaviors.

2.7. Summary

Most security issues mentioned in this section utilize some methods to steal cryptocurrencies from exchanges, trading platforms, user wallets, DApps, and blockchain platforms, some attackers use blockchain as a tool of ransom, money laundering. Then, attackers transfer cryptocurrencies to their own blockchain accounts. Although there are some protection methods, it is unable to find the masterminds behind modern cyber-crimes, which may pose a constant threat to users. Countries can improve the level of security by supervising malicious users. Although the transfer process of stolen cryptocurrencies can be analyzed through historical transaction data [43–47]. Due to the anonymity of blockchain, it is hard to create a mapping between blockchain address and user's real identity, which brings huge challenges to the security supervision of blockchain.

In the above-mentioned four-layer structure, the consensus layer is very important to blockchain. Its main function is to allow all nodes of the entire network to reach a consensus under the premise of mutual distrust. A large number of studies on this layer aim to improve the efficiency of the consensus mechanism. There is almost no research related to this survey, and neither normal users nor malicious users have little interaction with this layer. So in this review, we only briefly introduce the common consensus mechanisms used in blockchain.

In the following three sections, we combine three aspects to contribute the security supervision, including the network layer, transaction layer, and application layer. The regulator can detect abnormal behaviors of nodes and discover anomalous nodes on the network layer [16–19,21–24], and combine with the whole chain or currency flow of suspicious and illegal transactions, which is obtained through the analysis of historical transaction records [25,26,37,40,179]. By using clustering, machine learning and association techniques, some addresses can be correlated with real entities [30,43,45,52,56,60,63–65]. Then they can track the specific IP address to obtain abnormal behaviors of the user by analyzing application network traffic [66,67,69–72,180]. The regulator can form a suspicious user portrait of this IP address through blockchain nodes, the use of DApps and the detailed DApps user behaviors. When the suspicious user performs abnormal behaviors, this would be promptly reflected to the relevant departments.

3. Node discovery technology on the network layer

On the network layer of blockchain, the regulator can obtain communication information of nodes through active detection [16–19,21,22], passive monitoring [24], etc. Combining the two methods can verify each other and make up for deficiencies [23].

3.1. Active detection

Active detection methods manipulate the node discovery mechanism in an aggressive manner. Due to the openness and transparency of blockchain, anyone can access blockchain network, and blockchain does not distinguish whether the node is a normal node or a malicious node. After the malicious node connects the network, it can monitor block information on the network layer. Blockchain platforms such as Bitcoin, Ethereum, EOS, etc., are open source projects. By analyzing the source code, the regulator can understand the process of networking, handshake, encryption, mining, trading, etc. The procedure of action detection is as follows [16,18,19,23]:

(i) Modifying the source code of client to form a probe node, and deploying it in the network as a controlled node.
(ii) Constructing specific probe packets based on the corresponding network protocol (e.g., constructing a Getaddr message in Bitcoin, constructing a FindNode message in Ethereum).
(iii) Repeatedly querying neighbors to gather more nodes information actively (e.g., IP address, node ID), and recording them into a database.
(iv) continuously iterating step (iii) to form the communication relationship network of blockchain.

By using active detection, the regulator can analyze the propagation path of threat information on the network layer and find the original node, and they can also determine the node communicating with the abnormal node as a node that may be unusual. Because IP addresses of neighbor nodes can be collected, the regulator can obtain the geographic location [25]. The active detection is faster and more targeted.

In the blockchain network, nodes use the relay mode to spread messages. The original node refers to the first node to send the transaction, and other nodes are relay nodes [20]. If the regulator finds the original node, they can associate the input addresses of the dubious transaction and IP address. But it is difficult to distinguish the original node and the relay node, because there is no difference in the forwarded information. Koshy et al. [20] analyzed that there are three forwarding modes for Bitcoin transaction data:

- Multi-layer forwarding (Multi-Relayer, Non-Rerelayed Transactions): transaction information is forwarded by multiple nodes, but each node forwards only once;
- Single-layer forwarding (Single-Relayer Transactions): transaction information is forwarded once or multiple times by one node;
- Multi-layer forwarding (Multi-Relayer, Rerelayed Transactions): transaction information is forwarded by multiple nodes, but there may be one node repeatedly forward the transaction, or at least two nodes repeatedly forward the transaction.

The authors found that normal transactions use Multi-layer forwarding mode. Some specific transactions use the other two modes, which can be used to infer the original node of the transaction. However, their approach is limited to the transactions that expose anomalous behavior like transactions relayed only once or transactions that were relayed multiple times by the same node. But these specific transactions only account for less than 9% of total transactions on Bitcoin, so the method has a limitation.

The approach of searching neighbor nodes in the active detection method can be divided into breadth-first and depth-first. Because the technique of storing neighbor nodes on Ethereum is more mature than Bitcoin, we use Ethereum as an example:

- Breadth-first: Dividing the value space of the node ID into 256 subspaces on average. In every round, generating a random node ID in each subspace, then selecting the closest 16 nodes to the 256 node IDs among the existing nodes in the node pool. Finally, requiring the neighbor nodes of 256 ∗ 16 nodes through FindNode messages.
- Depth-first: By analyzing the K-bucket mechanism, constructing a number of node IDs with increasable distances from the existing node in the node pool, and trying to get the information of all neighbor nodes of this node. The strategy uses K buckets as the discovery unit, which can increase the utilization rate of existing nodes as much as possible and increase the speed of node discovery.

Feld et al. [16] used active detection methods to traverse Bitcoin's P2P network recursively in a protocol-compliant manner and perform an analysis of the interconnected degree of Bitcoin's P2P network under an AS-level(AS, Autonomous System) perspective. Their client actively sent constructed messages to neighbor nodes to obtain the list of IP addresses of their connected nodes. Most nodes resided in the same autonomous system left Bitcoin's vitality, resilience and security flawed. But the method only covered a small part of Bitcoin clients, and a node may be counted twice due to the regular disconnect caused by the node's ISP(Internet Service Provider).

3.1.1. Novel monitoring tool

In order to facilitate the data analysis of blockchain nodes, various blockchain tools are developed for scanning and monitoring networks, such as BITCOIN-NODE-SCANNER [17] for Bitcoin, NodeFinder [19] for Ethereum, NodeScanner and NeighborFinder [18] for Monero. These tools also bring convenience to the security supervision of blockchain, such as finding the geographic location of abnormal nodes in time, understanding the network behavior habits (e.g., circulation, response time) of malicious nodes. Three tools are detailed as follows:

- BITCOIN-NODE-SCANNER: the tool is based on the Bitcoin protocol, and helps to obtain as many IP addresses of nodes while acting as a fully functional client. The tool pays more attention to the network statistics (e.g., country distribution, port number, protocol version, number of nodes, agent type, etc.). Approximately one million users and 8500–23000 full-node peers were collected.
- NodeFinder: a node discovery tool based on Geth. In order to be an effective scanner, NodeFinder ignores the maximum peer limit by never sending too many peers disconnects, and it frees up their peer slots when it has gathered the nodes information from peer connection. The authors centrally deployed 30 instances of NodeFinder and used one machine to perform the process of discovery and identification for 3 months. They found a noisy network environment, and less than half of nodes were contributing to the main network. Due to the lack of ground truth data, their analysis results were not validated.
- NodeScanner and NeighborFinder: as for NodeScanner, it extracts the set of unique IP addresses from the received incoming connections to form the peer list; As for NeighborFinder, it infers the neighbors of the target node through the different answer delays to the SYN(Synchronize Sequence Numbers) packets which are sent to all of its neighbors to check its connections. Some information can be analyzed through the two tools, such as Monero network structure, connection relationship, number of nodes, types and distribution of nodes.

3.1.2. Blockchain with Tor

Due to the popularity of blockchain, many attackers know the disadvantages of blockchain, like pseudonyms, which is not completely anonymity. They try to use anonymity services (e.g., Tor, proxy servers) to avoid being tracked by the regulators. Fortunately, some studies show that attackers can still be found even if anonymous services such as Tor are used.

Biryukov et al. [21] deployed a number of Bitcoin nodes and medium-bandwidth Tor Exit relays, then periodically advertised the newly deployed nodes, therefore they were contained into the maximum number of buckets at the client-side, and made non-attacker's Bitcoin nodes ban the Tor Exit nodes. Finally, they may force users who decide to connect the Bitcoin network through anonymity services to connect exclusively through their Tor Exit nodes or Bitcoin nodes, thereby completely isolating the client from the rest of Bitcoin. However, the method would not be able to prohibit all relays or VPNs(Virtual Private Networks) from the network nodes due to the status of DoS protection mechanism in Bitcoin.

Biryukov et al. [22] tried to associate pseudonyms of Bitcoin users behind NAT or firewalls with IP address where the transaction is produced. At first, they used the punishment mechanism in Bitcoin to remove users who connected Bitcoin through Tor. Because users need to broadcast their nodes in Bitcoin network, the authors analyzed the access nodes which are used by interested users to connect the network. Then they checked the relaying nodes of the network server to find out the earliest node that forwards specific user information. At the same time, they collected the transaction information broadcast by these access nodes. If the two information is from the same node, indicating that the user of this node sent this transaction information.

3.2. Passive monitoring

Passive monitoring refers to inject some controlled nodes into different locations in the network, aiming to discover nodes by capturing all the traffic sent and received [23]. The difference between passive monitoring and active detection is that passive monitoring receives messages passively instead of continually sending messages actively. If the environment allows, passive monitoring has the following method [24]. Blockchain data is transmitted via TCP or UDP, and has some features in network traffic. First, by analyzing blockchain protocol and communication process, the regulator can extract the traffic features of communication between nodes. Second, by restoring communication flow and analyzing traffic data, they can capture the traffic with hit features and parse the traffic data to obtain the communication content according to the blockchain protocol format. Then some relevant information is stored in the database, such as IP, port, timestamp, version, and communication content (e.g., neighbor node list). Through data association, it can accurately map potential threat users to physical entities and locations. The regulator can also analyze the topological relationship of abnormal nodes. Compared with active methods, passive methods have the advantages of avoiding the extra network burden and discovering nodes behind network [23].

Li et al. [24] obtained communication data from the passive monitoring perspective, and they proposed a passive method using traffic association and machine learning to conduct online Ethereum node detection. Three datasets were collected:

(i) Set1: gathering Ethereum routing nodes from active detection and the third-party public websites;
(ii) Set2: monitoring NetFlow traffic and finding out the nodes connect to set1. Nodes in set2 are considered as Ethereum nodes, such as light clients, online wallets or other routing nodes;
(iii) Set3: repeating the previous step on set2 in NetFlow traffic to compose set3.

But the accuracy of set3 is much lower than set2, so the authors set set1 and set2 as training sets and used machine learning to train set3 to obtain set4, which has higher accuracy. Set1, set2 and set4 formed the Ethereum true nodes. They mapped the distribution of Ethereum nodes from the passive perspective, but it is difficult to check whether the nodes are complete.

Table 4
Traffic features of five public blockchain.

Blockchain	Traffic features	
Bitcoin	The first 4 bytes of data part of packet are 0xf9beb4d9	
Ethereum	V4 version protocol	1. The 98th byte of data part of packet represents message type, and the value is 1 to 4; 2. The last 4 bytes of data packet are unix timestamp, and the last 5th byte is the fixed value 0x84.
	V5 version protocol	1. The first 22 bytes of data part of packet are fixed values: 0x74656d706f72617279206469736f76657279207635; 2. The 88th byte of data part of packet represents the message type, and the value is 1 to 8.
EOS	1. The length of data packet is generally between 300 and 500 bytes; 2. The total length of the data packet minus 4 is equal to the value of the first 4 bytes of packet header; 3. The 5th byte is 0; 4. It is a fixed value starting from the eighth byte and represents the id of the main network: aca376f206b8fc25a6ed44dbdc66547c36c6c33e3a119ffbeaef943642f0e906	
IPFS	1. The data part of the first packet is fixed at 20 bytes: 0x132f6d756c746973747265616d2f312e302e300a; 2. The data part of the second packet is fixed at 14 bytes: 0x0b2f746c732f312e302e300a.	
Steem	The first four packets have a fixed number of bytes, 33, 576, 16 and 16 bytes.	

Table 5
The summary of typical technologies of the network layer.

Node detection	Active detection	Passive monitoring	Platform	Topology	Main statistics	Description
Koshy et al. [20]	✓	×	Bitcoin	×	–	Mapping addresses directly to IP, finding three forwarding modes.
Feld et al. [16]	✓	×	Bitcoin	×	(i) Routable clients distribution. (ii) Intersection between peerlists.	Analyzing nodes under an AS-level perspective.
Park et al. [17]	✓	×	Bitcoin	✓	(i) Geographical location. (ii) RTT measurement using ping to all IP.	A comparative measurement of nodes in Bitcoin through BITCOIN-NODE-SCANNER.
Kim et al. [19]	✓	×	Ethereum	×	(i) Message I/O. (ii) Disconnect reasons. (iii) Peer latency.	A effective node discovery tool that ignores connection restrictions.
Cao et al. [18]	✓	×	Monero	✓	(i) Nodes location distribution. (ii) Degree distribution. (iii) Node connectivity.	Developing a toolset to investigate the Monero network.
Biryukov et al. [21]	✓	×	Bitcoin	×	–	Using the DoS protection mechanism to associate users who use Tor.
Biryukov et al. [22]	✓	×	Bitcoin	×	–	Associating pseudonyms of users behind NAT or firewalls with IP where the transaction is produced.
Li et al. [24]	×	✓	Ethereum	✓	(i) Routing nodes stability. (ii) Routing nodes distribution.	Using traffic association and machine learning to conduct online node detection.
Gao et al. [23]	✓	✓	Ethereum	✓	(i) Small world characteristics. (ii) IP address aliasing. (iii) Node ID aliasing.	Combining active and passive detection technique.

Different blockchain platforms use different blockchain protocols, so the traffic feature of each blockchain are also different. Traffic features of several mainstream blockchain platforms are shown in Table 4.

Passive monitoring method can be used in home gateway routers, public WiFi, ISP, national network ports, etc. But if more comprehensive information is needed, it should be monitored at the national network ports, which is easier for the staff who supervises the blockchain network traffic.

3.3. Comprehensive monitoring

Active detection passive monitoring have their own advantages and disadvantages, the former is more focused and the latter is broader.

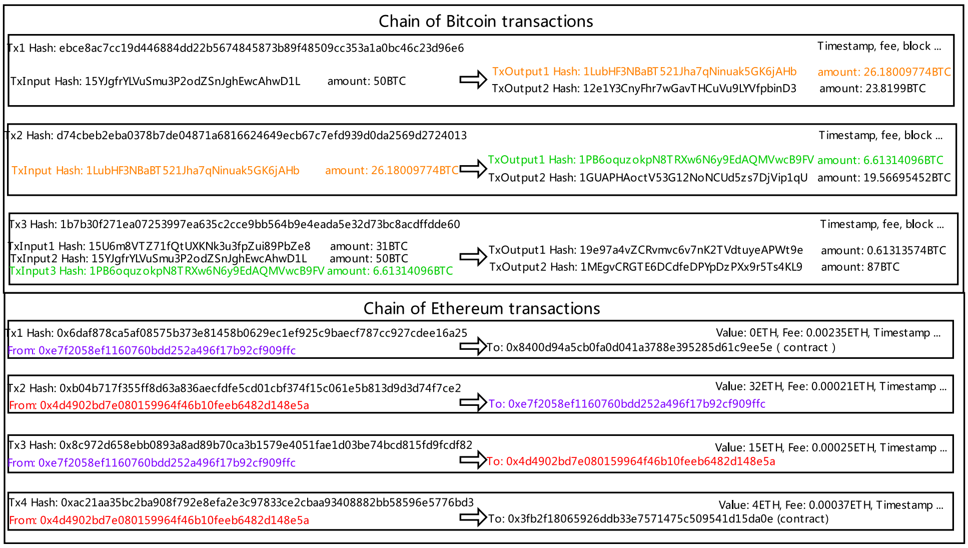
The communication of Ethereum P2P network is composed of some network protocols, named devp2p. Two key components are Node Discovery Protocol and RLPx Transport Protocol. Gao et al. [23] used passive monitoring to separate inactive nodes and non-routable nodes, then they modified the source code to make clients repeatedly query neighbors and aggressively gather more nodes through active detection. The active approaches are faster and more targeted. They found that there are about one hundred abnormal nodes in the network.

3.4. Summary

We summarize the surveyed studies of the network layer in Table 5 from the following aspects: detection method, platform, topology (network topology diagram), statistics. Besides, we provide a description of each method.

4. Data analysis technology on the transaction layer

The global ledger that stores historical transaction data of blockchain is completely public. Any nodes can join the network and download complete historical transactions. Fig. 4 shows the traceability of blockchain transactions, the detailed information of the transaction is recorded in blockchain, and the currency flow of the transaction can be obtained according to the timestamp. Different blockchains use different accounting models, such as Bitcoin and Ethereum use



Chain of Bitcoin transactions

Tx1 Hash: ebce8ac7cc19d446884dd22b5674845873b89f48509cc353a1a0bc46c23d96e6 Timestamp, fee, block ...
TxInput Hash: 15YJgfrYLVuSmu3P2odZSnJghEwcAhwD1L amount: 50BTC ⇨ TxOutput1 Hash: 1LubHF3NBaBT521Jha7qNinuak5GK6jAHb amount: 26.18009774BTC
TxOutput2 Hash: 12e1Y3CnyFhr7wGavTHCuVu9LYVfpbinD3 amount: 23.8199BTC

Tx2 Hash: d74cbeb2eba0378b7de04871a6816624649ecb67c7efd939d0da2569d2724013 Timestamp, fee, block ...
TxInput Hash: 1LubHF3NBaBT521Jha7qNinuak5GK6jAHb amount: 26.18009774BTC ⇨ TxOutput1 Hash: 1PB6oquzokpN8TRXw6N6y9EdAQMVwcB9FV amount: 6.61314096BTC
TxOutput2 Hash: 1GUAPHAoctV53G12NoNCUd5zs7DjVip1qU amount: 19.56695452BTC

Tx3 Hash: 1b7b30f271ea07253997ea635c2cce9bb564b9e4eada5e32d73bc8acdffdde60 Timestamp, fee, block ...
TxInput1 Hash: 15U6m8VTZ71fQtUXKNk3u3fpZui89PbZe8 amount: 31BTC
TxInput2 Hash: 15YJgfrYLVuSmu3P2odZSnJghEwcAhwD1L amount: 50BTC ⇨ TxOutput1 Hash: 19e97a4vZCRvmvc6v7nK2TVdtuyeAPWt9e amount: 0.61313574BTC
TxInput3 Hash: 1PB6oquzokpN8TRXw6N6y9EdAQMVwcB9FV amount: 6.61314096BTC TxOutput2 Hash: 1MEgvCRGTE6DCdfeDPYpDzPXx9r5Ts4KL9 amount: 87BTC

Chain of Ethereum transactions

Tx1 Hash: 0x6daf878ca5af08575b373e81458b0629ec1ef925c9baecf787cc927cdee16a25 Value: 0ETH, Fee: 0.00235ETH, Timestamp ...
From: 0xe7f2058ef1160760bdd252a496f17b92cf909ffc ⇨ To: 0x8400d94a5cb0fa0d041a3788e395285d61c9ee5e (contract)

Tx2 Hash: 0xb04b717f355ff8d63a836aecfdfe5cd01cbf374f15c061e5b813d9d3d74f7ce2 Value: 32ETH, Fee: 0.00021ETH, Timestamp ...
From: 0x4d4902bd7e080159964f46b10feeb6482d148e5a ⇨ To: 0xe7f2058ef1160760bdd252a496f17b92cf909ffc

Tx3 Hash: 0x8c972d658ebb0893a8ad89b70ca3b1579e4051fae1d03be74bcd815fd9fcdf82 Value: 15ETH, Fee: 0.00025ETH, Timestamp ...
From: 0xe7f2058ef1160760bdd252a496f17b92cf909ffc ⇨ To: 0x4d4902bd7e080159964f46b10feeb6482d148e5a

Tx4 Hash: 0xac21aa35bc2ba908f792e8efa2e3c97833ce2cbaa93408882bb58596e5776bd3 Value: 4ETH, Fee: 0.00037ETH, Timestamp ...
From: 0x4d4902bd7e080159964f46b10feeb6482d148e5a ⇨ To: 0x3fb2f18065926ddb33e7571475c509541d15da0e (contract)

Fig. 4. The chains of real transactions. Addresses with the same color in different transactions are same.

UTXO and account model, respectively. Any historical transactions are traceable. We divide the analysis of transaction layer into transaction analysis and user identity analysis.

4.1. Transaction analysis

All transactions are open and transparent, so the regulator can obtain the information of dubious transactions, such as account balance, currency flow, transaction records, and address information of potential threat users.

Any transaction can be traced to the miner's mining address through historical transaction records. Mining information and cryptocurrency transfer information are recorded in the ledger. Through blockchain mechanism and history transaction data, the regulator can get transaction graph (G1), address graph (G2), and entity graph (G3) [25,28,181], which are popular in research. The three graphs are detailed as following:

- Transaction graph (G1): the vertices represent transactions, directed edges represent the amount of BTC from the output of a transaction to the input of a transaction, and the value and timestamp are recorded on the edge.
- Address graph (G2): the vertices represent blockchain addresses (i.e., a public key), edges represent transactions between blockchain addresses, and the amount of transaction is the weight on the edge. All transactions can be added to the graph.
- Entity graph (G3): it is divided from the address graph and contains multiple graphs. The entity is characterized by a set of addresses, which can be interpreted as an organization or a user, so it is also called user graph.

Fig. 5 shows examples of transaction graph (left) and entity graph (right). Researchers utilize some algorithms to cluster the addresses belonging to the same user and use a series of addresses as nodes to draw the graph.

There are some challenges in data analytics of blockchain data, especially in Ethereum, EOS, etc. The data is stored at clients in heterogeneous and complex data structures, which is either binary or encrypted. Zheng et al. [182] extracted raw data consisting of 8100000 blocks of Ethereum, including three types: blocks, traces, and receipts. They processed and categorized the Ethereum data into six datasets: block and transaction, internal Ether transaction, contract information, contract calls, ERC20 token transactions, ERC721 token transactions, which are convenient to analyze blockchain.

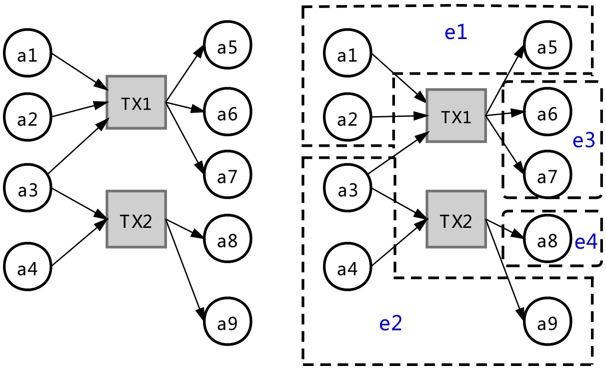


Fig. 5. Example of transaction graph (left) and entity graph (right).

Blockchain's currency flow and details of transactions can be analyzed through transaction data correlation, graph theory, heuristics. Because the data mechanism of each blockchain platform is different, we summarize research in terms of blockchain platform.

4.1.1. Bitcoin

Due to UTXO, transactions of Bitcoin are different from transactions of other blockchains. Fig. 6 shows four types of Bitcoin transactions [183]. Each rectangle can be thought of as a certain amount of BTC. Transaction 1 to 4 have 1/1, 2/2, 1/2, 2/1 input/output addresses respectively. Because input and output sources cannot be separated, there are ten different bitcoin addresses in the figure.

The UTXO accounting model has some unique mechanisms, which will help to link related transactions based on the time written to the blocks:

- The BTC in one address cannot be split into two parts to transfer;
- The input address of a transaction is the output address of the previous transaction;
- The output address of a transaction is the input address of other transactions;
- A transaction records all the input and output addresses in detail.

Many research methods use these rules as heuristic algorithms for finding currency flow [26,181,184].

For the complete BTC flow, It is generated by mining. At the beginning of creation of Bitcoin, miners can get a reward of 50 BTC,

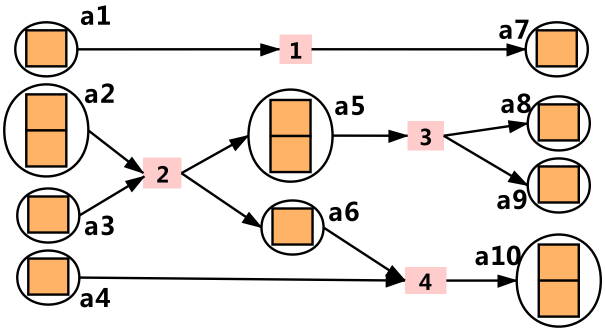


Fig. 6. Four basic type transactions. Each rectangle can be thought of as a certain amount of BTC.

and now mining successfully is rewarded by 12.5 BTC, and the reward is halved about every four years. The mining revenue can be calculated as follows:

$$50 * (1/2)^{\left[\frac{blockid}{210000}\right]} \tag{1}$$

There are many blockchain browsers that download all historical transaction records and display them on the web [185–187]. The regulator can directly retrieve suspicious addresses and transactions, transaction chains and other information related to the suspicious address will be shown in the interface.

Reid et al. [25] utilized a transaction graph (G1) and user graphs (G3) to conduct an anonymity analysis of Bitcoin. In the user graph, the input public keys in the same transaction are connected to build auxiliary networks. It may be possible to identify, such as communities, central users and hoarders within user graphs. After analyzing, the user network is found as a large circular structure. Then they inferred the information of Bitcoin users through various methods, for example, establishing mappings between Bitcoin users and IPs or organizations (The Bitcoin Faucet, voluntary disclosures of public-keys by users), investigating identity leakage, context discovery, flow and temporal analyses.

Zhao et al. [26] used a graph-based method to analyze the clustering of user identity and the attributes of currency flow. In order to group the addresses into unique sets, the Bitcoin mechanisms above were used to cluster, and a new algorithm was proposed as following:

(i) For the next transaction, if the input addresses are not in the database, map the addresses to a new ID, and traverse addresses that already exist and point to the ID of the root node, then save them in a group.
(ii) If there are intermediate nodes in the path, the nodes are recorded in the database, which is used for node's path reference.
(iii) Check the root node ID set in (i) whether is empty, if it is not empty, map the root node IDs to a new node ID. Otherwise, terminate the algorithm.

The method can help identify suspicious transaction behaviors and currency flows of Bitcoin users.

Maesa et al. [184] analyzed user graph (G3) of Bitcoin and proposed a scalable clustering algorithm to support the construction of the address graph (G2). The time evolution of graphs was considered through time window method. The analysis verified some properties of Bitcoin network, like the "rich get richer" property and the existence of central nodes that play a dominating role between different parts of Bitcoin network. They also analyzed Bitcoin network from some statistics perspectives such as clusters size, average out-degree, transaction amounts.

Mining pool is the main component of blockchain network, which plays an important role in network performance and security aspects. Wang et al. [27] considered that the details of mining pool behaviors (e.g., empty blocks, mining revenue and transaction collection strategies) may affect the usage of Bitcoin users (e.g., transaction fees, transaction delay and transaction acceptance rate). In terms of computing power allocation, mining revenue, transaction delay, transaction collection strategies, they analyzed more than 1.56 hundred thousand historical transaction blocks and 120 million unconfirmed transaction data obtained from real-time network traffic, and found that the top 5 mining pools entities continuously controlled more than 51 percent of daily computing power. Some attackers may attack blockchain through mining pools, such as 51% attack, so it is necessary for the regulator to understand its malicious behaviors.

4.1.2. Ethereum

Smart contracts and the history of calls made to these contracts set Ethereum apart from other cryptocurrencies,

and users can use it to accomplish things that cannot be done on Bitcoin, such as DApps, domain service. Because attackers use smart contracts to attack users and Etherum, which brings more challenges to security supervision of Ethereum. In order to meet these challenges, several graphs are proposed:

- Money transfer graph (G4): the graph is similar to the transaction graph (G1), but the vertices in this graph represent external owned accounts (EOAs) and smart contracts, edges represent the direction of transactions, and the amount of transaction is the weight on the edge.
- Contract/Account creation graph (G5): the vertices in this graph represent EOAs and smart contracts, which the same as those in G4, the directed edges represent an account that creates a smart contract or account.
- Contract invocation graph (G6): the graph is used to analyze how contracts interact with one another. The ordered edges represent an account that invokes a contract, the total number of invocation is the weight on the edge.
- User-to-user graph (G7): the vertices indicate EOAs, the directed edges indicate a transfer of Ethereum.
- Contract-to-contract graph (G8): the vertices indicate smart contracts, the directed edges indicate 'create', 'call' or transfer between contracts.
- User-contract graph (G9): the graph is bipartite, one side only contains EOAs and another side only contains smart contracts. The directed edges indicate how users create and use contracts, and also indicate the Ether transfer from contracts to EOAs.

There are nearly six hundred thousand smart contracts that have been deployed to Ethereum. A smart contract can be created by either an EOA or another smart contract, and little is known about the features of the relationships. Kiffer et al. [188] modified the Ethereum Geth client to log all smart contract operations (e.g., deployment, invocation, destruction). At first, They measured the behavior of smart contracts and found that malicious contracts repeatedly call self-destruction operations and other abnormal behaviors. Contract interaction graph (G6) was used to explain how contracts are created. They also found that the number of contracts created by contracts exceeded the number created by users after 2017.

In Ethereum, the transaction may have internal transactions due to the execution of smart contracts, and internal transactions are not stored in blockchain. Researchers collect all internal transactions by replaying external transactions in their customize EVM which is inserted into the instrumentation code. There are five operations that lead to internal transactions, including CREATE, CALL, CALLCODE, DELEGATECALL and SELFDESTRUCT.

For further research, based on all transaction data, Chen et al. [29] utilized graph theory to analyze three major activities on Ethereum, money transfer, smart contract creation, and smart contract invocation, and they constructed MFG (G4), CCG (G5), CIG (G6) to represent them.

A novel approaches based on cross-graph analysis was proposed to address attack forensics and anomaly detection in Ethereum. Attack forensics aims to find all accounts dominated by the attacker if given a malicious smart contract, which was introduced as follows:

(i) Collecting all contracts created by the root by computing the weak connected component (WCC) containing the malicious contract from CCG.
(ii) For each node in WCC, locating all callers in CIG, and backtracking in CIG until reaching an external owned account.
(iii) The found Nodes in WCC and nodes invoking the nodes in the WCC are all dominated by the attacker.

The authors also proposed a method to detect abnormal contract creation. T1, T2, T3 are thresholds.

(i) For an account u, obtaining all contracts created by u from CCG, if the number is smaller than T1, u is considered to be benign.
(ii) For each created smart contract c, obtaining all edges pointing to it in CIG. If the total number of invocations to all contracts, belonging to the WCC, is smaller than T2 and the amount of Ether transferred by c is smaller than T3, u is considered abnormal.

The two approaches were evaluated through real cases demonstrates their effectiveness.

From another angle, Bai et al. [31] characterized the trading relationship through three types of graphs: user-to-user (G7), contract-to-contract (G8) and user-to-contract graphs (G9), and used time window to analyze the evolutionary behaviors of these graphs, including the sliding time window and the incremental time window.

4.1.3. Monero

Monero pays more attention to privacy than other cryptocurrencies. It adopts CryptoNote protocol, which achieves through "multi-layer linkable spontaneous anonymous group signature". Monero uses ring signatures, ring confidential transactions, and encrypted addresses to confuse the source, amount and destination of all transactions, which achieves true anonymization. Due to the unique privacy properties of Monero, it has rapidly gained popularity.

Kumar et al. [32] focused on the traceability analysis of Monero. The dataset consisted of 961463 non-coinbase transactions and 47428 RingCTs. The authors presented three heuristics. The experimental results showed that 87% of inputs were traceable. First, H1: leveraging Zero Mix-ins may lead to a cascade effect, in other words, the traceability of an input affects another input in a later transaction. Second, H2: leveraging output merging mainly trace RingCTs. If users want to spend his XMR(one digital token of Monero), he would have to merge these funds. Third, H3: utilizing temporal analysis. The three heuristics were introduced as following:

- Zero mixin removal heuristic (H1): as shown in Fig. 7(a), input in Tx_1 is traceable because Tx_1 uses no mix-ins. Input in Tx_2 is also traceable due to the one mix-in. Tx_3 uses two mix-ins, including Tx_1 and Tx_2, so input in Tx_3 is also traceable.
- Output merging heuristic (H2): as shown in Fig. 7(b), Tx_4 consists of one input and two outputs, Tx_5 consists of two inputs I1, I2 and one output. Both I1, I2 uses one mix-in, including outp_1, outp_2, respectively. Outp_1 and outp_2 can be identified as the real input keys being spent in Tx_5.
- Guess Newest Heuristic (H3): given a set of input keys, because a TXO would not remain unspent for an infinite time, the real key is one with the highest block height.

Moser et al. [33] improved Zero Mixin Removal Heuristic (H1) mentioned in [32] as following:

- Intersection removal heuristic (H4): if N rings include the same N members, it is unable to determine which output has been spent in which ring, but as one ring contains all the members, they can be marked as spent and others to these outputs as mix-ins.

Hinteregger et al. [34] generalized the heuristic further as following:

- Cross chain analysis heuristic (H5): they look for sets S of rings that satisfy the attribute: if S contains n sets, the union of these sets have n elements. This question is similar to the matching problem for bipartite graphs G = (V1, V2, E), where N(x) represents the set of neighbors of x, and $|S| \leqslant |N(S)|$. If the matching is perfect, holding $\forall S : S \subseteq V_1$.

But with introducing mandatory minimum ringsize and RingCT, only part of transactions can be traced through transaction analysis techniques. Triangular distribution protocol was also added to Monero, and it described that more than 25% of mix-ins must be taken from recent outputs, which made temporal analysis lose effect. Wijaya et al. [35] proposed a new method to reduce the size of anonymity sets of the transaction mix-ins. The approach is based on the leniency of the Monero daemon to create transactions through Monero wallet. By creating identical mix-ins and multiple transactions, outputs of these transactions may be used as mix-ins.

4.1.4. EOS

In contrast to Bitcoin and Ethereum, EOSIO is the first blockchain platform which is high throughput by using DPoS, so it has attracted significant attention from people. However, there were many attacks against EOSIO in recent years, exploiting vulnerabilities in DApps, which has led to millions of dollars lost.

Huang et al. [36] researched the currency flows, account creation and contract invocation for the dataset, which contains more than 3 billion transactions, over 1 million EOSIO accounts. Since timestamp was important to capture abnormal behaviors, the authors improved Money Flow Graph as follows: $EMFG = (V, E, D, w), E = \left(v_i, v_j, D_k\right), v_i, v_j \in V, D_k \subseteq D$, each edge has at least one time attributes D_k, which indicates when the transfer occurs. The improved method is also used in Account Creation Graph and Contract Invocation Graph. They proposed a method combined both account relations (e.g., the number of accounts created) and behavior similarities (i.e., the invocation time and frequency, the contract target of the transactions) to identify suspicious bot accounts from community-level.

Due to the mechanism of EOS, there is another behavior in EOSIO that each account can vote for other accounts to select 21 super nodes as block producers. In addition to the above three graphs, Zhao et al. [37] defined the account vote graph and contract authorization graph:

- Account vote graph (G10): investigating the relationship between producers and voters. The graph is a directed graph G=(V, E, W), V represents the set of nodes enrolling in the voting activity, each edge (v_i, vj) represents v_i votes for v_j, and w means the corresponding voting times for each edge.
- Contract authorization graph (G11): the vertices indicate accounts participating in contract authorization activity, the directed edges indicate one account delegates its authorization to conduct the contract of others, and w means the number of executions.

According to the results, they found that some voters often vote for the same candidates in different actions, and some voting gangs exist.

4.1.5. Zcash

Zcash is similar to Monero and Bitcoin, which can obscure the source, destination, and amounts of transactions. But only when the two addresses of the transaction are both shielded addresses, Zcash's privacy guarantees will effect, otherwise, the transfer is akin to Bitcoin transactions. The shielded addresses are based on zk-SNARKs, which are practical zero-knowledge proofs. Fig. 8 shows transaction types of Zcash. If the transaction involves shielded addresses (e.g., $t\_a_1 - to - z\_a_1$, $z\_a_1 - to - z\_a_4$, $z\_a_4 - to - (t\_a_7, t\_a_8)$), it is carried out in the JoinSplit structure. As for the interactions with the shielded pool, including

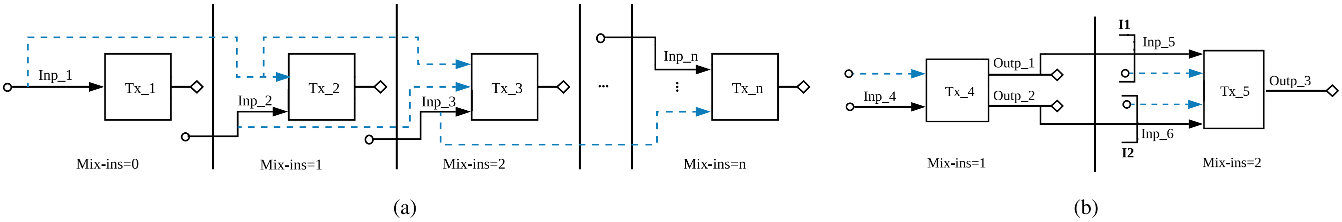


Fig. 7. Two heuristics. In (a), leveraging zero mix-ins may lead to cascade effect. In (b), an example for Output Merging Heuristic (H2). Full lines represent the real input keys being spent, dotted lines represent the mix-ins.

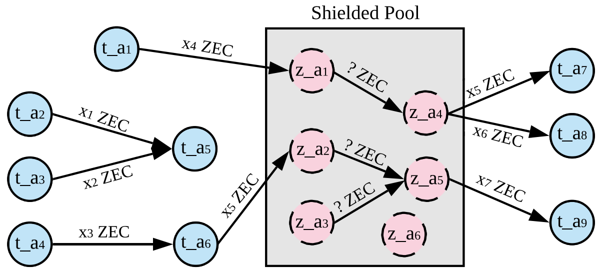


Fig. 8. The transaction types of Zcash. The circles with full lines represent the transparent addresses, the circles with dotted lines represent the shielded addresses.

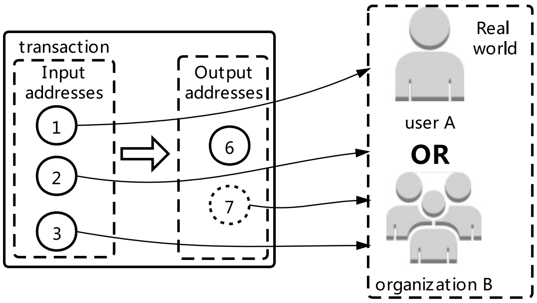


Fig. 9. Association chart of blockchain addresses.

withdrawals (e.g., $z\_a_5 - to - t\_a_9$ in Fig. 8) out of and deposits into (e.g., $t\_a_6 - to - z\_a_2$) the pool.

Quesnelle et al. [38] found that only 19.6% contained a JoinSplit operation and defined round-trip transaction (RTT), which can be understood as most coins sent to shielded addresses were send to back to transparent addresses. The results showed that 31.5% of coins entering the shielded pool related to RTTs. RTTs should meet the following conditions:

- Round-trip transaction heuristic (H6): the shielding amount in Tx_1 is equal to the de-shielding amount in Tx_2. The height of Tx_2 is higher than Tx_1.

Considering Zcash's block reward mechanism, Biryukov et al. [42] presented three clustering heuristics to link publicly visible addresses. Since block rewards were sent to shielded addresses first, there were two general patterns for miners and mining pools to use their rewards:

- Pattern T: pool converted the mined coins to a transparent address, then paying the miners.
- Pattern Z: pool pays the miners from a shielded address to a transparent address.

Three heuristics were proposed for a different pattern, the third aimed for pattern Z without a top miner (i.e., a base_tx). By using these heuristics, 84.1% of the volume of withdrawal transactions can be linked. But Pattern Z heuristic (H8) only usable for short periods (e.g., 4days, 2000blocks). The heuristics were as follows:

- Pattern T heuristic (H7): if a constant public address is found, a set of addresses controlled by the same entity can be found, and they can be linked to the shielded addresses of the mining pool by calculating the total amount of received coins;
- Pattern Z heuristic (H8): Each shielded transaction is scanned to find the transactions which have lots of outputs. If the amount of overlapping addresses between the output addresses of the transaction and the existing set of miner addresses exceeds 40, the transaction is sent by the same mining pool, and expand address set with the new addresses. Repeat this process until most of transactions can be linked to a mining pool;
- Pattern Z without a top miners heuristic (H9): all shielded transactions are considered as unknown transactions, and finding transactions which have tens of outputs. Each transaction is checked with H8, then compared the amount of the overall received coins with the value in different mining pools.

Biryukov et al. [41] proposed novel linkability techniques by using value matching and value fingerprinting:

- Transaction fee heuristic (H10): 10000 Zatoshi was the default transaction fee, so if the shielding amount was n * 10000 Zatoshi larger than the de-shielding value (n means the value makes n hops in the shielded pool), their values were unique in the observed block range, and the height of the transparent transaction was higher than the shielded transaction, the two transactions were considered to be linked.
- Fingerprinted values heuristic (H11): value fingerprint was defined as the last 7 or 4 digits in Zatoshi. if the shielding amount matched fingerprints with a de-shielding value (if 5 of the last 7 digits or the last 4 digits were the same, the two fingerprints matched), no other amount matched fingerprints with their values in the observed block range, and the height of the transparent transaction was higher than the shielded transaction, the two transactions were considered to be linked.

4.1.6. Summary

The various characteristics of different blockchain platforms bring challenges for analyzing the currency flow of suspicious transactions. For Bitcoin, graph theory is used to construct transaction graph, address graph, and entity graph, and distinguish normal and abnormal transactions from the graph based on the statistics of vertex and edge, such as degree distribution, connectivity, etc. As Ethereum adds smart contracts, many attackers use contract vulnerabilities or malicious contracts to steal Ether, so the creation and invocation graph about contracts are proposed. On the basis of the above graphs, the research on EOS adds time factor to obtain more information about transactions and nodes. The anonymity of Monero is different from Bitcoin, Ethereum, so more clustering algorithms are used to find currency flow instead of graph theory. We summarize the research in Table 6 from the following aspects: graph/heuristic category, data integrity, platform/scenarios, main techniques, and other useful analysis.

Table 6
Transaction and contract analysis.

Data analysis	Graph/ Heuristic	Data integrity	Platform/ scenarios	Main techniques	Case study	Other useful analysis
Zheng et al. [182]	×	✓	Ethereum, all data	(i) Data processing	×	(i) Providing well-processed datasets, containing six parts
Reid et al. [25]	G1, G3	✓	Bitcoin, transaction	(i) Graph theory (ii) Information correlation (iii) Visualization technique	(i) A theft of 25000 BTC	(i) Egocentric analysis (ii) Context discovery (iii) Temporal analysis
Zhao et al. [26]	G2	× block 210000 to block 314700	Bitcoin, transaction	(i) Graph theory (ii) Address clustering	(i) Mt.Gox	(i) Currency flow (ii) A new approach for clustering transactions
Maesa et al. [184]	G2, G3	✓	Bitcoin, transaction	(i) Graph theory (ii) Address clustering (iii) Time window	×	(i) Connectivity analysis over time (ii) Degree distribution (iii) Centrality analysis
Wang et al. [27]	×	× from 02/2016 to 01/2019	Bitcoin, mining pool	(i) Data correlation (ii) Data analysis	×	(i) Distribution of computing power (ii) Mining Revenue (iii) Transaction Delay (iv) Block size
Kiffer et al. [188]	G6	✓ without internal transactions	Ethereum, contract	(i) Clustering (ii) Data analysis	×	(i) Smart contract topology (ii) Contract life cycle (iii) Contract similarity
Chen et al. [29]	G4, G5, G6	✓ have internal transactions	Ethereum, contract, transaction	(i) Graph theory (ii) Cross-graph analysis (iii) Backtracking (iv) Visualization technique	(i) a malicious contract BD37 for a DoS attack (ii) 7C20 and 3898	(i) Novel methods for attack forensics and anomaly detection (ii) Ether distribution (iii) Important nodes
Bai et al. [31]	G7, G8, G9	✓ have internal transactions	Ethereum, contract, transaction	(i) Graph theory (ii) Data analysis (iii) Time window (iv) Motifs	×	(i) Degree distribution and properties (ii) The rich gets richer
Kumar et al. [32]	H1, H2, H3	✓	Monero, transaction	(i) Clustering technique	×	(i) Three new heuristics for traceability
Moser et al. [33]	H3, H4	✓	Monero, transaction	(i) Clustering (ii) Data analysis	(i) The AlphaBay	(i) Improving the previous heuristic (ii) Quantifying mining activity
Hinteregger et al. [34]	H1, H4, H5	✓	Monero, transaction	(i) Clustering (ii) Data analysis	×	(i) Generalizing the previous heuristic further
Huang et al. [36]	Enhanced G4, G5, G6	✓	EOSIO, contract, transaction, bot account	(i) Graph theory (ii) Data analysis (iii) Visualization technique	(i) Permission misuse (ii) Unchecked input and predictable state attacks	(i) Degree distribution (ii) Bots account detection from two level
Zhao et al. [37]	G4, G5, G10, G11	× the first 15 million blocks	EOSIO	(i) Graph theory (ii) Data analysis (iii) Visualization technique	×	(i) Degree distribution
Quesnelle et al. [38]	H6	✓	Zcash	(i) Data correlation	×	(i) RTT information
Biryukov et al. [42]	H7, H8, H9	✓	Zcash, mining transactions	(i) Linkability technique	×	(i) Miners linkability
Biryukov et al. [41]	H10, H11	✓	Zcash	(i) Linkability technique (ii) Direct and fingerprinted value linking (iii) Sliding block window	(i) Danaan-gift attack (ii) Dust attack	(i) Discovery of three subliminal channels for hidden transactions: the inner, outer, and pedersen subliminal channel

4.2. User identity analysis

In addition to individual users in blockchain, there are many large central nodes such as exchanges and trading websites. The regulator can discover the user identity of a dubious transaction address through some potential rules in blockchain technology, which is called heuristic. These heuristics can be used for clustering addresses of some blockchain platforms. The common rules are as follows, and Fig. 9 is an overview of the two rules.

- Multi-input heuristic (H12): because the premise of transaction is that the user has the private keys of all input addresses, so all input addresses of a transaction belong to the same user or the same institution;
- Change address heuristic (H13): because BTC cannot be split in a transaction, and Bitcoin uses change address to solve this problem, so the change address and input addresses of a transaction belong to the same user or institution. And the change address appears only once as an output address;

Through clustering method, Zhao et al. [26] analyzed 35587286 addresses collected from Bitcoin transaction records, and clustered 13062822 distinct address sets. They also used breadth-first search to quantitatively capture the currency flow to determine the most probable direction the stolen coins go in.

Since all transactions are artificial, which are similar to transactions in real life, so there are some transaction regular patterns. For example, some breakfast shops are only open in the morning, and the transaction amount is between 0 and 30 dollars. Blockchain users also follow some

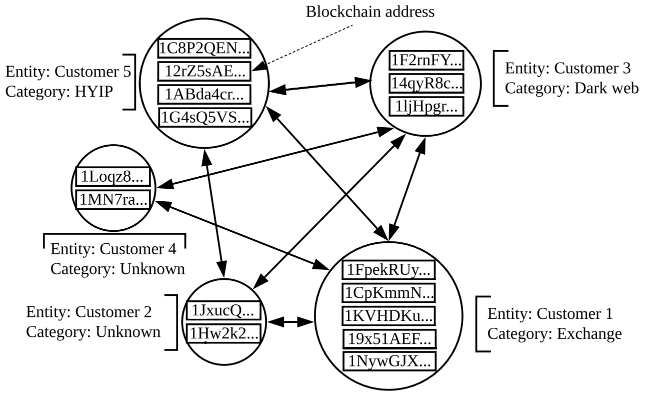


Fig. 10. An example of different categories of blockchain clusters.

transaction rules such as the range of transaction amount, transaction frequency, currency flow, connectivity, etc. Researchers can cluster addresses and make malicious user portraits through these heuristics.

After Androulaki et al. [51] analyzed Bitcoin system, they implemented a simulator that let students use Bitcoin as daily transaction currencies in the university. The authors utilized multi-input heuristic, change address heuristic and behavior-based clustering techniques to analyze daily transactions. They found that even using one-time addresses, which is the protection strategy of user privacy, it was possible to associate 40% of the student's identity with the Bitcoin address, with 80% accuracy.

Users can generate many blockchain addresses for transactions. It is difficult to find an account's identity because its addresses instead of the identity are enough for an account to use. But the regulator can resort to discussion boards, forums, and search engines, because many blockchain users may interact with important accounts and post their addresses on these places. So it is able to obtain the identity from its name tags. The characteristics of user's usage of blockchain can also be analyzed through research on the blockchain community, and these characteristics are used as features to distinguish users with different purposes of use. Moreover, the available smart contract's source code (e.g., comments, keywords) can be used to infer the identity of suspicious blockchain users.

Fig. 10 shows an example of different categories of blockchain clusters. Different categories have different labels, such as Customer 5 is labeled with HYIP (high yield investment programs), Customer 2 is labeled with unknown, meaning the cluster has not been identified [62].

We divide user identity analysis into three parts according to the research content: identifying various entity categories, analyzing user identity for special transactions, and visualization tools used for malicious transactions.

4.2.1. Identifying various entity categories

Supervisors can obtain a lot of useful information about suspicious addresses and transactions through graph theory, clustering, machine learning, etc.

4.2.1.1. Graph theory technique. Ron et al. [181] used a variant of the Union-Find graph algorithm to generate address graph and entity graph. Typical behaviors of entities are analyzed on Bitcoin network, such as the number of addresses owned by entities, the number of different entities, the number of BTC owned by addresses, transaction times, etc. They analyzed the transaction volume, BTC, and addresses of active top19 entities, and found that 55% BTC is not circulating in Bitcoin system. The flows in the two graphs had some characteristic behaviors: long consecutive chains of transactions, fork-merge patterns that may include self-loops, setting aside BTC's and final distribution of large sums via a binary tree-like structure.

Fleder et al. [189] found that people can obtain Bitcoin user information (a person's true name or username) from donation sites, forums, social networks by scraping Bitcoin addresses, so that they can associate numerous unrelated IDs with a real user. The mechanism for matching actual users to transactions uses rough information (transaction time and value) mentioned in the scraping data. But this approach only narrows the mapping range instead of mapping precisely.

4.2.1.2. Clustering technique. Meiklejohn et al. [190] used Bitcoin to assign identities to clusters, the 832 labeled data were collected from 26 exchanges and 10 wallet service accounts through buying BTC from 25 different vendors. And two heuristics were used to cluster Bitcoin addresses: multi-input heuristic (H12), change address heuristic (H13). The approach was able to identify 1.9 million public keys with real-world identities or services.

There are three types of transaction relationships in Ethereum: Ether trading transaction between EOAs, smart contract creation, contract invocation. The heterogeneous network of Ethereum is formed of two different accounts and three transaction relationship types.

Sun et al. [30] utilized Metapath2vec to learn the eigenvectors of nodes, and millions of paths were sent into the skip-gram model for training to obtain a 128-dimensions eigenvector of each node. Then, they used PCA(Principal Component Analysis) and TSNE(T-Distribution Stochastic Neighbour Embedding) algorithms to reduce dimension of the eigenvectors and utilized Birch algorithm to cluster the nodes. The entities were labeled through the tags in Ethereum, attacks, and social media networks. Finally, a novel malicious user detection method was proposed, which was based on the vector space distance of the nodes:

(i) Given two malicious nodes and their cluster, calculate the distance of the nodes in their cluster from the two malicious nodes respectively.
(ii) Select the n nodes closest to the two nodes by distance from the two malicious nodes to form two sets, which are marked as potentially malicious addresses.
(iii) Repeat these steps for two nodes at the intersection of the two sets.

Since Zcash is similar to Bitcoin, Kappos et al. [39] improved multi-input heuristic (H12) and change addresses heuristic (H13) used in Bitcoin. As for withdrawals out of and deposits into the pool, two new heuristics were proposed (H16, H17). For other entities, they used the heuristic (H6) mentioned in [38] to cluster. Through these heuristics, reducing the size of the anonymity set by 65.6%.

- Zcash-specific multi-input heuristic (H14): if more than two transparent addresses are inputs in a transaction (regardless of transaction type, transparent, shielded, or mixed), the inputs belong to the same entity.
- Zcash-specific change heuristic (H15): if a JoinSplit transaction contains one or more transparent input addresses, and only one transparent output address in the transaction, the addresses belong to the same account.
- Founder heuristic (H16): any withdrawal transactions carrying 250.0001 ZEC is done by the founders;
- Miner heuristic (H17): if one withdrawal has more than 100 transparent output addresses and one of them belongs to a mining pool, this transactions will be labeled as a mining withdrawal, and all non-pool addresses are labeled as belonging to miners.

Zhang et al. [40] improved the heuristic mentioned in [39]. They considered two outputs in a transaction meaning the transaction is pure value-transferring rather than procedural ones (e.g., mining reward distribution) or functional ones (e.g., mixing service). Two heuristics were proposed to simplify transaction network. Compared to [39], the clustering rate improved by 9%:

- Variable change heuristic (H18): if a transparent transaction consists of two outputs, and the values of the two addresses differ by 20 times, the smaller one is the change address.
- Mining heuristic (H19): miners list and mining pools list can be updated if the following conditions are met:

  (i) A transaction only has one input shielded address and contains more than 50 output transparent addresses, the transparent addresses except mining pools' transparent addresses belong to miners;
  (ii) If a coinbase transaction contains an output address with 10 ZECs, the output address belongs to a mining pool.
  (iii) A transaction only has one input transparent address and contains more than 50 output transparent addresses, the transparent addresses except mining pools' transparent addresses belongs to miners, the input transparent address belongs to the mining pool.

4.2.1.3. Graph theory with clustering technique. Pham et al. [28] constructed two graphs: user graph and transaction graph. They extracted different features from the two graphs, such as in-degree, out-degree, unique in-degree, average in–transaction, average time interval between in-transactions, active duration, etc. They detected anomalies in bitcoin network through three methods, including power degree & densification laws, local outlier factor and K-means clustering algorithm, so that the authors can classify the abnormal behaviors of users and transactions. Due to unlabeled data, there is no universally agreed method to do testing for the result of this research.

Sharma et al. [179] used the two heuristics for address linking mentioned in [190]. Besides transaction graph and address graph, they proposed a new graph for Bitcoin analysis, which was called as cluster graph:

- Cluster graph (G12): the vertices represent a cluster of addresses, which are linked by some heuristics, and the directed edges represent transactions.

4.2.1.4. Machine learning technique. Although it is important to cluster blockchain addresses, it is more important to identify entities behind potentially illegal ones. Besides clustering methods and graph theory, supervised machine learning methods are also used for detecting abnormality of blockchain addresses, such as Logistic Regression, SVM, AdaBoost, Random Forest, XGBoost, LightGBM.

Nair et al. [191] extracted out-degree, in-degree, transaction rate and other features of the address graph, and input them into Isolation forest Classifier to detect anomaly transactions. The experimental result showed that abnormal transactions account for about 1% of all transactions. However, due to the large dataset, only limited features were extracted to detect abnormal transactions.

4.2.1.5. Machine learning with clustering technique. The transaction is publicly accessible on blockchain and block explorers. Nonetheless, the malicious user identities are difficult to obtain. Several addresses may be found by scraping websites, but these addresses are not enough to support the training data. Through clustering, more training data can be obtained, thereby achieving higher accuracy.

Yin et al. [52] extracted 12 types of observations from different websites (e.g., tor market, scams, ransomware, stolen bitcoins, etc.). The dataset contained a total of 854 observations with categorical identifiers. Each observation has different number of addresses and transactions. Thirteen classifiers were trained based on 854 observations, it was found that Bagging (78.46%) and Gradient Boosting (80.76%) were the best classifiers, which were applied to classify 10000 unknown observations. According to the experimental results, 5.79% addresses and 10.02% coins were related to cyber-criminal entities. But the dataset was unbalanced, some entities were oversampled (e.g., exchange, personal-wallet), and some entities were undersampled (e.g., ransomware, stolen-bitcoins). In this paper, three clustering methods were used to enrich the dataset, including Co-spend clustering (H12), intelligence-based clustering and behavioral clustering:

- Intelligence-based clustering heuristic (H20): information is obtained from outside of blockchain, including court documents, exchanges, data leaks.
- Behavioral clustering heuristic (H21): clustering related addresses according to known patterns, which are used by software wallet, web wallet, etc.

In [63], the authors proposed new features to build a classification model. Their dataset was derived from a heuristic, naming multi-input transactions [51], which contained 26313 Bitcoin addresses with labels and owners. Then three parts features were extracted: basic statistical features (from previous work), extra statistical features (i.e., total amounts, statistical measures of transactions), and moment features (which represents the temporal distribution of transactions and transaction intervals). As for address-based scheme, the Micro/Macro F1 scores reached 87% / 86%. As for entity-based scheme, the Micro/Macro F1 scores reached 91% / 78%. They ranked the features through importance scores, and the top 10 features were as following:

- the number of transaction per day
- the mean value of the number of outputs in the spent transactions
- the mean value of the number of inputs in the spent transactions
- the number of received transactions
- the moments of transaction interval distribution
- the standard deviation of balance in BTC after each transaction
- the duration between the first transaction and the last transaction in terms of days
- the ratio of Bitcoin addresses that appear in both inputs and outputs
- the moments of received transaction distribution
- the mean value of balance in BTC after each transaction

4.2.1.6. Graph theory with machine learning technique. Besides address graph entity graph, Jourdan et al. [64] utilized discrete-time graph to analyze leakage of information about entities, which is obtained from the entity graph through temporal aggregation, the new graph was as following:

- Discrete-time graph (G13): the graph is a directed weighted bipartite graph. During the selected time period, entity transactions are considered, the edges represent the aggregation of transactions. The time window maybe a day, a week, or a month.

Motifs were introduced in Bitcoin. Direct N-motif was defined as following:

- Direct N-motif heuristic (H22): in entity graph, N-motif is a path of length 2N, starting and ending with an entity. Direct meant at least one input from each transaction (TX) is output to the previous transaction.

A case of 3-motif was shown in Fig. 11. Five features classes were proposed for entity classification: Address-specific features, analogous features, centrality-based features, temporal features, and motif features involving both edge features and vertex features. They used LightGBM(Light Gradient Boosting Machine) to classify 272 Bitcoin entities, about 30 million addresses, and global accuracy of 92% could be achieved.

Inspired by [64], Zola et al. [65] combined a novel cascading machine learning method with motifs to attach Bitcoin anonymity. Their dataset was composed of 311 entities, including exchange, service, gambling, mining pool, mixer and marketplace. Through the novel method, they can obtain an average global accuracy of 99.68%. Cascading machine learning model was introduced as following:

(i) Built three separate classifiers, based on additionally address, 1 motif, and 2 motif dataframes, to create a set of six new features for each classifier.

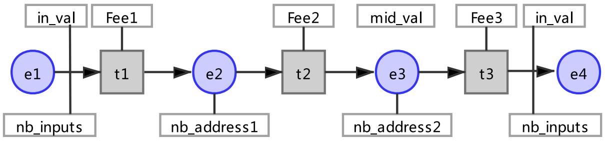


Fig. 11. A case of 3-motif features (rectangular white boxes), which contains a path of length 6 on the entity-transaction graph. The e means entity, t means transaction. The nb_inputs means the number of inputs in t1 of the motif. The in_val means the total value of inputs of t1 in BTC. Fee1 means the fees of the transaction. Mid_val means the value transferred in the middle.

(ii) Combined new features with the original features (belonged to the entity dataframe), the new classifier was generated to conduct entity classification.

In Ethereum, Bai et al. [31] also researched on motifs, which reveal the microscopic behaviors on the formation of Ethereum. They investigated 3-node motifs of user-to-user graph using the sliding window. There were 13 motifs for the directed graph, seven of them were closed-loop modes.

4.2.1.7. Summary. We summarize the main technologies used in general user identity correlation, scenarios, other analyses in Table 7.

4.2.2. Analyzing user identity for special transactions

For special transactions and powerful users in grand larceny or other events in blockchain history, the regulator can continuously monitor the follow-up transaction information of the suspicious transaction and find out the masterminds behind crimes. We divide the surveyed paper into several categories according to the address attributes of research.

4.2.2.1. Ransomware. Zhao et al. [43] performed a measurement analysis of CryptoLocker, a family of ransomware. At first, CryptoLocker addresses are expanded by Bitcoin mechanisms (i.e., multi-input heuristic (H12) and change addresses heuristic (H13)). In the period of CryptoLocker's operating, they examined ransom payment timestamps both longitudinally and transversely to distinguish trends and changes in timestamp distributions. The authors constructed a transaction graph containing 993 CryptoLocker addresses and systematically analyzed CryptoLocker's money laundering strategies. Finally, CryptoLocker was speculated connections to other Bitcoin cyber-crime, such as the Sheep Marketplace scam.

Several heuristics and tedious information collecting steps complicate the analysis of ransomware. Akcora et al. [60] automatically detect ransomware related addresses and transactions through advanced data analytics techniques (e.g., topological data analysis). They constructed Bitcoin graph as a heterogeneous network, which combined the advantages of transaction graph and address graph. The novel graph was as following:

- Bitcoin graph model (G14): the directed weighted graph is denoted as G = (V,E,B), where V is a set of nodes, including two node types: transactions and addresses. E is a set of edges, which represent BTC transfer between a transaction and an address node. B = Address, Transaction represents a set of node types. When $|B| = 1$, similar addresses are grouped into nodes, and common addresses between two nodes are put into the set of edges.

A topological data analysis (TDA) Mapper approach was used to create six filtered cluster graphs by using six features of each address. A suspicion score was assigned to an address according to the number of ransomware addresses in each cluster. According to their experimental results, their approach achieved higher precision and recall compared to the existing methods. The method of calculating the suspicion score was as following:

- Suspicion score algorithm (A1):

  (i) Scores of all addresses are initialized with 0.
  (ii) If the inclusion threshold times the amount of labeled ransomware addresses is less than the number of ransomware addresses in the cluster, and the amount of all addresses in the cluster is less than the size threshold times the number of labeled ransomware addresses in the cluster, the suspicion scores of addresses in the cluster increments by one.
  (iii) If the scores are higher than the quantile threshold, the addresses are filtered out.

4.2.2.2. Ponzi. Chen et al. [46] collected 131 Ponzi scheme contracts and 1251 non-Ponzi scheme contracts from Etherscan. To establish the detection model, they downloaded transactions and bytecodes through the APIs provided by etherscan.io. After analyzing the Ether flow graph of Rubixi (a typical Ponzi scheme contract) and the Ether flow graph of LooneyLottery (a typical non-Ponzi scheme contract), They extracted seven key features in contracts (i.e., Known rate, Balance, N_investment, N_payment, Difference index, Paid rate, N_maxpay) and some code features (e.g., opcode) to distinguish Ponzi scheme. The performance of the detection model can be improved by combining opcode features with account features.

Compared with a previous study, Toyoda et al. [56] collected above 2000 HYIP (high yield investment programs, which is also called Ponzi scheme) Bitcoin addresses from a major Bitcoin online forum, rather than 43 addresses [55]. They collected HYIP-related topics, posts with TXID in the topics, and extracted HYIP Bitcoin addresses through manually judging the context of posts. For each address, related transactions were retrieved from the blockchain. A supervised machine learning was used to judge HYIP addresses, the following features were extracted, frequency of transactions, ratio of coinbase transactions, mean value of the number of inputs, frequency of i in USD appeared in transactions, etc. The TPR(True Positive Rate) of their experiment can reach 95%.

4.2.2.3. Stolen cryptocurrency. There are many abnormal transactions in blockchain, and the masterminds may manipulate the Bitcoin price through the activities of some suspicious accounts. Through the real-time price of transactions, Chen et al. [45] divided all the accounts of Mt. Gox Bitcoin exchange into three categories: extremely high account (EHA), extremely low account (ELA) and normal account (NMA). Then they constructed user graph (G3) for three account types and analyzed the structure of graphs from various aspects, such as nodes and edges classification, clustering coefficient, degree distribution. The authors compared successive snapshots of graphs to detect important changes of graph structure, some base graphs are extracted through singular value decomposition (SVD). Furthermore, they drew the daily subgraph of the core abnormal accounts and found many abnormal transaction patterns (i.e., self-loop, bi-direction, star), which are important evidences of market manipulation in trading.

Besides features mentioned in [191], Goldsmith et al. [192] extracted some new features for analyzing hack subnetworks, which are consisted of nodes that received hacked funds. There is a fact that criminals would change their stolen BTC for fiat funds, and the authors traversed each graph from the starting nodes to exit points, which achieved the funds to cash out. The exit point was defined as terminal nodes, which were defined based on the radio of sending to receiving activity (ρ parameter). The number of transactions represented how active the hacker was. For each subnetwork, the authors extracted some new features:

- Logarithm of hack balance
- Logarithmic difference of the average percent of funds
- Second difference of the average percent of funds
- Logarithmic difference of the standard deviation of the percent of funds
- Number of terminal nodes as function of ρ.

And they found that 6 hacks were carried out by 2 hacking groups.

Table 7
General user identity correlation.

Identity correlation	Graph/ Heuristic	Data integrity	Label methods	Scenarios	Main technologies	Other analysis
Androulaki et al. [51]	H12, H13	× the first 140000 blocks	Simulating the usage in University	Bitcoin	(i) Heuristics (ii) Behavior-based clustering	×
Ron et al. [181]	G2, G3	✓	Address analysis after clustering	Bitcoin	(i) Union-Find algorithm	(i) Statistics calculated over two graphs (ii) The graph of the largest transactions
Fleder et al. [189]	G1, G3	✓	Scraping from emails, forum, donation sites	Bitcoin	(i) Graph theory (ii) Identity association	(i) Silk Road (ii) TX fingerprinting
Meiklejohn et al. [190]	H12, H13	✓	By Trading BTC, then expand by clustering	Bitcoin	(i) Clustering technique	×
Sun et al. [30]	×	× the first 10 millions TXs without internal	Through tags in Etherum, attacks, social networks	Ethereum	(i) Metapath2vec (ii) Node Embedding (iii) Birch algoritms (iv) PCA, and TSNE	(i) A new malicious user detection approach
Kappos et al. [39]	H6, H14, H15, H16, H17	✓	By Trading ZEC and scraping from websites	Zcash	(i) Clustering technique	(i) The Shadow Brokers
Zhang et al. [40]	G1, H18, H19	× from height 29400 to 29500	Clustering	Zcash	(i) Graph theory (ii) Clustering	(i) Poor use of shieldedpool
Pham et al. [28]	G1, G3	× from beginning to 04.07.2013	Clustering	Bitcoin	(i) Graph theory (ii) k-means Clustering (iii) Network analysis	(i) Features engineering for two graphs and k-means (ii) Statistics
Sharma et al. [179]	G1, G2, G12, H12, H13	× 300000 blocks	From known merchants, ransomware	Bitcoin	(i) Graph theory (ii) Heuristics (iii) Address linking	(i) Computational graph analytics (ii) Graph pattern matching
Nair et al. [191]	G2, G3	×	Machine learning	Bitcoin	(i) Unsupervised machine learning	(i) 1% of transactions are abnormal
Yin et al. [52]	H12,H20, H21	× dataset is based on observations	From tor market, scams, ransomware, mixing, stolen BTCs, exchange, gambing, merchant, wallet, mining pool	Bitcoin	(i) Scikit-Learn classification	(i) 80.76% accuracy
Lin et al. [63]	H12	× leverage existing dataset	leverage dataset and derive from several heuristic	Bitcoin	(i) Features engineering (ii) machine learning	(i) 87% and 91% F1 scores for address-based and entity-based scheme
Jourdan et al. [64]	G2, G3, G13, H12, H22	✓	From WalletExplorer	Bitcoin	(i) Discrete time graph (ii) LightGBM (iii) Motifs	(i) 92% accuracy
Zola et al. [65]	G2, G3	✓	From WalletExplorer	Bitcoin	(i) Cascading machine learning (ii) Motifs	(i) 99.68% accuracy

4.2.2.4. Phishing. The attackers have been making a notable amount of money through phishing scams on blockchain. The total number of labeled phishing addresses posted on etherscan.io and the addresses of Ethereum are 2041 and 500 million, respectively, so extreme data imbalance is one of the biggest challenges for phishing detection. Wu et al. [47] proposed an approach to detect phishing scams on Ethereum its historical transaction records. Due to the aforementioned issues of extreme data imbalance, they adopted one-class SVM(Support Vector Machine) by turning this problem into a single classification task, which was an unsupervised anomaly detection approach. Through the novel trans2vec algorithm, the transactions were embedded into node representation vectors, which were used as feature inputs for the task of phishing scam detection. The central idea of tran2vec algorithm (A2) was as following:

(i) Tran2vec adopted Skip-gram in random walk strategies based on transaction amount and timestamp of each edge.
(ii) The amount value of transaction implies the relationship between the two involved nodes, and if the timestamp of transaction is later, the greater the influence on the relationship of the nodes.

Ostapowicz et al. [61] automatically collected available data about fraudulent accounts and transactions and utilized supervised learning methods to detect them. By using Etherscan API, over 2200 wallets were marked as "Hack/Phishing". For each wallet, 13 features were extracted from the transactions, such as amount of incoming/outgoing transactions, amount of unique incoming/outgoing transactions, average value of the incoming/outgoing transaction, average gas price, etc. Three supervised learning algorithms (i.e., SVM, RF(Random Forest), XGBoost(eXtreme Gradient Boosting)) were used to conduct early warning systems.

4.2.2.5. Spamming scheme. Maesa et al. [136] aimed to verify the conjecture: the small world theory discrepancies are caused by artificial users' behavior. They proposed pseudo-spam transactions, pseudo-spam chain, and the research scope ranges from the case study to generic chains. They analyzed the outliers in the in-degree distribution of the Bitcoin users graph to support their conjecture. Pseudo-spam transactions satisfy the following properties:

- Only one input address in the transaction.
- At least two output addresses in the transaction.
- The receiving amount of two output addresses is 0.00001 BTC.

Besides pseudo-spam transactions, sextortion emerged in 2018, a new spamming scheme. The attacker utilized sexual extortion messages requiring payments in Bitcoin. There were bitcoin addresses in sextortion spams. Paquet-Clousston et al. [53] projected addresses mentioned in 4340736 sextortion spams onto transaction graph, which helps them to investigate and track monetary flows. They found sets of emails with high textual similarity through email bucketing heuristic (H23), which used two points to determine whether the email should be in the same bucket:

- The last words of emails are same;
- In all pairs of template emails representing a bucket, if the Jaccard similarity is higher than a threshold t, the two buckets are merged.

1810 buckets were produced and 96 buckets represented 99.6% of the total number of emails. They extracted 12533 Bitcoin addresses in 96 buckets, but only 245 addresses received funds. These addresses are associated with clusters in transaction graph. From the association between Bitcoin addresses and sextortion campaigns, they analyzed passwords breaches, reuse of bitcoin addresses across campaigns, holding periods and monetary flows. But their dataset represented a small proportion of a global campaign.

4.2.2.6. Sex ads. Blockchain is also used for sex ads. S. Portnoff et al. [54] designed two technique that links some sex ads to Bitcoin transactions and wallets:

(i) Any two ads were determined whether they were written by the same or different author through two supervised learning stylometry models (i.e., WritePrints, Jaccard).
   - WritePrints Limited: the model used counts of characters, words, punctuation, etc.
   - Jaccard and structure: the model used text-based features, including word unigrams, word bigrams, character n-grams, parts of speech, proper names, the location and spacing of line breaks in the posts.
   - Evaluation: 1164663 unique ads were used to verify their method, and achieving 89.54% true positive rate.

(ii) since GoCoin processed all transactions on Backpage, the authors saved snapshots of mempool state and obtained 16767921 transactions. The authors proposed the following heuristic:
   - Linking ads to transaction heuristic (H24): if the transaction amount equals the cost of posting the ads and the difference between the timestamp when the transaction first occurred in the mempool and the timestamp of the ads' appearance on the page, the transaction may belong to the ads.

4.2.2.7. Hidden services. Tor hidden services can make server-side anonymity, which provides criminals with a safer opportunity for crimes. Fortunately, the criminals using Bitcoin for Tor may be tracked through some approaches in [57]. They crawled Bitcoin addresses from onion landing pages of various hidden services. The beginning number of Bitcoin address is 1 or 3, so the used regex is shown as Eq. (2).

$$* [13][a - km - zA - HJ - NP - Z1 - 9]25, 34 \tag{2}$$

After collating the Bitcoin address, they obtained 88 unique Bitcoin addresses that represent the Tor hidden services dataset. The addresses and identities of users are collected from online social networks (i.e., Twitter and BitcoinTalk forum). Then, the authors used wallet-closure analysis to expand Bitcoin addresses. Due to the mixing services or CoinJoin transactions, they performed a cleaning process: all closures sharing at least one address were merged, at the same time those closures were removed, ending up with unique closures that had no intersections and were mutually-exclusive. As for the link between users and hidden services, they cross-matched transactions for hidden services and transactions for users. There are two main limitations to this study. The first one is that users may create fake online identities to hide their real identity. Second, the cleaning process for mixing services may exclude users who did not use mixing services.

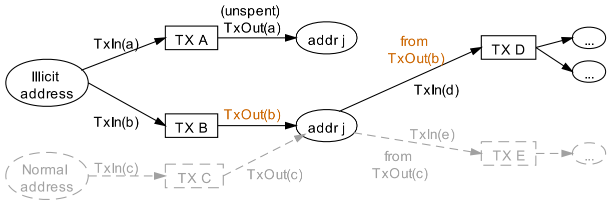


Fig. 12. An example transaction graph starting from an illicit address. Solid lines means that TX inputs and outputs are involved in illicit currency flow. Dot lines means they are not involved and will not be traced.

The Dark Web is a major distribution channel of unlawful goods, illicit financial transactions and harmful information. Perpetrators can hide their identities in the dark web. Lee et al. [59] designed MFScope to understand cryptocurrency abuse on the dark web. MFScope collected over 27 million dark webpages and obtained 10 million blockchain addresses, including Bitcoin, Ethereum, Monero. After filtering out invalid blockchain addresses, 5440 Bitcoin, 50 Ethereum, 61 Monero addresses are obtained. Then, they manually checked whether those addresses are used for trading illicit goods by reviewing up the webpages containing the address. The Cross-domain Analysis module in MFScope used the illicit addresses as keywords to conduct a Google search and manually analyzed the search result to obtain extra information about those addresses. Then, they performed taint-based currency flow to track the money and quantify the number of currency flow to the destinations. An example of the new transaction graphs was shown in Fig. 12.

- The illicit transaction graph (G15): the graph is a rooted directed graph with an illicit address as a root node, which is different from the graphs of the previous studies. Since TX E spent the output TxOut(c), which does not carry currency from the illicit address, TX E is abandoned and TX D is the net transaction in this currency flow.

4.2.2.8. Several attributes. Due to the pseudo-anonymity of Bitcoin, the dark web uses monero and Zcash more than Bitcoin, and Bitcoin is more likely to be involved in mixing transactions. Eldefrawy et al. [58] collected more than 2.3 million Bitcoin addresses from the dark web, and Only 47697 Bitcoin addresses were labeled with tags that cover suspicious, malicious and illegal activities. To identify mixing transactions (i.e., CoinJoins), the authors used a heuristic based on their characteristics. CoinJoin identification heuristic (H25) is as following:

(i) The number of participants in the transaction is more than half the number of outputs.
(ii) The number of participants in the transaction is less than or equal to the number of inputs.
(iii) There should be one or more than one match between inputs and outputs.

The match can be expressed as Eq. (3). Some participants served as liquidity providers, in exchange, they received a percentage P of the most value (MCV) in the transaction. N is the number of participants, and chng is the change address. (1+(n-1)P) means that a liquidity provider may receive fees from other participants. Their study also existed limitations. First, the Bitcoin addresses were collected from the dark web between June 2016 and December 2017, which meant that the coverage of dark web was fragmentary. Second, only a few addresses can be labeled with suspicious or malicious activities. Third, it was unable to cross-check the accuracy of their ratiocination.

$$\forall inpt \in S : inpt \in [MCV \cdot (1-P)+chng, \quad MCV \cdot (1+(n-1)P)+chng+fees] \tag{3}$$

4.2.2.9. Summary. For special transactions, how to obtain the relevant labeled address is a big challenge for researchers. We compare the surveyed research from the following aspects: labeled data source (difficult to obtain labeled data), application scenarios, amount of related addresses, main technologies, which are shown in Table 8.

4.2.3. Visualization tools used for transactions

Spagnuolo et al. [44] parsed the blockchain through their proposed framework, Bitlodine, which clustered blockchain addresses that likely belonged to the same user or organization, labeled users automatically with information on their identity and actions. Then they utilized transaction graph (G1) and user graph (G3) to visualize complex information extracted from the Bitcoin network. Bitlodine was verified by several real-world use cases, such as Silk Road cold wallet, CryptoLocker ransomware. But their work has the assumption that owners do not share private keys.

McGinn et al. [48] designed a tool for dynamically visualizing Bitcoin transactions, mempool, blockchain, peer, which was consisted of transaction and address graphs. The visualizations can help to discover unexpected transaction patterns such as money laundering and several distinct denials of service attacks on the Bitcoin network. The tools allowed researchers to rapidly understand the structure of such behavioral patterns for accelerated analysis and classification investigation. The new graph was as following:

- High-fidelity visualizations (G16): transactions are represented as nodes with black color, the size of node is related with the amount of coinbase transaction, the inputs are represented as nodes with orange color, the outputs are blue nodes, the size is depended on value. A gray associative edge represents the address

In [193], BiVA was designed through two novel information theory-based algorithms on the new graph, which is called 2-mode network. First, they improved the Markov entropic centrality, and realized an entropic centrality algorithm for asymptotic time, which aimed to the impact of a particular wallet address. Second, they utilized AgglomerativeClustering to look at the probability to visit neighbors of nodes with low entropic centralities. The new graph was as following:

- 2-mode network graph (G17): the graph contains two types of nodes with two colors: transactions and addresses, which are labeled with two different labels. The directed edge represents the direction of a transaction

Kinkeldey et al. [49] designed BitConduite, which was a visual exploration tool supported for aggregation of Bitcoin addresses to high-level entities. As for entity activity, eight measures were available on BitConduite, such as time of first/last transaction, time active in days, number of transactions, etc. Five linked views were utilized to compose the tool:

- Filter view: provide the temporal distribution of TXs.
- Tree view: represent a hierarchy of partitions.
- Cluster view: automatically create entities with similar characteristics by k-means.
- Entity browser.
- Transactions view: show a timeline about transactions for a single entity.

We compare the four Visualization tools in terms of five aspects: the used graph/heuristic category, data source, application scenarios, main technologies and other analyses in Table 9.

For the transaction or identity analysis on the transaction layer, due to the lack of ground truth information, the correctness of much research could not be verified.

5. Network traffic analysis on the application layer

Through the analysis and correlation on the network layer and transaction layer, some users' IP can be mapped to blockchain addresses. The application layer of blockchain includes normal users and users with abnormal behaviors. Combined with passive monitoring methods, it can assist the association between users' IP and blockchain addresses.

The main objects of network traffic analysis on the application layer are abnormal user behavior, abnormal node behavior caused by malicious user, blockchain services and tools. We will introduce the surveyed papers from the above three aspects.

5.1. User behavior analysis

Due to the emergence of smart contracts, the function of blockchain is not limited to cryptocurrencies, but a complete world of DApps. As of May 2020, there are about 3110 DApps on blockchain. DApps cover a broad spectrum of functions such as finance, gambling, social, gaming, art, etc.

Although the communication process is encrypted, it is possible to identify different DApps or different user behaviors by combining passive measurement, machine learning and other methods. DApps data and records of operation are cryptographically stored in the decentralized blockchain. By continuously paying attention to behaviors of specific suspicious users on DApps, the regulator can create suspicious user profiles and infer their hobbies, lifestyle habits, life status and other information to supervise them.

Researchers study the algorithm performance and recognition accuracy through different machine learning frameworks, combined with empirical parameter setting, analysis of target error and parameter selection methods. Research repeatedly experiment and determine the optimal model to identify abnormal transactions and suspicious behaviors of users.

Shen et al. [69] performed feature fusion of time series, packet length series, burst series, and transformed them into high-dimensional features. After training with SVM, KNN(K-NearestNeighbor), and RF machine learning algorithms, the accuracy of DApps traffic classification reaches more than 90%. But the classification of DApps encrypted traffic is too coarse-grained. The training time and testing time of this method are much longer than other methods, and the efficiency of the classifier becomes a major problem.

Aiolli et al. [70] selected several Bitcoin wallets and user behaviors that existing in each wallet to classify. They extracted time statistics features of different flow directions, such as length, maximum, minimum, average, etc. They used SVM and random forest to train and identified user behaviors of Bitcoin wallets, the accuracy reached 95%.

5.2. Abnormal node behavior caused by user

In addition to monitoring the abnormal behavior of the user, it can also detect the abnormal behavior of the node controlled by the user. Use the behavior sequence to represent the node and classify it as a normal or abnormal node to help the supervisor follow-up supervision.

In the public blockchain, there are millions of nodes. Some nodes may try to cheat for illegal interests and have abnormal behavior patterns. Tang et al. [66] extracted sequence data from the activities of the peer to represent peer behaviors and proposed a novel deep learning based method to classify blockchain peer behavior. Behavior sequence consists of the daily transaction amounts of a peer. Peers were classified into several categories according to the jitter levels of transaction amounts in the recent 90 days. Harmful or malicious peers

Table 8
Special user identity correlation.

Special transactions	Graph/ Heuristic/ algorithm	Data source	Application scenarios	Amount of related addresses	Main technologies	Other analysis
Zhao et al. [43]	H12, H13, G1	Investigate reddit thread, and clustering	Ransomware, CryptoLocker	993 addresses	(i) Graph theory (ii) Visualization techniques (iii) Timestamp analysis	(i) The Sheep Marketplace scam
Akcora et al. [60]	H12, G14, A1	From three widely adopted studies	Ransomware families	24486 addresses from 27 ransomware families	(i) Graph theory (ii) Topological data analysis (iii) Time window (iv) Topological data analysis (v) Clustering	(i) Ransomware behavior similarity analysis (ii) Filtering approaches for address elimination
Chen et al. [46]	×	From Etherscan	Ponzi	131 Ponzi and 1251 non-Ponzi schem contracts	(i) XGBoost (ii) Features extraction	(i) Detect other Ponzi scheme contracts
Toyoda et al. [56]	H12, H13	Scrape and collect from BitcoinTalk	Ponzi, HYIP	Above 2000 HYIP addresses	(i) Supervised machine learning technique: the rate conversion and the sampling technique	×
Chen et al. [45]	G3	Leaked Mt.Gox data	Mt.Gox	14916 abnormal accounts and 3 million TXs	(i) Graph theory (ii) Static and temporal network analysis (iii) SVD	(i) Account classification based on transaction behaviors (ii) Abnormal transaction patterns analysis
Goldsmith et al. [192]	G1	From professional crime investigators	Hack subnetworks	(ii) hacking groups containing 6 hack subnetworks	(i) Graph theory (ii) Features engineering	(i) Hacking group Alpha (ii) Hacking group Beta
Wu et al. [47]	G1, A2	From authoritative websites: EtherScamDB, Etherscan	Phishing	1259 phishing addresses	(i) Network embedding: tran2vec (ii) Random Walks (iii) search strategies: amount-based and time-based (iv) SVM	×
Ostapowicz et al. [61]	×	From Etherscan	Hack/Phishing	2200 wallets	(i) Supervised learning techniques	(i) Sensitivity analysis for dependence of particular explanatory variables
Maesa et al. [136]	×	Detecting by the definition of pseudo-spam	Pseudo-spam	578316 PS-TXs and 3805 PS-chains	(i) Topological techniques	(i) Pseudo-spam chain statistic analysis
Paquet et al. [53]	G1, H12, H23	Setting up a tailored spam filter	Sextortion spams	Above 4 million sextortion spams	(i) Clustering techniques (ii) Semantic analysis techniques (iii) Data analysis	(i) Spammers' pricing strategy (ii) Reuse of addresses across campaigns (iii) Sextortion Revenues (iv) Monetary flow
S.Portnoff et al. [54]	×	From Backpage and GoCoin	Sex ads	Above 1.9 million unique ads and 0.47 million authors	(i) Heuristics (ii) Labeling techniques: WritePrints, Jaccard (iii) Linking techniques	(i) Case study: 33 Backpage ads (ii) Price reconstruction for ads
Jawaheri et al. [57]	×	Regex search on the landing pages (Ahmia), malicious activities	Hidden Services	88 hidden services, 4183 twitter, and 40970 forum	(i) Linking techniques (ii) Cross-matching techniques	(i) Economic activity: money flow, operational lifetime
Lee et al. [59]	H12, H13, G15	Crawling pages from Ahmia and FreshOnions, extracting address from these pages	Dark web	88 illicit addresses, 4471 possible illicit addresses, 884 legitimate addresses	(i) Clustering techniques (ii) Graph theory (iii) Cross-domain analysis (iv) Taint-based analysis	(i) Quantifying illicit financial flows (ii) Bitcoin investment scam (iii) Trafficking (iv) Revealing hidden financial hubs
Eldefrawy et al. [58]	H25	From previously published onion datasets, DNS resolver logs, web crawl data	Ponzi, hacker, drugs, ransom, casino, etc.	47697 addresses indicative of suspicious activities	(i) Clustering techniques	(i) Detect mixing transactions (ii) Bitcoin neighborhood analysis, including entire blockchain and Dark web

vibrated fiercely, which was more difficult to predict. PeerClassifier was an end to end neural network to classify and predict behavior pattern in a consortium blockchain network.

Due to the huge scale of blockchain, researchers need to spend a tremendous amount of effort, if they want to manually study the behaviors of all nodes. Huang et al. [67] proposed a novel algorithm termed Behavior Pattern Clustering, which automatically clusters the behaviors of all nodes into categories. The experimental results showed that their proposed cluster center initialization can achieve better performance than random initialization. The main characteristics of BPC are as follows:

5.2.0.1. Behavior pattern clustering, h26. BPC sorts the sequences data of node behaviors, and selects k sequences uniformly from the sorted

Table 9
Visualization tools for transactions and identities.

Visualization tools	Graph/ Heuristic	Data source	Application scenarios	Main technologies	Other analysis
BitIodine [44]	G1, G3, H12, H13	Transactions: local Bitcoin client. Labeled data: scraping from BitcoinTalk forum, Bitcoin-OTC marketplace, Casascius, blockchain.info	Transactions, cluster	(i) Clustering techniques (ii) Visualization techniques	(i) Silk Road black market (ii) CryptoLocker ransomware (iii) Analysis with Gephi
A top-down visualization [48]	G16	Transactions: raw blockchain database	Transactions, mempool, peer	(i) Data association (ii) Visualization techniques	(i) Transaction patterns: automated laundering operations, the evolution of a denial of service attacks (tumor)
BiVA [193]	G17	Transactions: from Bitcoin Core	Transactions	(i) Graph theory (ii) AgglomerativeClustering (iii) Breadth first search	(i) Bitcoin flow traceback (ii) Case study: Extorition of Ashley Madison breach victims
BitConduite [49]	H12, G2	Transactions: Bitcoin Core client. Labeled data: from websites: Walletexplorer, blockchain.info	Entity, temporal distribution of TXs, cluster, transactions	(i) Clustering (ii) Visualization (iii) Data association	(i) Case study: Classifying entities from 2009 to 2011, Halving Day 2016 (ii) Expert workshop about tool

list; It utilizes DTW(Dynamic Time Warping) distance between sequences; BPC selects the sequence with the smallest distance to its [n/k]th nearest neighbor in the cluster as the cluster center.

Biryukov et al. [68] proposed a novel research idea to identify transactions through network traffic analysis. They clustered transactions based on the message propagation mechanics instead of analyzing historical transactions. The novel process was as following:

(i) For a transaction, consider the first N nodes which relayed it to their listening node. The weight of the first forwarded node is the highest, and the weights of other nodes decrease in turn.
(ii) Utilizing the Pearson correlation to calculate a matrix, which represents a block-diagonal matrix corresponding to the transaction source.
(iii) The matrix is clustered using the K-means algorithm so that transactions issued from one device can be clustered through a passive method.

5.3. Blockchain services and tools

Blockchain services and tools provide users with a variety of services based on the characteristics of blockchain. Supervisors can start from these services and find information related to malicious users for association.

In Fig. 2, there is a service called Ethereum Name Service (ENS) in Ethereum, which is similar to Domain Name Service (DNS). Beres et al. [180] focused on quasi-identifiers of Ethereum users (i.e., the active time in a day, the gas price selection, and the location in the transaction graph). They collected addresses belonging to regular users from Twitter, Humanity DAO, and Tornado Cash mixer contracts. Combined with the information in the Ethereum Label Word Cloud, Ethereum addresses were labeled with service category, such as exchange, gambling, stablecoins. The transaction timestamps of the daily transaction activity showed the account owner's daily activity patterns. Transaction graph analysis was used to characterize users (e.g., interaction with services or addresses, revealing address clusters when transferring funds). They listed three heuristics:

- The deposit and withdraw can be linked if an address is both in the deposit and withdraw.
- The deposit and withdraw can be linked if a deposit-withdraw pair unique and manually set gas prices.
- The addresses can be linked if there is a transaction between a deposit and a withdraw address in a mixer.

Tokens are an important tool in blockchain and involves lots of money. Users usually use third-party tools to manipulate tokens. They may use wallets to transfer tokens, leverage exchange markets to purchase or sell tokens and employ blockchain explorers to check their transactions.

Frowis et al. [71] proposed two methods to recognize token contracts by analyzing Ethereum virtual machine bytecode:

- The first method relies on the method IDs of standard interfaces, but this method is prone to both false positives and false negatives.
- The second method applies symbolic execution and taint analysis to detect the pattern of token transfers, but this approach suffers from the limitations of symbolic execution and their pattern definition.

Chen et al. [72] investigated inconsistent token behaviors with regard to ERC-20, the most popular token standard. Inconsistent behaviors lead to user confusion and financial loss. They proposed a method to automatically detect inconsistency behaviors by contrasting the behaviors derived from three different sources:

- The manipulations of core data structures recording the token holders and their shares.
- The actions indicated by standard interfaces.
- The behaviors suggested by standard events.

5.4. Summary

We summarize the research of network traffic classification on the application layer from three aspects. By classifying DApps and DApps user behaviors, the Supervisor can acquire sensitive DApps or detecting suspicious user behaviors, supervising them for a long time to obtain additional information about suspicious users; Discover anomalous nodes by detecting the abnormal sequence behavior of user-controlled nodes, thereby expanding the set of abnormal node to make the monitoring scope wider, and combine transaction analysis to obtain suspicious currency flow; Through obtaining the user information from the blockchain services or tools so that the suspicious address obtained on the transaction layer can be associated with the real world user, which helps Supervisors to find real abnormal users. We give an overview of the surveyed studies with respect to application scenarios, main technologies and description in Table 10.

6. Discussion

The blockchain attracts a large number of users and developers due to its unique characteristics, such as data security, long-term retention,

Table 10
Network traffic analysis on the application layer.

Network traffic analysis	Application scenarios	Main techniques	Description
Shen et al. [69]	DApps classification	(i) Feature fusion techniques (ii) Features engineering (iii) Machine learning	(i) 15 representative DApps of diverse categories (ii) Long time for building a model and testing
Aiolli et al. [70]	Bitcoin wallets and user behavior	(i) Features engineering (ii) Machine learning	(i) 8 Bitcoin wallets covering android and IOS, 3 common user behaviors
Tang et al. [66]	Peer behavior	(i) Behavior sequence extraction (ii) Hierarchical stacking of LSTM	(i) Classify all peers into 3 classes according to the jitter level of behavior sequence
Huang et al. [67]	Node behavior	(i) Behavior pattern clustering (ii) Similarity measure selection: DTW distance	(i) Automatically cluster behavior pattern of nodes (ii) Evaluate several potential sequence similarity measures
Biryukov et al. [68]	TXs clustering based on network traffic analysis	(i) Passive monitoring (ii) Clustering algorithm	(i) Cluster TXs based on the first N nodes introducing them into network (ii) Utilize transaction propagation
Beres et al. [180]	ENS, TXs, accounts and mix services on Ethereum	(i) User profiling techniques based on user quasi-identifiers (ii) New heuristics (iii) G1	(i) The new heuristics for linking mixer deposits and withdraws
Frowis et al. [71]	Token systems detection in Ethereum	(i) Combine symbolic execution and taint analysis (ii) G6	(i) Public bytecode analysis: the characteristic program behavior and the common interface of ERC-20
Chen et al. [72]	Inconsistent behaviors detection of tokens in Ethereum	(i) Mapping variables recognition algorithm (ii) Trace recording techniques	(i) Analyze the reasons of inconsistent behaviors (ii) Case studies: HYDRO, SMT, ZXBT, GTN, Tablow Club, and MCRT

etc. The number of industries and applications based on blockchain is increasing. Blockchain is a double-edged sword. On the one hand, it brings convenience and protects the privacy of people. On the other hand, it is used to conduct illegal behaviors and acts of terrorism, which threatens the security of normal users and even countries. So how to detect malicious users and supervise them need further research urgently.

In this section, we will discuss blockchain supervision from three aspects: the relationship between the formulation of traditional law and blockchain, challenges, and future directions.

6.1. The relationship between the formulation of traditional law and blockchain

In addition to the research on new technologies for blockchain supervision, the relationship between blockchain and traditional law should also be considered. At present, there are almost no laws on blockchain security supervision in most countries.

As an emerging technology, if blockchain needs to be independent of the central authority of states and deviate from the scope of traditional law, one of the following conditions should be fulfilled.

(i) Blockchain sets the conditions for joining the chain: only users with formal identities, no criminal records and dedicated to the development of the blockchain can join the blockchain.
(ii) Blockchain has a strong governance and regulatory framework. When a security incident occurs, blockchain can directly supervise the users of malicious nodes, and regulate behaviors of blockchain users, protecting the rights of the states and users;
(iii) The legitimate rights of the states and most users are not threatened by blockchain technology, attackers or applications. If the rights are threatened, traditional laws cannot mitigate the threat or reduce losses;

However, blockchain is a P2P network, and the nodes in blockchain are equal. Anonymity and openness are advantages of blockchain. It is unrealistic to restrict the participation of nodes or add the governance and regulatory framework to blockchain, and the security incidents that have occurred in blockchain have caused huge economic losses to the states and users. Legitimate rights and interests have been damaged. Therefore, as for the supervision of blockchain, the states need to formulate relevant laws, and different laws need to be formulated according to the conditions of different countries.

6.2. Challenges

After investigating the attacks in blockchain and summarizing the existing technology of blockchain security supervision, we see that there are still some challenges, including:

(i) Nodes can join or exit blockchain network at will. Attackers can create new nodes at any time with a low cost, such as by changing IPs. How to associate the new attacking nodes with identified harmful nodes needs to be resolved.
(ii) New anonymous services, blockchain anonymity mechanisms, and mixing services technologies also bring challenges to blockchain security supervision.
(iii) Although blockchain addresses and related user identity can be crawled on some well-known websites or forums, this information may be false identity processed by attackers. And compared to all transactions in blockchain, the obtained user identity information only accounts for a small part, and it is difficult to find a deeply hidden attacker.
(iv) As for user identity analysis, due to the lack of ground truth information, the correctness of much research could not be verified.
(v) For the detection of threat information on blockchain, real-time monitoring can be achieved by establishing a browser that analyzes the additional information of transaction. For decentralized applications with message dissemination function, there is no good method to supervise politically sensitive content or other harmful information.

6.3. Future directions

In response to the above challenges, we propose some future directions, in addition, we discuss several possible research approaches.

(i) *Research on abnormal node detection technology:* Since Public Chain cannot directly restrict the access of harmful node, just like Private Chain or Consortium Chain, it can only detect as early as possible through abnormal node detection technology and provided to relevant regulatory authorities for supervision or shielding. On the network layer, we can obtain the routing table structure of abnormal nodes and their connection relationships, analyze the characteristics of this information, and apply methods such as machine learning or deep learning to other nodes in the network, thus finding the latest abnormal nodes.

(ii) *Based on the similarity in fields:* In order to deal with new anonymous services, blockchain anonymity mechanisms, and mixing services technologies, it may be possible to find similarities from the existing research, then expand the technology to solve the new problem. We can also use transfer learning (e.g., instance based TL, feature based TL, parameter based TL, and relation based TL) to solve problems in the new field.

(iii) *Acquisition and verification of label data:* In order to avoid problems such as less labeled data about user identity and unknown accuracy, we can study new data retention technologies, continuously expand the label database, and combine data association technology to automatically verify the authenticity of label data from multiple aspects. The new data retention technology should also meet the real-time condition. When a new suspicious transaction or blockchain address appears, it can quickly locate the user identity and reduce the loss to normal users and the country.

(iv) *Threatening DApps and content detection technology:* Due to the large scale of DApps, it is difficult to detect threat information for each DApp. The essence of DApp is the smart contract and it is recorded on blockchain. First, we can detect malicious smart contracts through symbol execution technology, such as finding sensitive bytecodes in smart contracts, and executing these methods through simulation to see if there are any abnormal operations, thereby filtering out a batch of malicious DApps. Second, we can process information through text categorization technology and block harmful information in time.

(v) *The balance between security supervision and user privacy:* Blockchain supervision will definitely bring issues of user privacy. The malicious IPs filtered out maybe not completely correct, and normal users may be mistaken as malicious users. In the process of legal interference, it is necessary to consider the degree of intervention, which can not infringe on the privacy of normal users, but also can supervise malicious users. How to achieve a balance between security supervision and user privacy has not yet been solved, which is also one of the future directions.

7. Conclusion

In this article, we present a comprehensive survey, focusing on the technologies for security supervision of blockchain. Preliminary to summarizing the technologies, we provide a review of blockchain technology and security issues according to the four-layer system architecture. In addition, the major content of security supervision of blockchain is explained. The technologies related to security supervision are comprehensively and systematically summarized from the following perspective: node discovery technology on the network layer, data analysis technology on the transaction layer, and network traffic analysis on the application layer, with a specific emphasis on the transaction layer (graph theory, heuristic, etc.). Various visualization tools have been introduced and compared. Based on our comprehensive survey, we identify the challenges and the future directions of this field. We also propose research methods that may be useful for future directions. We hope that this survey is expected to contribute to the development of this research area, and serve as an efficient guideline for exploring potential research directions.

CRediT authorship contribution statement

Yu Wang: Conceptualization, Methodology, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration. **Gaopeng Gou:** Conceptualization, Writing - review & editing, Supervision, Project administration. **Chang Liu:** Conceptualization, Writing - original draft, Writing - review & editing. **Mingxin Cui:** Supervision. **Zhen Li:** Supervision, Project administration, Funding acquisition. **Gang Xiong:** Conceptualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported by The National Key Research and Development Program of China (No.2020YFB1006100, No.2020YFE0200500 and No.2018YFB1800200) and Key research and Development Program for Guangdong Province under grant No. 2019B010137003.

References

[1] Best DM, Bohn S, Love D, Wynne A, Pike WA. Real-time visualization of network behaviors for situational awareness. In: International symposium on visualization for cyber security; 2010.

[2] CoinMarketCap, https://coinmarketcap.com.

[3] Gartner. Top 10 strategic technology trends for 2017, https://www.gartner.com/en/information-technology/insights/trends-predictions.

[4] Digital marketing ramblings, https://expandedramblings.com.

[5] Swan M. Blockchain: Blueprint for a new economy. O'Reilly; 2015.

[6] Shen X, Pei QQ, Liu XF. Survey of block chain. Chin J Netw Inf Secur 2016.

[7] Noneage, https://www.noneage.com/.

[8] Chidaolian, http://www.chidaolian.com/.

[9] Digital currency AML (Anti Money Laundering) research report, https://coinholmes.com/static/pdf/1.pdf.

[10] Ajello NJ. Fitting a square peg in a round hole: Bitcoin, money laundering, and the fifth amendment privilege against self-incrimination. Brooklyn Law Rev 2015;80.

[11] Kleiman JA. Beyond the silk road: Unregulated decentralized virtual currencies continue to endanger US national security and welfare. J Geol Soc 2013;140(5):769–79.

[12] Lin TCW. Compliance, technology, and modern finance. Brooklyn J Corp Financ Commer Law 2016;11:159–82.

[13] Kiviat TI. Beyond bitcoin: Issues in regulating blockchain transactions. Duke Law J 2015;65:569–608.

[14] Tsukerman M. The block is hot: A survey of the state of bitcoin regulation and suggestions for the future. Berkeley Technol Law J 2015;30:1127–69.

[15] Turpin JB. Bitcoin: The economic case for a global, virtual currency operating in an unexplored legal framework. Indiana J Glob Legal Stud 2014;21:335–68.

[16] Feld S, Schnfeld M, Werner M. Analyzing the deployment of Bitcoin's P2P network under an AS-level perspective. In: Proceedings of the 5th international conference on ambient systems, networks and technologies (ANT 2014), the 4th international conference on sustainable energy information technology, vol. 32; 2014. p. 1121–6.

[17] Park S, Im S, Seol Y, Paek J. Nodes in the bitcoin network: Comparative measurement study and survey. IEEE Access 2019;7:57009–22, https://doi.org/10.1109/ACCESS.2019.2914098.

[18] Cao T, Yu J, Decouchant J, Luo X, Verissimo P. Exploring the monero peer-to-peer network. IACR Cryptol ePrint Arch 2019;2019:411.

[19] Kim SK, Ma Z, Murali S, Mason J, Bailey M. Measuring ethereum network peers. In: The internet measurement conference 2018; 2018. p. 91–104.

[20] Koshy P, Koshy D, Mcdaniel P. An analysis of anonymity in bitcoin using P2P network traffic. In: International conference on financial cryptography and data security; 2014. p. 469–85.

[21] Biryukov A, Pustogarov I. Bitcoin over Tor isn't a good idea. In: IEEE symposium on security and privacy; 2015. p. 122–34.

[22] Biryukov A, Khovratovich D, Pustogarov I. Deanonymisation of clients in Bitcoin P2P network. In: ACM sigsac conference on computer & communications security; 2014. p. 15–29.

[23] Gao Y, Shi J, Wang X, Tan Q, Zhao C, Yin Z. Topology measurement and analysis on ethereum P2P network. In: International symposium on computers and communications; 2019. p. 1–7.

[24] Li Z, Hou J, Wang H, Wang C, Kang C, Fu P. Ethereum behavior analysis with netflow data. In: Asia Pacific network operations and management symposium; 2019. p. 1–6.

[25] Reid F, Harrigan M. An analysis of anonymity in the Bitcoin system. In: PASSAT/SocialCom 2011, privacy, security, risk and trust (PASSAT), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing; 2011. p. 1318–1326.

[26] Chen Z, Yong G. A graph-based investigation of bitcoin transactions. In: IFIP international conference on digital forensics; 2015. p. 79–95.

[27] Wang C, Chu X, Yang Q. Measurement and analysis of the bitcoin networks: A view from mining pools. Cryptogr Secur 2019. arXiv.

[28] Pham T, Lee S. Anomaly detection in the bitcoin system - a network perspective. Soc Inf Netw 2016. arXiv:.

[29] Chen T, Zhu Y, Li Z, Chen J, Li X, Luo X, et al. Understanding ethereum via graph analysis. In: International conference on computer communications; 2018. p. 1484–92.

[30] Sun H, Ruan N, Liu H. Ethereum analysis via node clustering. In: Network and system security - 13th international conference. Lecture notes in computer science, vol. 11928, Springer; 2019, p. 114–29, https://doi.org/10.1007/978-3-030-36938-5_7.

[31] Bai Q, Zhang C, Xu Y, Chen X, Wang X. Evolution of ethereum: A temporal graph perspective. In: 2020 IFIP networking conference, networking 2020. IEEE; 2020, p. 652–4, https://ieeexplore.ieee.org/document/9142770.

[32] Kumar A, Fischer C, Tople S, Saxena P. A traceability analysis of monero's blockchain. In: 22nd European symposium on research in computer security, 2017. Lecture notes in computer science, vol. 10493, Springer; 2017, p. 153–73, https://doi.org/10.1007/978-3-319-66399-9_9.

[33] Möser M, Soska K, Heilman E, Lee K, Heffan H, Srivastava S, et al. An empirical analysis of traceability in the monero blockchain. PoPETs 2018;2018(3):143–63, https://doi.org/10.1515/popets-2018-0025.

[34] Hinteregger A, Haslhofer B. An empirical analysis of monero cross-chain traceability. In: Financial cryptography and data security - 23rd international conference FC 2019. Lecture notes in computer science, vol. 11598, Springer; 2019, p. 150–7, https://doi.org/10.1007/978-3-030-32101-7_10.

[35] Wijaya DA, Liu JK, Steinfeld R, Liu D. Monero ring attack: Recreating zero mixin transaction effect. In: 17th IEEE international conference on trust, security and privacy in computing and communications / 12th IEEE international conference on big data science and engineering. IEEE; 2018, p. 1196–201, https://doi.org/10.1109/TrustCom/BigDataSE.2018.00165.

[36] Huang Y, Wang H, Wu L, Tyson G, Luo X, Zhang R, et al. Understanding (mis)behavior on the EOSIO blockchain. Proc ACM Meas Anal Comput Syst 2020;4(2):37:1–28, https://doi.org/10.1145/3392155.

[37] Zhao Y, Liu J, Han Q, Zheng W, Wu J. Exploring EOSIO via graph characterization. 2020, CoRR, abs/2004.10017, https://arxiv.org/abs/2004.10017.

[38] Quesnelle J. On the linkability of zcash transactions. 2017, CoRR, abs/1712.01210, http://arxiv.org/abs/1712.01210.

[39] Kappos G, Yousaf H, Maller M, Meiklejohn S. An empirical analysis of anonymity in zcash. In: 27th USENIX security symposium, USENIX security 2018. USENIX Association; 2018, p. 463–77, https://www.usenix.org/conference/usenixsecurity18/presentation/kappos.

[40] Zhang Z, Li W, Liu H, Liu J. A refined analysis of zcash anonymity. IEEE Access 2020;8:31845–53, https://doi.org/10.1109/ACCESS.2020.2973291.

[41] Biryukov A, Feher D, Vitto G. Privacy aspects and subliminal channels in zcash. In: Proceedings of the 2019 ACM SIGSAC conference on computer and communications security, CCS 2019. ACM; 2019, p. 1795–811, https://doi.org/10.1145/3319535.3345663.

[42] Biryukov A, Feher D. Privacy and linkability of mining in zcash. In: 7th IEEE conference on communications and network security, CNS 2019. IEEE; 2019, p. 118–23, https://doi.org/10.1109/CNS.2019.8802711.

[43] Liao K, Zhao Z, Doupé A, Ahn G. Behind closed doors: measurement and analysis of cryptolocker ransoms in bitcoin. In: 2016 APWG symposium on electronic crime research, ECrime 2016. IEEE; 2016, p. 1–13, https://doi.org/10.1109/ECRIME.2016.7487938.

[44] Spagnuolo M, Maggi F, Zanero S. Bitiodine: Extracting intelligence from the bitcoin network. In: Financial Cryptography. 2014, p. 457–68.

[45] Chen W, Wu J, Zheng Z, Chen C, Zhou Y. Market manipulation of bitcoin: Evidence from mining the Mt. Gox transaction network. In: 2019 IEEE conference on computer communications, INFOCOM 2019; 2019. p. 964–72.

[46] Chen W, Zheng Z, Cui J, Ngai E, Zhou Y. Detecting ponzi schemes on ethereum: Towards healthier blockchain technology. In: Proceedings of the 2018 world wide web conference on world wide web, WWW 2018, Lyon, France, April 23–27, 2018; 2018. p. 1409–18.

[47] Wu J, Yuan Q, Lin D, You W, Chen W, Chen C, et al. Who are the phishers? Phishing scam detection on ethereum via network embedding. Soc Inf Netw 2019. arXiv.

[48] Dan MG, Birch D, Akroyd D, Molina-Solana M, Guo Y, Knottenbelt WJ. Visualizing dynamic bitcoin transaction patterns. Big Data 2016;109–19.

[49] Kinkeldey C, Fekete J, Blascheck T, Isenberg P. Visualizing and analyzing entity activity on the bitcoin network. 2019, CoRR, abs/1912.08101, http://arxiv.org/abs/1912.08101.

[50] Bohr J, Bashir M. Who uses bitcoin? An exploration of the Bitcoin community. In: Twelfth conference on privacy, security and trust; 2014. p. 94–101.

[51] Androulaki E, Karame GO, Roeschlin M, Scherer T, Capkun S. Evaluating user privacy in bitcoin. In: Financial cryptography and data security - 17th international conference, FC 2013; 2013. p. 34–51.

[52] Yin HS, Vatrapu R. A first estimation of the proportion of cybercriminal entities in the bitcoin ecosystem using supervised machine learning. In: 2017 IEEE international conference on big data. IEEE Computer Society; 2017, p. 3690–9, https://doi.org/10.1109/BigData.2017.8258365.

[53] Paquet-Clouston M, Romiti M, Haslhofer B, Charvat T. Spams meet cryptocurrencies: Sextortion in the bitcoin ecosystem. 2019, CoRR, abs/1908.01051, http://arxiv.org/abs/1908.01051.

[54] Portnoff RS, Huang DY, Doerfler P, Afroz S, McCoy D. Backpage and bitcoin: Uncovering human traffickers. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, Halifax, NS, Canada, August 13 - 17, 2017. ACM; 2017, p. 1595–604, https://doi.org/10.1145/3097983.3098082.

[55] Toyoda K, Ohtsuki T, Mathiopoulos PT. Identification of high yielding investment programs in bitcoin via transactions pattern analysis. In: Globecom IEEE global communications conference; 2017.

[56] Toyoda K, Takis Mathiopoulos P, Ohtsuki T. A novel methodology for HYIP operators' bitcoin addresses identification. IEEE Access 2019;7:74835–48.

[57] Jawaheri HA, Sabah MA, Boshmaf Y, Erbad A. Deanonymizing tor hidden service users through bitcoin transactions analysis. Comput Secur 2020;89. https://doi.org/10.1016/j.cose.2019.101684.

[58] Eldefrawy K, Gehani A, Matton A. Longitudinal analysis of misuse of bitcoin. In: Applied cryptography and network security - 17th international conference, vol. 11464. 2019, p. 259–78, https://doi.org/10.1007/978-3-030-21568-2_13.

[59] Lee S, Yoon C, Kang H, Kim Y, Kim Y, Han D, et al. Cybercriminal minds: An investigative study of cryptocurrency abuses in the dark web. In: 26th annual network and distributed system security symposium. The Internet Society; 2019.

[60] Akcora CG, Li Y, Gel YR, Kantarcioglu M. Bitcoinheist: Topological data analysis for ransomware prediction on the bitcoin blockchain. In: Proceedings of the twenty-ninth international joint conference on artificial intelligence. ijcai.org; 2020, p. 4439–45, https://doi.org/10.24963/ijcai.2020/612.

[61] Ostapowicz M, Zbikowski K. Detecting fraudulent accounts on blockchain: A supervised approach. In: Web information systems engineering - WISE 2019 - 20th international conference. Lecture notes in computer science, vol. 11881, Springer; 2019, p. 18–31, https://doi.org/10.1007/978-3-030-34223-4_2.

[62] Harlev MA, Yin HS, Langenheldt KC, Mukkamala RR, Vatrapu R. Breaking bad: De-anonymising entity types on the bitcoin blockchain using supervised machine learning. In: 51st Hawaii international conference on system sciences; 2018. p. 1–10.

[63] Lin Y, Wu P, Hsu C, Tu I, Liao S. An evaluation of bitcoin address classification based on transaction history summarization. In: IEEE international conference on blockchain and cryptocurrency. IEEE; 2019, p. 302–10.

[64] Jourdan M, Blandin S, Wynter L, Deshpande P. Characterizing entities in the bitcoin blockchain. In: 2018 IEEE international conference on data mining workshops. 2018, p. 55–62, https://doi.org/10.1109/ICDMW.2018.00016.

[65] Zola F, Eguimendia M, Bruse JL, Urrutia RO. Cascading machine learning to attack bitcoin anonymity. In: IEEE international conference on blockchain. IEEE; 2019, p. 10–7, https://doi.org/10.1109/Blockchain.2019.00011.

[66] Tang H, Jiao Y, Huang B, Lin C, Wang B. Learning to classify blockchain peers according to their behavior sequences. IEEE Access 2018;6:71208–15.

[67] Huang B, Liu Z, Chen J, Liu A, Liu Q, He Q. Behavior pattern clustering in blockchain networks. Multimedia Tools Appl 2017;20099–110.

[68] Biryukov A, Tikhomirov S. Transaction clustering using network traffic analysis for bitcoin and derived blockchains. In: IEEE INFOCOM 2019 - IEEE conference on computer communications workshops, INFOCOM workshops 2019, Paris, France, April 29 - May 2, 2019. IEEE; 2019, p. 204–9, https://doi.org/10.1109/INFCOMW.2019.8845213.

[69] Shen M, Zhang J, Zhu L, Xu K, Du X, Liu Y. Encrypted traffic classification of decentralized applications on ethereum using feature fusion. In: Proceedings of the international symposium on quality of service; 2019. p. 18:1–18:10.

[70] Aiolli F, Conti M, Gangwal A, Polato M. Mind your wallet's privacy: identifying Bitcoin wallet apps and user's actions through network traffic analysis. In: Proceedings of the 34th ACM/SIGAPP Symposium on applied computing; 2019. p. 1484–91.

[71] Frowis M, Fuchs A, Bohme R. Detecting token systems on ethereum. In: Financial cryptography. 2019, p. 93–112.

[72] Chen T, Zhang Y, Li Z, Luo X, Wang T, Cao R, et al. Tokenscope: Automatically detecting inconsistent behaviors of cryptocurrency tokens in ethereum. In: Proceedings of the 2019 ACM SIGSAC conference on computer and communications security. ACM; 2019, p. 1503–20.
[73] Gao W, Hatcher WG, Yu W. A survey of blockchain: Techniques, applications, and challenges. In: 2018 27th international conference on computer communication and networks; 2018.
[74] Zheng Z, Xie S, Dai H, Chen X, Wang H. An overview of blockchain technology: Architecture, consensus, and future trends. In: 2017 IEEE international congress on big data, bigdata congress 2017. IEEE Computer Society; 2017, p. 557–64, https://doi.org/10.1109/BigDataCongress.2017.85.
[75] Ferdous MS, Chowdhury MJM, Hoque MA, Colman A. Blockchain consensuses algorithms: A survey. 2020, CoRR, https://arxiv.org/abs/2001.07091.
[76] Nguyen G, Kim K. A survey about consensus algorithms used in blockchain. JIPS 2018;14:101–28, https://doi.org/10.3745/JIPS.01.0024.
[77] Macrinici D, Cartofeanu C, Gao S. Smart contract applications within blockchain technology: A systematic mapping study. Telemat Inform 2018;35(8):2337–54, https://doi.org/10.1016/j.tele.2018.10.004.
[78] Cui Y, Pan B, Sun Y. A survey of privacy-preserving techniques for blockchain. In: Artificial intelligence and security - 5th international conference. Lecture notes in computer science, vol. 11635, Springer; 2019, p. 225–34, https://doi.org/10.1007/978-3-030-24268-8_21.
[79] Khalilov MCK, Levi A. A survey on anonymity and privacy in bitcoin-like digital cash systems. IEEE Commun Surv Tutorials 2018;2543–85. http://dx.doi.org/10.1109/COMST.2018.2818623.
[80] Feng Q, He D, Zeadally S, Khan MK, Kumar N. A survey on privacy protection in blockchain system. J Netw Comput Appl 2019;45–58. http://dx.doi.org/10.1016/j.jnca.2018.10.020.
[81] Monrat AA, Schelén O, Andersson K. A survey of blockchain from the perspectives of applications, challenges, and opportunities. IEEE Access 2019;7:117134–51, https://doi.org/10.1109/ACCESS.2019.2936094.
[82] Berdik D, Otoum S, Schmidt N, Porter D, Jararweh Y. A survey on blockchain for information systems management and security. Inf Process Manag 2021.
[83] Li X, Jiang P, Chen T, Luo X, Wen Q. A survey on the security of blockchain systems. Future Gener Comput Syst 2020;841–53. http://dx.doi.org/10.1016/j.future.2017.08.020.
[84] Saad M, Spaulding J, Njilla L, Kamhoua CA, Shetty S, Nyang D, et al. Exploring the attack surface of blockchain: A systematic overview. 2019, CoRR, abs/1904.03487, http://arxiv.org/abs/1904.03487.
[85] Wang H, Wang Y, Cao Z, Li Z, Xiong G. An overview of blockchain security analysis. In: Cyber security. 2019.
[86] Hasanova H, Baek U, Shin M, Cho K, Kim M. A survey on blockchain cybersecurity vulnerabilities and possible countermeasures. Int J Netw Manag 2019;29(2). http://dx.doi.org/10.1002/nem.2060.
[87] Atzei N, Bartoletti M, Cimoli T. A survey of attacks on ethereum smart contracts (SoK), In: International conference on principles of security and trust; 2017.
[88] Zhu L, Zheng B, Shen M, Yu S, Gao F, Li H, et al. Research on the security of blockchain data: A survey. 2018, CoRR, http://arxiv.org/abs/1812.02009.
[89] Zaghloul E, Li T, Mutka M, Ren J. Bitcoin and blockchain: Security and privacy. IEEE Internet Things J 2020;7(10):10288–313. http://dx.doi.org/10.1109/JIOT.2020.3004273.
[90] Anita N, Vijayalakshmi M. Blockchain security attack: A brief survey. In: 10th international conference on computing, communication and networking technologies. 2019, p. 1–6. http://dx.doi.org/10.1109/ICCCNT45670.2019.8944615.
[91] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. 2008.
[92] Bitcoin project, https://github.com/bitcoin/bitcoin.
[93] Gavin W. Ethereum: a secure decentralised generalised transaction ledger. In: Ethereum project yellow paper. 2014, p. 1–32. http://dx.doi.org/10.1017/CBO9781107415324.004.
[94] Decentralized applications, https://github.com/ethereum/wiki/wiki/Decentralized-apps-(dapps).
[95] DApps 2020, https://www.stateofthedapps.com/.
[96] Buterin V, et al. A next-generation smart contract and decentralized application platform. 2014, https://ethereum.org/whitepaper/.
[97] EOS.IO technical. 2018, https://github.com/EOSIO/Documentation/blob/master/TechnicalWhitePaper.md.
[98] An incentivized, blockchain-based, public content platform. 2017, https://steem.com/SteemWhitePaper.pdf.
[99] Yong Y, Wang F-Y. Blockchain: the state of the art and future trends. Acta Automat Sinica 2016;42(4):481–94.
[100] BCSEC, https://bcsec.org.
[101] Litke A, Anagnostopoulos D, Varvarigou T. Blockchains for supply chain management: Architectural elements and challenges towards a global scale deployment. Logistics 2019;3:5. http://dx.doi.org/10.3390/logistics3010005.
[102] Thampi SM, Sekaran KC. Survey of search and replication schemes in unstructured P2P networks. 2010, CoRR, http://arxiv.org/abs/1008.1629.
[103] Ratnasamy S, Francis P, Handley M, Karp RM, Shenker S. A scalable content-addressable network. In: Proceedings of the ACM SIGCOMM conference on applications, technologies, architectures, and protocols for computer communication. 2001, p. 161–72, https://doi.org/10.1145/383059.383072.
[104] Stoica I, Morris RT, Karger DR, Kaashoek MF, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for internet applications. In: Proceedings of the ACM SIGCOMM 2001 conference on applications, technologies, architectures, and protocols for computer communication. ACM; 2001, p. 149–60, https://doi.org/10.1145/383059.383071.
[105] Anderson L, Holz R, Ponomarev A, Rimba P, Weber I. New kids on the block: an analysis of modern blockchains. 2016, CoRR, abs/1606.06530, http://arxiv.org/abs/1606.06530.
[106] Bitcoin Developer. Bitcoin p2p networking, https://developer.bitcoin.org/reference/p2p_networking.html.
[107] EthFans Node Project, https://github.com/EthFans/wiki/wiki.
[108] Bitnodes 2020, https://bitnodes.io/.
[109] Ethernodes 2020, https://ethernodes.org/.
[110] Bitcoin Community. Change wiki, https://en.bitcoin.it/wiki/Change.
[111] Apostolaki M, Zohar A, Vanbever L. Hijacking bitcoin: Routing attacks on cryptocurrencies. In: 2017 IEEE symposium on security and privacy, SP 2017. IEEE Computer Society; 2017, p. 375–92, https://doi.org/10.1109/SP.2017.29.
[112] Apostolaki M, Marti G, Müller J, Vanbever L. SABRE: protecting bitcoin against routing attacks. In: 26th annual network and distributed system security symposium, NDSS 2019. 2019, https://www.ndss-symposium.org/ndss-paper/sabre-protecting-bitcoin-against-routing-attacks/.
[113] Litke P, Stewart J. BGP hijacking for cryptocurrency profit. 2014, https://www.secureworks.com/research/bgp-hijacking-for-cryptocurrency-profit.
[114] Yan H, Oliveira R, Burnett K, Matthews D, Zhang L, Massey D. Bgpmon: A real-time, scalable, extensible monitoring system. In: Cybersecurity applications technology conference for homeland security. 2009, http://dx.doi.org/10.1109/CATCH.2009.28.
[115] Heilman E, Kendler A, Zohar A, Goldberg S. Eclipse attacks on bitcoin's peer-to-peer network. In: 24th USENIX security symposium, USENIX security 2015. USENIX Association; 2015, p. 129–44, https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/heilman.
[116] Singh A, Ngan T, Druschel P, Wallach DS. Eclipse attacks on overlay networks: Threats and defenses. In: INFOCOM 2006. 25th IEEE international conference on computer communications, joint conference of the IEEE computer and communications societies. 2006, https://doi.org/10.1109/INFOCOM.2006.231.
[117] 74% of bitcoin-related sites suffered a DDoS attack, https://www.bleepingcomputer.com/news/security/74-percent-of-all-bitcoin-related-sites-suffered-a-ddos-attack/.
[118] Bitcoin Community. Coinbase wiki, https://en.bitcoin.it/wiki/Coinbase.
[119] Schrijvers O, Bonneau J, Boneh D, Roughgarden T. Incentive compatibility of bitcoin mining pool reward functions. In: Financial cryptography and data security - 20th international conference. Lecture notes in computer science, vol. 9603, Springer; 2016, p. 477–98, https://doi.org/10.1007/978-3-662-54970-4_28.
[120] Delgado-Segura S, Pérez-Solà C, Navarro-Arribas G, Herrera-Joancomartí J. Analysis of the bitcoin UTXO set. IACR Cryptol ePrint Arch 2017;1095, http://eprint.iacr.org/2017/1095.
[121] Vujicic D, Jagodic D, Randic S. Blockchain technology, bitcoin, and Ethereum: A brief overview. In: International symposium infoteh-jahorina; 2018. p. 1–6.
[122] Karame G, Androulaki E, Capkun S. Two bitcoins at the price of one? Double-spending attacks on fast payments in bitcoin. IACR Cryptol ePrint Arch 2012;2012:248, http://eprint.iacr.org/2012/248.
[123] Karame G, Androulaki E. Double-spending fast payments in bitcoin. In: The ACM conference on computer and communications security. 2012, p. 906–17, https://doi.org/10.1145/2382196.2382292.
[124] Proof of stake versus proof of work. 2015, https://bitfury.com/content/downloads/pos-vs-pow-1.0.2.pdf.
[125] Vector76 attack, https://en.bitcoin.it/wiki/Double-spending#Vector76_attack.
[126] Hajdarbegovic N. Bitcoin miners ditch ghash.io pool over fears of 51% attack. 2014, https://www.coindesk.com/bitcoin-miners-ditch-ghash-io-pool-51-attack.
[127] Yang X, Chen Y, Chen X. Effective scheme against 51% attack on proof-of-work blockchain with history weighted information. In: IEEE international conference on blockchain. 2019, p. 261–5. http://dx.doi.org/10.1109/Blockchain.2019.00041.
[128] Sun H, Ruan N, Su C. How to model the bribery attack: A practical quantification method in blockchain. In: Computer security - ESORICS 2020 - 25th European symposium on research in computer security. 2020, p. 569–89. http://dx.doi.org/10.1007/978-3-030-59013-0_28.
[129] Roll Back Attack about blacklist in EOS, https://medium.com/@slowmist/roll-back-attack-about-blacklist-in-eos-adf53edd8d69.
[130] "Transaction congestion attack": Attackers could paralyze EOS network with minimal cost. 2019, https://blog.peckshield.com/2019/01/15/eos_CVE-2019-6199/.
[131] Luu L, Teutsch J, Kulkarni R, Saxena P. Demystifying incentives in the consensus computer. In: Proceedings of the 22nd ACM SIGSAC conference on computer and communications security. ACM; 2015, p. 706–19, https://doi.org/10.1145/2810103.2813659.
[132] He N, Zhang R, Wu L, Wang H, Luo X, Guo Y, et al.
[133] Defeating EOS gambling games: The techniques behind random number loophole. 2018, https://blog.peckshield.com/2018/11/22/eos/.

[134] An incorrect check of EOS transaction status will cause “false-top-up” valunerability, https://medium.com/@slowmist/hard-fail-status-attack-for-eos-7cfa73ae7d4b.

[135] Protecting from replay attacks after recent BCH hard fork. 2018, https://blog.peckshield.com/2018/11/19/bch/.

[136] Maesa DDF, Marino A, Ricci L. An analysis of the Bitcoin users graph: inferring unusual behaviours. In: Complex networks & their applications V - proceedings of the 5th international workshop on complex networks and their applications; 2016. p. 749–60.

[137] Hyperledger project. 2015, https://www.hyperledger.org/.

[138] “Fake transfer notice” loophole details explained, 140k EOS tokens lost by eosbet. 2018, https://blog.peckshield.com/2018/10/26/eos/.

[139] “Fake EOS attack” upgraded, 60k EOS tokens lost by eoscast. 2018, https://blog.peckshield.com/2018/11/02/eos/.

[140] King S, Nadal S. PPCoin: Peer-to-peer crypto-currency with proof-of-stake. In: Self-published paper. 2012.

[141] Nextcoin, https://github.com/nxcoin/nxcoin-project.

[142] Advanced decentralized blockchain platform. 2018, https://tron.network/static/doc/white_pape_v_2_0.pdf.

[143] Gervais A, Karame GO, Wüst K, Glykantzis V, Ritzdorf H, Capkun S. On the security and performance of proof of work blockchains. IACR Cryptol ePrint Arch 2016;2016. http://eprint.iacr.org/2016/555.

[144] Saleh F. Blockchain without waste: Proof-of-stake. Soc Sci Electron Publishing 2018.

[145] Kiayias A, Konstantinou I, Russell A, David B, Oliynykov R. A provably secure proof-of-stake blockchain protocol. IACR Cryptol ePrint Arch 2016;889, http://eprint.iacr.org/2016/889.

[146] Larimer D. Delegated proof of stake (DPOS). In: Bitshare, White paper. 2014.

[147] Castro M, Liskov B. Practical byzantine fault tolerance. In: Proceedings of the third USENIX symposium on operating systems design and implementation (OSDI) 1999. USENIX Association; 1999, p. 173–86, https://dl.acm.org/citation.cfm?id=296824.

[148] Zheng K, Liu Y, Dai C, Duan Y, Huang X. Model checking PBFT consensus mechanism in healthcare blockchain network. In: 2018 9th international conference on information technology in medicine and education; 2018.

[149] Kwon J. Tendermint: Consensus without mining. 2014.

[150] Douceur JR. The sybil attack. In: Peer-to-peer systems, first international workshop, IPTPS 2002, Cambridge, MA, USA, March 7–8, 2002, revised papers. Lecture notes in computer science, vol. 2429, Springer; 2002, p. 251–60, https://doi.org/10.1007/3-540-45748-8_24.

[151] Natoli C, Gramoli V. The balance attack against proof-of-work blockchains: The R3 testbed as an example. 2016, CoRR, arXiv:abs/1612.09426.

[152] Nomura research institute: Survey on blockchain technologies and related services, https://www.meti.go.jp/english/press/2016/pdf/0531_01f.pdf.

[153] Sharma PK, Chen M, Park JH. A software defined fog node based distributed blockchain cloud architecture for IoT. IEEE Access 2018;6:115–24, https://doi.org/10.1109/ACCESS.2017.2757955.

[154] Yasin A, Liu L. An online identity and smart contract management system. In: 40th IEEE annual computer software and applications conference, compsac workshops 2016. 2016, p. 192–8, https://doi.org/10.1109/COMPSAC.2016.2.

[155] Shae Z, Tsai JJP. On the design of a blockchain platform for clinical trial and precision medicine. In: 37th IEEE international conference on distributed computing systems. 2017, p. 1972–80, https://doi.org/10.1109/ICDCS.2017.61.

[156] Ethereum RPC, https://github.com/ethereum/go-ethereum/tree/be9172a7ac5cf8a6919a36213531c51ecb5cc6ef/rpc.

[157] Carlin D, OrKane P, Sezer S, Burgess J. Detecting cryptomining using dynamic analysis. In: 2018 16th annual conference on privacy, security and trust; 2018.

[158] Bahack L. Theoretical bitcoin attacks with less than half of the computational power (draft). IACR Cryptol ePrint Arch 2013;2013:868, http://eprint.iacr.org/2013/868.

[159] Solat S, Potop-Butucaru M. Zeroblock: Preventing selfish mining in bitcoin. 2016, CoRR, abs/1605.02435, http://arxiv.org/abs/1605.02435.

[160] Luu L, Saha R, Parameshwaran I, Saxena P, Hobor A. On power splitting games in distributed computation: The case of bitcoin pooled mining. In: IEEE 28th computer security foundations symposium. 2015, p. 397–411, https://doi.org/10.1109/CSF.2015.34.

[161] Eyal I, Sirer EG. Majority is not enough: Bitcoin mining is vulnerable. In: Financial cryptography and data security - 18th international conference. Lecture notes in computer science, vol. 8437, 2014, p. 436–54, https://doi.org/10.1007/978-3-662-45472-5_28.

[162] Myetherwallet domain-hijacking. 2018, https://blog.peckshield.com/2018/04/26/mew-dns-hijacking/.

[163] Know your ransomware: CTB-locker. 2017, https://www.secalliance.com/blog/ransomware-ctb-locker/.

[164] Fernández-Caramés TM, Fraga-Lamas P. Towards post-quantum blockchain: A review on blockchain cryptography resistant to quantum computing attacks. IEEE Access 2020;21091–116. http://dx.doi.org/10.1109/ACCESS.2020.2968985.

[165] Maunder M. Cryptomining supply chain attack hits government websites. 2018, https://www.wordfence.com/blog/2018/02/cryptomining-javascript-supply-chain-attack/?utm_source=list&utm_medium=email&utm_campaign=021118.

[166] Moen D. Wordpress supply chain attacks: An emerging threat. 2018, https://www.wordfence.com/blog/2018/01/wordpress-supply-chain-attacks/.

[167] Pranesh SA, Kannan VV, Viswanathan N, Vijayalakshmi M. Design and analysis of incentive mechanism for ethereum-based supply chain management systems. In: 11th international conference on computing, communication and networking technologies. 2020, http://dx.doi.org/10.1109/ICCCNT49239.2020.9225602.

[168] Lai E, Luo W. Static analysis of integer overflow of smart contracts in ethereum. In: Proceedings of the 2020 4th international conference on cryptography, security and privacy. 2020, p. 110–5. http://dx.doi.org/10.1145/3377644.3377650, https://doi.org/10.1145/3377644.3377650.

[169] ALERT: New batchoverflow bug in multiple ERC20 smart contracts (CVE-2018-10299). 2018, https://blog.peckshield.com/2018/04/22/batchOverflow/.

[170] New proxyoverflow bug in multiple ERC20 smart contracts (CVE-2018-10376). 2018, https://blog.peckshield.com/2018/04/25/proxyOverflow/.

[171] New multioverflow bug identified in multiple ERC20 smart contracts (CVE-2018-10706). 2018, https://blog.peckshield.com/2018/05/10/multiOverflow/.

[172] Inlinereflex attacks. 2018, https://blog.peckshield.com/2018/12/18/inlineReflex/.

[173] Hu Y, Seneviratne S, Thilakarathna K, Fukuda K, Seneviratne A. Characterizing and detecting money laundering activities on the bitcoin network. 2019, CoRR, abs/1912.12060, http://arxiv.org/abs/1912.12060.

[174] Wannacry ransomware attack. 2017, https://en.wikipedia.org/wiki/WannaCry_ransomware_attack.

[175] Treasury H. UK National risk assessment of money laundering and terrorist financing. 2017, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/468210/UK_NRA_October_2015_final_web.pdf.

[176] Jake F. Wannacry ransomware attack. 2020, https://www.investopedia.com/terms/d/dark-wallet.asp.

[177] Matzutt R, Hiller J, Henze M, Ziegeldorf JH, Wehrle K. A quantitative analysis of the impact of arbitrary blockchain content on bitcoin. In: Financial cryptography and data security. 2018.

[178] Conti M, Mancini LV, Spolaor R, Verde NV. Analyzing android encrypted network traffic to identify user actions. IEEE Trans Inf Forensics Secur 2016;11(1):114–25, https://doi.org/10.1109/TIFS.2015.2478741.

[179] Sharma A, Bhatia A. Bitcoin's blockchain data analytics: A graph theoretic perspective. 2020, CoRR, abs/2002.06403, https://arxiv.org/abs/2002.06403.

[180] Béres F, Seres IA, Benczúr AA, Quintyne-Collins M. Blockchain is watching you: Profiling and deanonymizing ethereum users. 2020, CoRR, abs/2005.14051, https://arxiv.org/abs/2005.14051.

[181] Ron D, Shamir A. Quantitative analysis of the full bitcoin transaction graph. In: Financial cryptography. 2013, p. 6–24.

[182] Zheng P, Zheng Z, Dai H-n. Xblock-ETH: Extracting and exploring blockchain data from ethereum. Cryptogr Secur 2019. arXiv.

[183] Akcora CG, Gel YR, Kantarcioglu M. Blockchain: A graph primer. 2017, CoRR, abs/1708.08749, http://arxiv.org/abs/1708.08749.

[184] Maesa DDF, Marino A, Ricci L. Uncovering the bitcoin blockchain: An analysis of the full users graph. In: 2016 IEEE international conference on data science and advanced analytics; 2016. p. 537–46.

[185] Bitcoin explorer, https://bsv.btc.com/.

[186] Ethereum explorer, https://eth.btc.com/.

[187] Monero explorer, https://xmr.tokenview.com/.

[188] Kiffer L, Levin D, Mislove A. Analyzing Ethereum's contract topology. In: Internet measurement conference; 2018, p. 494–9.

[189] Fleder M, Kester MS, Pillai S. Bitcoin transaction graph analysis. 2015, CoRR.

[190] Meiklejohn S, Pomarole M, Jordan G, Levchenko K, McCoy D, Voelker GM, et al. A fistful of bitcoins: characterizing payments among men with no names. In: Proceedings of the 2013 internet measurement conference. ACM; 2013, p. 127–40, https://doi.org/10.1145/2504730.2504747.

[191] S. Nair B, Kumar R. Anonymity analysis of bitcoin transactions using unsupervised machine learning. Int J Res Sci Innov 2018.

[192] Goldsmith D, Grauer K, Shmalo Y. Analyzing hack subnetworks in the bitcoin transaction graph. 2019, CoRR, abs/1910.13415, http://arxiv.org/abs/1910.13415.

[193] Oggier FE, Phetsouvanh S, Datta A. Biva: Bitcoin network visualization & analysis. In: 2018 IEEE international conference on data mining workshops, ICDM workshops, Singapore, Singapore, November 17–20, 2018. IEEE; 2018, p. 1469–74, https://doi.org/10.1109/ICDMW.2018.00210.