Anomaly Detection in Blockchain Networks: A Comprehensive Survey

Muneeb Ul Hassan, Mubashir Husain Rehmani, and Jinjun Chen

***Abstract*—Over the past decade, blockchain technology has attracted a huge attention from both industry and academia because it can be integrated with a large number of everyday applications working over features of modern information and communication technologies (ICT). Peer-to-peer (P2) architecture of blockchain enhances these applications by providing strong security and trust-oriented guarantees, such as immutability, verifiability, and decentralization. Despite of these incredible features that blockchain technology brings to these ICT applications, modern research have indicated that these strong guarantees are not sufficient enough and blockchain networks may still prone to various security, privacy, and reliability related issues. In order to overcome these issues, it is important to identify the anomalous behaviour within time. Therefore, nowadays anomaly detection models are playing an important role in protection of modern blockchain networks. These anomaly detection models autonomously detect and predict anomaly in the network in order to protect network from unexpected attacks. In this article, we provide an in-depth survey regarding integration of anomaly detection models in blockchain technology. For this, we first discuss that how anomaly detection can aid in ensuring security of blockchain based applications. Then, we demonstrate certain fundamental evaluation matrices and key requirements that can play a critical role while developing anomaly detection models for blockchain. Afterwards, we present a thorough survey of various anomaly detection models from perspective of each layer of blockchain to provide readers an in-depth overview of integration that has been carried out till date. Finally, we conclude the article by highlighting certain important challenges alongside discussing that how they can serve as a future research directions for new researchers in the field.**

***Index Terms*—Blockchain, Anomaly Detection, Fraud Detection**

I. Introduction

Blockchain technology was first introduced by Satoshi Nakamoto as a distributed decentralized storage ledger for Bitcoin to overcome the problem of double spending [1]. Certain key functionalities of blockchain technology such as immutability, security, decentralization, tamper-resistance, and distributed consensus aided to establish trust among transacting parties of Bitcoin. This tremendous success of Bitcoin as a cryptocurrency grabbed attention of researchers, who then started to explore the underlying technology behind this cryptocurrency named as blockchain. Research works investigated that although blockchain integrated with information and communication technologies (ICT) have a vast number of use cases in daily life applications, such as supply chain, finances, IoT operations, cloud services, etc. Since then, applications of blockchain technology are being explored by researchers and this number is continuously increasing [2].

From a technological point of view, blockchain is an immutable append-only ledger which works over the phenomenon of decentralized peer-to-peer (P2P) networking [3]. Blockchain technology can also be considered as a repository of blocks which are connected together to form a chain like structure. Every block in the blockchain ledger consists of multiple transactions, these transactions could be of any type ranging from financial transactions to transportation data. Apart from transactions, every block also has the time stamp and the value of hash of past block, which are them combined with transactional values to compute a combined hash value, which later serves as a header of that specific block [4]. This header hash is then again linked with the next block and a chain like structure is formed which cannot be tempered due to this strong guarantee. In a public blockchain network, all transactions are transparent and are publicly available, thus anyone in the network can view these transactions and can cross-verify any fraudulent activity. This property of blockchain networks make them more secure and trustworthy for a large number of use cases where trust is required [5].

To understand blockchain a bit further, blockchain can further be classified into multiple layers named as data layer, network layer, incentive layer, and smart contract layer. Each layer in the blockchain taxonomy has its own functionalities and responsibilities, e.g., data layer is responsible to organize all data blocks in a chain-like structure, contract layer is responsible for successful deployment of smart contracts, etc. Similarly, all other layers have their own functionalities according to the requirements. A detailed discussion about layer-oriented taxonomy of blockchain has been provided in Section. II-C.

Despite of these benefits, blockchain technology is not 100% secure and it is still prone to certain attacks and issues [6], [7]. For example, a large number of Ponzi schemes have been developed in order to steal money from legitimate users, a large number of malicious accounts are being created regularly to carry out money laundering. Similarly, in certain cases some malicious forks are being created and deployed to overcome computational power and to carry out double spending in the network [8]. Therefore, for successful functioning of a blockchain network, it is important to detect the occurrence of these vulnerabilities in the most precise and timely manner. To provide the successful detection and prediction of such attacks over blockchain, the field of anomaly detection for blockchain comes in action. The major functionality is to detect or

M. Ul Hassan and J. Chen are with the Swinburne University of Technology, Hawthorn VIC 3122, Australia (e-mail: muneebmh1@gmail.com; jinjun.chen@gmail.com).

M.H. Rehmani is with the Department of Computer Science, Munster Technological University, Ireland (e-mail: mshrehmani@gmail.com).

even predict the future occurrence of any vulnerability in the network in order to take a timely action against them. A large number of anomaly detection models are being created and deployed by researchers for various blockchain. From a generic point of view, these models can be categorized on the basis of different layers of blockchain. E.g., certain models work over prediction of anomalous commands in the smart contracts, so these models comes under anomaly detection for smart contract. Similarly, some models work over detection of malicious block deployment, so these models can be categorized under the umbrella of anomaly detection in data layer. Overall, it is important to mention anomaly detection is one of the important field in order to secure future blockchain network, and a vast amount of work is being carried out in this field from various perspective which we will discuss in this survey.

A. Contributions of This Article

Certain surveys have been published in the field blockchain, however, to the best of our knowledge, none of them provides an in-depth overview of anomaly detection in blockchain technology. To summarize, the major contributions of our work are as follows:

- We focus over providing generalist audience a brief overview regarding the field of anomaly detection in blockchain by discussing all fundamentals and preliminaries involved in this direction. E.g., providing basic discussion about blockchain technology, anomaly detection, and the need of anomaly detection in blockchain.
- We provide detailed discussion about classification of anomalous attacks, their detection models, and existing works in blockchain technology alongside providing some fundamental matrices and key requirements for their robust and timely identification.
- We highlight critical challenges in blockchain based anomaly detection that needs to be solved alongside providing a brief overview of future directions associated with these challenges.

B. Comparison with Related Surveys

Our survey on anomaly detection in blockchain network is distinctive from all past surveys because we cover the aspect of anomaly detection in detail from basics to integration perspective at different layers of blockchain technology. Discussing about previous works, a vast number of surveys have been published which revolve around blockchain technology form various perspective attacks such as privacy, security, etc. In this section, we compile a list of these surveys and alongside presenting a thorough comparison that how our work is novel and distinct from all previous works. A brief table presenting major contribution and coverage of the survey from perspective of anomaly detection scopes is given in Table. I.

In literature, a detailed survey discussing various security threats and their machine-learning based countermeasures have been presented by Mohamed *et al.* in [9]. Authors first presented a brief overview regarding Bitcoin and its underlying technologies alongside discussing various security threats, and afterwards, authors provide a detailed discussion about existing solutions proposed by researchers till now. Similarly, another similar survey discussing the privacy and anonymity of Bitcoin and similar cryptocurrencies have been presented by authors in [10]. In the article authors investigated and explored various privacy and anonymity solutions alongside highlighting their effectiveness in cryptocurrencies. A short review discussing intrusion detection systems (IDS) in blockchain technology have been presented by Meng *et al.* in [11]. Authors first discussed basics about integration of IDS in blockchain, and then discuss various solutions and scopes of this direction to give readers an overall point of view. Similar to this, a work discussing security services via blockchain have been presented by Salman *et al.* The article first emphasized that blockchain can be used as a critical tool for security services and then the authors presented a through literature review of services which are being enhanced via blockchain technology.

A work over discussing privacy preservation strategies in blockchain based Internet of Things (IoT) systems have been presented by authors in [13]. Authors presented a detailed taxonomy of all possible solutions alongside discussing technical works which have implemented these solutions. The next work [14] is a master's thesis rather than a survey work, which discussed the implementation of seven anomaly detection models in blockchain scenario. The presented work only covers the technical implementation rather than a detailed taxonomy and layer oriented survey. Moving towards generic surveys of blockchain, a very comprehensive survey discussing blockchain from theory to application perspective of IoT have been presented by Wu *et al.* in [15]. The work classified all practical implementation and their extensions in detail in order to give readers an in-depth point of view. Another similar survey providing a thorough discussion about blockchain, and its useful scenarios have been presented by authors in [16]. In this article, authors started discussion from basics of blockchain and Bitcoin and then moved to advanced technologies.

From security perspective, very detailed works have been presented and published by researchers. E.g., a short article discussing security verification have been presented by Liu *et al.* in [17]. Authors picked and discussed 53 articles in which this aspect of security verification via smart contract of blockchain was discussed. Another article on very similar topic of security and attacks of blockchain has also been presented by authors in [18]. Similar to this, a very detailed survey covering the direction of attack survey of blockchain has been presented by Saad *et al.* in [19]. Comparably, a comprehensive work providing details about threats, vulnerabilities, and resilience models from security perspective of blockchain have also been presented by authors in [20]. The work presents a detailed taxonomy of all possible security vulnerabilities alongside discussing their defences.

Alongside this, a very short work discussing anomaly detection in blockchain via data mining methods have been presented by authors in [21]. The work only provide a surface level discussion and did not went in-depth of anomaly

TABLE I
COMPARISON OF EXISTING SURVEY WORKS IN THE SIMILAR FIELD OF BLOCKCHAIN AND ANOMALY DETECTION WITH THEIR MAJOR CONTRIBUTION AND CERTAIN SCOPES.

Acronyms: Classification of Blockchain Anomalies (CoBA), Anomaly in Contract Layer (AiCL), Anomaly in Data Layer (AiDL), Anomaly in Network Layer (AiNL), Anomaly in Incentive Layer (AiIL), Existing Works in Blockchain Anomaly Detection (EWiBAD). Tick(✔) Shows that the mentioned topic is covered, Cross(✗) shows that the provided domain is not covered, and Asterisk(✻) shows that the particular topic is partially covered.

Ref No.	**Year**	**Major Contribution of Surveys**	**Scope**					
			CoBA	**AICL**	**AIDL**	**AiNL**	**AiIL**	**EWiBAD**
[9]	2018	A comprehensive survey about security concerns and their countermeasures in Bitcoin.	✗	✗	✗	✻	✻	✻
[10]	2018	A thorough literature about privacy and anonymity of Bitcoin-like systems.	✗	✗	✗	✗	✻	✻
[11]	2018	A detailed study about integration of intrusion detection systems with blockchain technology.	✗	✗	✗	✻	✗	✗
[12]	2019	A detailed survey of blockchain-based works in various security services.	✗	✗	✗	✗	✗	✗
[13]	2019	A comprehensive discussion about privacy issued in IoT scenarios operating on blockchain.	✗	✗	✗	✗	✗	✗
[14]	2019	A thesis works evaluating various anomaly detection mechanisms in blockchain.	✻	✗	✻	✗	✻	✻
[15]	2019	A survey on integration strategies of blockchain technology with IoT and beyond from application perspective.	✗	✗	✗	✗	✗	✗
[16]	2019	A comprehensive survey about blockchain, its functioning, and applicability in various scenarios.	✗	✗	✗	✗	✗	✗
[17]	2019	A survey over security assurance and correction verifications of smart contracts deployed on blockchain.	✗	✗	✗	✗	✗	✗
[18]	2020	A thorough survey on security attacks and their countermeasures for blockchain based IoT and IIoT systems.	✗	✗	✗	✗	✗	✗
[19]	2020	Worked over exploration of attack surfaces and attacks vectors in blockchain environment.	✗	✗	✗	✗	✗	✗
[20]	2020	A detailed investigation of blockchain from perspective of security reference architecture for blockchain.	✗	✗	✗	✗	✗	✗
[21]	2020	A book chapter over anomaly detection approaches in blockchain technology.	✔	✻	✻	✗	✻	✻
[22]	2020	A detailed investigation of integration of privacy preservation via differential privacy strategy in blockchain technology.	✗	✗	✗	✗	✗	✗
[23]	2020	A survey over necessities of privacy services, security issues, and applications of blockchain technology.	✗	✗	✗	✗	✗	✗
[24]	2020	A comprehensive survey of consensus algorithms in blockchain based systems.	✗	✗	✗	✗	✗	✗
[25]	2020	A thorough technical overview of blockchain smart contracts.	✗	✗	✗	✗	✗	✗
This Work	2021	A comprehensive survey on anomaly detection in blockchain technology from perspective of identification, integration, requirement, and methodologies for anomaly detection in blockchain.	✔	✔	✔	✔	✔	✔

detection models from layers and integration perspective. Apart from anomaly, a work providing an in-depth evaluation and discussion about integration of differential privacy mechanism in layers of blockchain technology have been presented by authors in [22]. Another work covering a brief overview of security and privacy threats alongside future applications have been published by authors in [23]. Apart from these threats, a detailed survey providing an overview of consensus models in blockchain networks have been given in [24]. Continuing this trend, a detailed survey covering advances in the field of blockchain based smart contracts have been published by Kemmoe *et al.* in [25].

C. *Organization of Article*

The remainder of article is organized as follows: Section 2 provides a generalist introduction of preliminaries of anomaly detection in blockchain, while section 3 provides fundamental guidelines to identify anomalous behaviour in blockchain from classification, matrices, and detection requirement perspective. Discussion from Section 4 to Section 7 provides a classification oriented in-depth technical overview of the works carried out in blockchain anomaly detection from perspective of different blockchain layer. Afterwards, Section 8 provides discussion about challenges and prospective future directions in the field of blockchain anomaly detection. Finally, section 9 concludes the article.

II. Preliminaries of Anomaly Detection in Blockchain

In this section, we discuss the basic concepts of anomaly detection in blockchain which will be helpful for readers to understand the technical aspects mentioned in the later sections.

A. *What is Anomaly Detection?*

Anomaly detection has its roots with various fields related to data such as data mining, data analytics, etc. Formally, it can be defined as a mechanism which is used to identify atypical patterns in data which are different from the normal behaviour of whole dataset [26]. These atypical patterns are usually known as outliers, which afterwards help trace the root cause of anomalous behaviour. From the perspective of ICTs, anomaly detection is usually viewed from two aspects, one aspect is that anomaly could be due any fault in the network, which is causing anomalous generation of data, while the other aspect says that anomaly could be due to some novel instance in the data [27]. This novel instance or fault could be harmful or harmless to the whole network depending upon its nature and reason for generation. Therefore, it is important to identify both of type of anomalous behaviours in a timely manner for efficient functioning of the network. By keeping the above definition in mind, in this article, we focus over identifying all prospective anomalies in blockchain network that can be generated due to some fault or due to some novel instance, whether malicious or not.

B. *What is Blockchain?*

Blockchain came into discussion after advent of decentralized cryptocurrency named as Blockchain by Satoshi Nakamoto [1]. The reason Bitcoin gained hype was because its backend technology was completely different from the alternative online currencies of that time. E.g., Bitcoin focused over enhancing both transparency and trust in the network without the involvement of centralized authority. Certain other decentralized alternatives were also being proposed at that time, however, Bitcoin took the lead by developing a consensus mechanism via which all nodes can agree of a unified digital ledger without any conflict to avoid any possible double-spending [10]. This decentralized consensus of Bitcoin was the mortal blow which attracted the attention of researchers and researchers started exploring the technology behind Bitcoin named as blockchain.

As the name suggested, blockchain is a chain of blocks, which are connected together by strong cryptographic guarantees. Each block has a header, which is the hash of its contents, and each block has the hash of its previous block in the body alongside other contents. Thus, tampering with anything in the block body will change the hash of header, which eventually will not be linked with the next block due to being tampered. In this way, blockchain technology ensures that the ledger remains tamper-proof. Alongside this, each of the blockchain node has a copy of this digital ledger, thus, a single node cannot play with the complete chain as well, which ensures that the ledger is immutable. All the nodes of blockchain operate in decentralized manner and have strong cryptographically secured communication between them which obeys the rules of P2P networking. All these functionalities combine to form a modern day technology, which is named as blockchain. To write it formally, blockchain can be termed as a traceable, time-stamped, append-only, tamper-proof, immutable digital ledger which is capable of storing data in a decentralized P2P manner [22]. Readers interested to explore more about basics of blockchain technology can study a very good resource by Belotti *et al.* in [16].

C. *Layered Architecture of Blockchain*

Since blockchain is a fully functional P2P network having the features of decentralized communication, incentives, and consensus, thus, to understand the functionality of this a bit further, researchers divided it into multiple layers. Various works have been carried out to identify and discuss different layers, for instance, Homoliak *et al.* in [20] and Wu *et al.* in [15] classified blockchain into four layers named as network layer, consensus layer, data/state layer, and application layer. Similarly, Belotti *et al.* [16] divided architecture of blockchain into five layers and added another layer named as 'Execution layer'. Similar to this, Xie *et al.* in [28] added another layer and proposed a 06 layered architecture of blockchain comprising of data layer, network layer, consensus layer, incentive layer, contract layer, and application. Apart from these pioneering works, some other works also highlighted and named other layers as well on the basis of functionality for a specific use case. Since, the focus of our work to is to detect anomaly in blockchain network, thus, after going through the available literature, it can be concluded that four layers of blockchain are more prone to anomaly attacks. Therefore, in this article, we discuss the detection of anomalies in blockchain from perspective of four layers named as data layer, network layer, incentive layer, and contract layer.

Each of the layer of blockchain have their own functionalities and are responsible for the allocated tasks. In this section, we briefly highlight the functionalities of four prominent layers which we will be discussing later in this article from anomaly detection perspective.

1) Data Layer: The first layer in the blockchain architecture is data layer which comprise of data blocks. The blocks in the data layer are time-stamped and are linked with one another via hashes to form a chain-like structure. A typical block in a blockchain network comprises of two parts named as header and body of block [28]. Block header mainly comprises of important metadata parameters, such as block hash, previous block hash, time-stamp, nonce value, Merkle root hash, etc. The contents in the block header can also differ according to the need of application. The second part of the block is block body which mainly comprises of the transactions which are picked from mining pool for the purpose of storing on blockchain. The hashes of these transactions are computed and are further combined to form a single Merkle root hash, which is also the part of block header.

2) Network Layer: Network layer in blockchain is responsible to carry out distributed communication and networking for blockchain peers. Since, blockchain is a P2P network in which all peers have same rights, thus, the functionality of this layer is to run such networking and communication models which ensure the timely distribution, forwarding, and verification of blocks in the network [20]. For instance, if a transaction is generated in the blockchain network, then the network layer is responsible to broadcast this transaction to all neighbouring peers. Similarly, verification acknowledgement of this transaction will also be returned via using functionalities of network layer. Afterwards, if the transaction turns out to be a valid transaction, then it will again be sent to broadcast to other peer nodes. Contrarily, an invalid transaction is denied and is not sent for further broadcast in the network. Within this transaction verification step, it is also important to highlight that usually the functionalities of cryptographic mechanisms are generally used to ensure transaction validity, which comprise of a signing via public-private key pair. A detailed discussion about transaction signing and verification is out of scope of this article, interested readers are suggested to study the discussion given in [16].

3) Incentive Layer: Incentive layer in blockchain revolves around financial incentives in the network, which serves as a major factor of motivation for participants of the network [16]. In a decentralized network with no centralized authority, maintaining motivation of participants is a major challenge, and this challenge in blockchain is solved by developing incentive models in incentive layer. For instance, in Bitcoin network, a specific number of Bitcoins are given as a reward upon for completion of a round of mining. This reward mechanism motivates Bitcoin users to actively participate in the mining process. Similarly, in other blockchain networks, similar rewards is issues upon completion of specific tasks, which serves as a driving force for the network. Apart from incentives, certain penalties and deposits do also comes under the scope of this layer. E.g., if some user behaves in a malicious manner, then he/she should have to pay the penalty amount for that malicious behaviour. The models and mechanisms corresponding to these penalty amounts are also linked with this incentive layer, because it directly deals with financial things.

4) Contract Layer: Contract layer, also known as smart contract layer, is responsible to bring programmable functionalities in the blockchain network. Traditional blockchain networks such Bitcoin only provides its users with few basic functionalities, such transactions, incentives, etc. However, modern blockchain networks, such as Hyperledger Fabric, Ethereum 2.0, etc., provides the functionalities of dynamic programming in which the users can write a logical program to execute it on the network. The program is written in the form of a contract, which is known as smart contract. This smart contract is a piece of executable code, which runs over the blockchain network, and performs the tasks assigned to it. There could be multiple types of smart contracts depending upon the nature of execution, application, and requirements. However, a detailed discussion about different types of smart contracts is out of scope of this article, interested readers are suggested to study an interesting article by [29]. To summarize, the layer dealing with all these functionalities is known as contract layer and it is playing a very critical role in development of modern day blockchain networks.

D. Anomaly Problem in Blockchain Layers

Since the advent of blockchain by Satoshi Nakamoto in 2008, adversaries are continuously trying to carry out malicious activities in the network ranging from cryptocurrency frauds to identity thefts for blockchain wallets. To efficiently run operations of blockchain network, it is important to identify and take action against these adversarial behaviours within time. In order to do so, anomaly detection models came into discussion, which are responsible for effective detection of anomalous behaviour of a specific node. Generically, we divide the anomaly detection models into six sub-categories on the basis of their functioning, which are named as generative architectures, classification based models, clustering based models, nearest neighbour models, statistical & analytical models, and reinforcement learning based models (see Fig. 1).

From the perspective of blockchain layers, it is important to mention that the anomalies in blockchain technology are not pretty generic because almost all layers of blockchain have their specific anomalies, thus, their detection mechanism do also vary. For example, an anomaly related to Ponzi scheme falls under category of incentive layers, while spreading of anomalous messages in the network falls under the category of network layer. Therefore, it is important to classify these anomalies according to the layer they fall under in order to carry out their efficient detection.

E. Technical Challenges while Integration of Anomaly Detection in Blockchain

Nevertheless, development of basic anomaly detection models is not much complicated if one have basic knowledge of machine learning. However, when it comes to anomaly detection in blockchain networks, then certain challenges arise due to the nature of blockchain. In this section, we discuss certain prospective challenges, that one can face while development of anomaly detection models for blockchain based scenarios.

1) Network-wide Consensus on Anomaly: The first challenge that one have to overcome while developing anomaly detection models for blockchain network is to carry out a network-wide consensus on anomaly. Since, the blockchain network has no centralized entity to determine rules, thus, it becomes hard to categorize a specific event to be an anomaly or not. Therefore, alongside designing a mechanism to detect outliers, one also have to make sure that this outlier is considered as an anomaly throughout the network. In short, complete network have to reach on a consensus that a particular event is an anomaly and appropriate actions should be taken against it. This becomes even more challenging, when some nodes in the network start behaving in a malicious manner. Therefore, considering the aspect of network-wide consensus alongside designing anomaly detection models is important in blockchain because of decentralization.

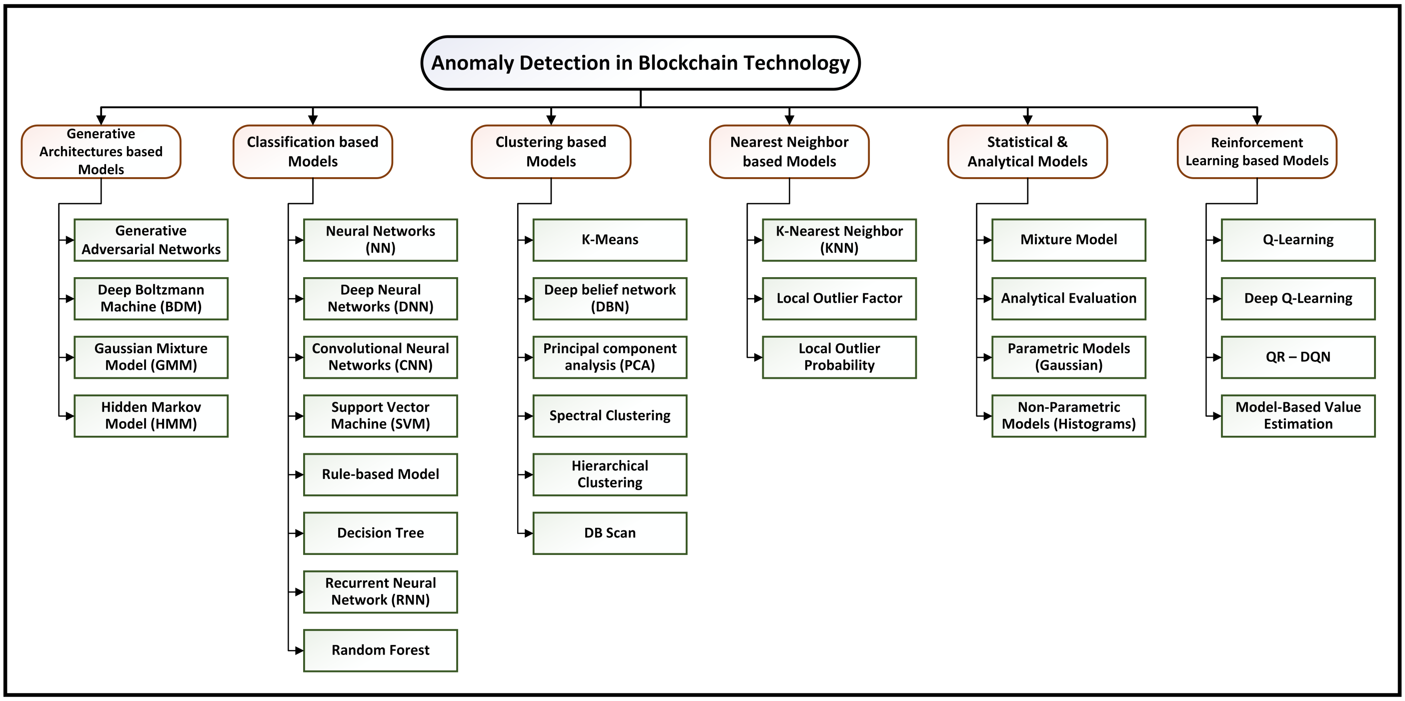
Anomaly Detection in Blockchain Technology
Generative Architectures based Models
Classification based Models
Clustering based Models
Nearest Neighbor based Models
Statistical & Analytical Models
Reinforcement Learning based Models
Generative Adversarial Networks
Deep Boltzmann Machine (BDM)
Gaussian Mixture Model (GMM)
Hidden Markov Model (HMM)
Neural Networks (NN)
Deep Neural Networks (DNN)
Convolutional Neural Networks (CNN)
Support Vector Machine (SVM)
Rule-based Model
Decision Tree
Recurrent Neural Network (RNN)
Random Forest
K-Means
Deep belief network (DBN)
Principal component analysis (PCA)
Spectral Clustering
Hierarchical Clustering
DB Scan
K-Nearest Neighbor (KNN)
Local Outlier Factor
Local Outlier Probability
Mixture Model
Analytical Evaluation
Parametric Models (Gaussian)
Non-Parametric Models (Histograms)
Q-Learning
Deep Q-Learning
QR – DQN
Model-Based Value Estimation

Fig. 1: Classification of Anomaly Detection Models in Blockchain Technology

2) Careful Selection of Outlier Features: Since blockchain is a novel paradigm and plenty of attacks are pretty new even for researchers, therefore, selecting the best features for outliers is one of the major challenge. For instance, if one wants to label the purpose of a smart contract in a blockchain network, then it becomes hard due to very less available references. In order to overcome this, researchers worked over developing automated methods for labelling new and unknown smart contracts for detection of possible anomalies from these labels [30]. However, majority of smart contracts are pretty identical and do not show significant difference, therefore, such methods are not well-established till now. Thus, it is always challenging to identify these features and one have to be extra careful in order to ensure that they do not categorize a legitimate user or transaction as anomaly.

3) Smart Contract Programmability & Execution due to Environmental Constraints: It is important to mention that majority of smart contracts of blockchain are being developed in bytecode rather than binary code, which makes it difficult to run traditional anomaly detection models on blockchain network in real-time environment [31]. Similarly, detecting anomalies at bytecode level becomes even more challenging because not all information is available at the level of bytecode, and certain information gets lost during the compilation process [32]. Therefore, designing such models, which efficiently detect anomalies from smart contract bytecode is a big challenge that every researchers working in blockchain anomaly detection faces regularly.

4) Lack of Defined Rules: With the advent of every new application of blockchain, the rules changes accordingly. E.g., a blockchain based smart grid will have different set of rules for anomaly as compared to a network of blockchain based electric vehicles. Therefore, the rules defined for anomaly detection in decentralized smart grid cannot be applied to other blockchain networks. Similarly, the generic rules defined for generalized blockchain networks cannot be applied to specific domain-oriented networks. Certain researchers worked over carrying out manual inspection of models and truncations in order to gather detailed information, however, it is a tiresome and inefficient process [31]. Therefore, it can be said that designing a set of rules is one of the major challenge for researchers working in the domain of anomaly detection in blockchain networks.

III. Fundamentals Guidelines to Identify Anomalous Behaviour in Blockchain

In this section, we discuss various fundamental guidelines that will pave the path towards understanding and development of anomaly detection models in blockchain.

A. Types of Anomalous Attacks on Blockchain

Broadly, the blockchain anomalies can be categorised to five subtypes on the basis of their orientation, such as account based, smart contract based, consensus based, transactions based, and system based. A detailed taxonomy of these anomalies can be visualized in Fig. 2. However, in order to provide our readers an in-depth functioning of some of the critical anomalies and attacks, we picked some of the most prominent and severe ones and discuss them in this section.

1) Malicious Transaction Pattern Detection: One of the most common anomaly in blockchain networks is uneven transactions. Due to pseudonymity property of blockchain, nodes usually feel safe to carry out large transactions, however, among these transactions, some transactions are also uneven, which are mostly carried out for some malicious purposed.

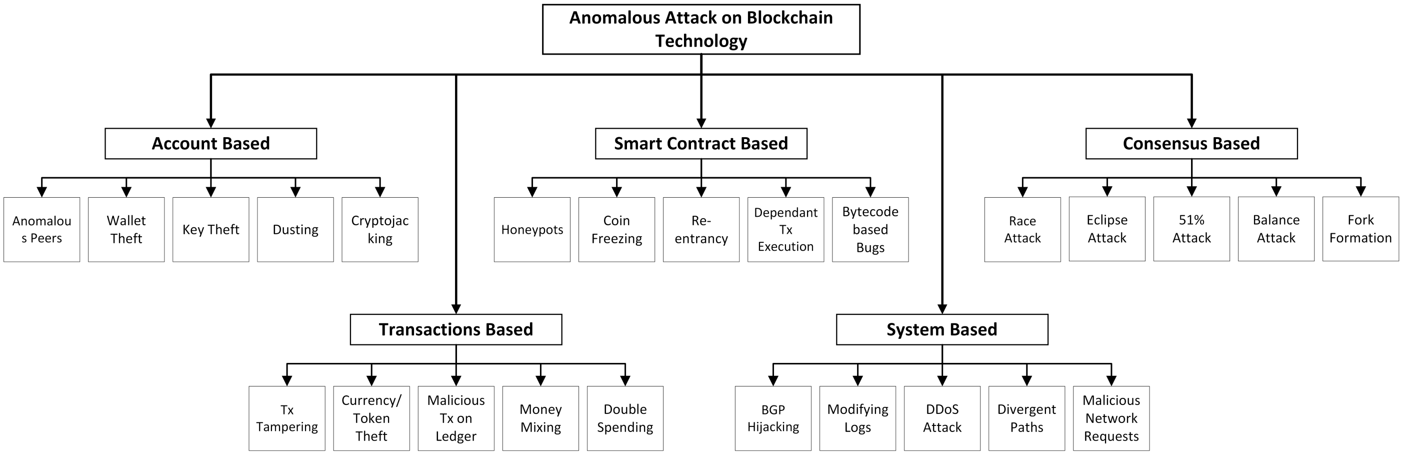

Fig. 2: Classification of Anomalous Attacks on Blockchain Technology

For example, some users try to do carry out money laundering while being anonymous in the blockchain network [33]. Therefore, it is important to identify such transactions within time in order to take appropriate action against them. Fortunately, due to decentralized nature of blockchain, these transactions can be identified by observing various transaction patterns.

2) Double Spending Detection: As the name suggests, double spending is related to spending or utilization of an asset more than once in a decentralized network. Since blockchain is decentralized and there is no central authority to verify each transaction, therefore, malicious nodes usually try to use this nature of blockchain to carry out double-spending [34]. A transaction on blockchain is finalized once it gets validated by the peer nodes, however, during the process of validation, some malicious nodes try to use same amount of funds to carry out multiple transactions. Due to strong consensus guarantees of blockchain network, it is not that easy to carry out double spending, however, sometimes hackers succeed in fooling the network. Therefore, for successful functioning of blockchain network, it is important to timely detect and even predict the occurrence of double spending in the network.

3) Money Mixing Detection: Generally, money mixing is a legitimate process in blockchain networks, which revolves around mixing different assets or tokens to overcome identifiability trail [35]. The mixing in blockchain is usually used to enhance transactions anonymity from malicious attackers. However, some maleficent nodes try to take unlawful advantages out of it and try to hide their transaction patterns to carry out immoral activities, such as money laundering, etc. Therefore, it is important to identify malicious money mixing in order to protect blockchain network from unlawful activities.

4) Currency/Token Theft Detection From Network: Apart from basic anomalies, things get intense when malicious users try to steal tokens directly from user by hacking or similar other ways. In a centralized payment model, such as banks, they continuously monitor large transactions and do not approve any huge transaction unless the owner approves or provide some sort of verification. However, in a decentralized blockchain network, there is no centralized mediatory to regulate these transactions. Similarly, currency theft in blockchain can also be linked with asset theft, such as transfer of copyrights, etc., in which a person transfer rights or ownership of its assets to another user. Since, blockchain is tamper-proof, so reversing this transfer is not easy at all if it gets validated. Therefore, the importance of theft detection in decentralized blockchain scenarios increase exponentially to prevent any large mishap.

5) Smart Contract Anomalies: Smart contracts play a critical role in functioning and development of modern day blockchain because they add the feature of programmability in blockchain networks. Through this programmability feature, one can use blockchain for numerous advantages ranging from tracking decentralized ownership of assets to verification of education degrees. However, the base of these smart contracts is programming, and programming is not guaranteed to be 100% perfect all the time. There is always a possibility of mistakes in the smart contract, and since smart contracts are irreversible, thus, these mistakes can cause big damage. Similarly, apart from unintentional mistakes, some adversaries try to set up honeypots in smart contracts, the sole purpose of whom is to perform fraudulent activities such as coin theft, etc [32]. Therefore, in both of the cases, it is important to identify any anomaly in blockchain smart contract before its execution.

6) Wallet Theft Identification: The term wallet or digital wallet in blockchain is used to nominate a functionality, which is responsible for storage of assets or cryptocurrencies. A digital wallet allows blockchain users to efficiently manage, store, and trade their available assets via decentralized blockchain network. Traditional blockchain model uses digital signature on basis of elliptic curve cryptography to ensure security of wallets, however, instances have shown that wallets have been compromised in the past due to key thefts, etc [36]. One of the most common mode of carrying out wallet theft is via phishing, in which hackers try to carry out a phishing attack on targeted node in order to steal credentials. Another way that attackers use to compromise blockchain wallets is excessive generation of wallets keys in order to compromise their security. Therefore, considering the above discussion, it can be said that it is critical to identify such theft as soon as possible in order to take appropriate action.

7) Malicious Network Requests: From an outside perspective, blockchain is a secure network, however, from an insider point of view, blockchain network is still prone to certain attacks, and malicious network requests are one of them. In such requests, adversary nodes try to tamper with the transactional values before pushing these values to peer nodes. In this way, hackers try to divide the network into multiple parts so that they would not be able to communicate with each other. These type of malicious requests are also known as routing attacks on the network, and they can further be divided into partition and delay based attacks on the basis of their nature. These malicious requests can cause a big harm in the network; thus, their timely identification and eradication is compulsory.

8) Divergent Path & Forks: Blockchains as a ledger are immutable according to their nature, which implies that the information on the ledger cannot be changed once it gets recorded. However, if an adversary try to play maliciously while trading or while developing a smart contract, then they can exploit this feature of immutability for their unlawful benefits and can initiate formation of new forks in the network. Attackers usually try to create divergent paths in order to take control over 51% of network, which basically leads to disastrous consequences. Therefore, for successful functioning of a blockchain network, the timely detection and prevention of forks is very important.

9) Race Attack: Race attack in the blockchain network is affiliated with transaction fraud, in which an attacker tries to create two transactions which are conflicting with each other. For instance, attacker initiates a transaction with a victim (for instance to purchase some asset), who accepts the transactions and sends the required asset or product. But before the transaction confirmation, the attacker at the same time, send back the same amount of cryptocurrency to the attackers other account and broadcast this transaction to the network for validation. This act of attacker makes the first transaction to the victim as invalid, and victim ends up losing the prospective money. This race attack can be used for large malicious purposed, therefore, opportune detection and prevention of such attacks are important in a blockchain network.

10) Tampering Blockchain Logs: In a certain blockchain based applications, logs play an important role in determining operational activity. For instance, in a manufacturing industry, whenever a new step is performed, it is recorded into blockchain ledger by log system, which is then used to ensure the quality of product. However, if some nodes try to act as an adversary, then they can try to tamper with the logs in order to misguide the scrutinizing body, which makes audit difficult or sometimes impossible. Therefore, in order to ensure successful functioning of such blockchain applications, a regular and through analysis of logs is required to overcome any mishap.

B. Evaluation Matrices being used to Identify Anomalies

In an anomaly detection system, one needs to be very precise about every factor they consider, e.g., one cannot overlook certain anomalies as it can lead to serious mishap. Similarly, on the other hand one cannot even classify a normal behaviour as an anomaly, which can also lead to strenuous trouble certain times. Therefore, while developing such models, researchers are required to carefully consider certain factors, such as accuracy, precision, etc. In this section, we provide a brief overview of such factors alongside their importance in anomaly detection.

1) Sensitivity & Specificity of Outcome: In anomaly detection, the factors related to sensitivity and specificity play a critical role in determining effectiveness of any model. Nevertheless, in an anomaly detection model, decision could be right or wrong which means true or false but quantifying the outcome to make the model more effective is the key. In order to understand the sensitivity and specificity of an anomaly detection model a bit further, researchers have devised certain terms, which are discussed as follows:

a) True Positive (TP): The outcome TP means that the anomaly model gave a positive outcome and identified a specific behaviour as an anomaly, and in reality the result is true and that behaviour was actually anomalous.

b) False Positive (FP): This term means that the anomaly model gave a positive outcome and identified the node/behaviour as an anomaly, but in reality it was not anomaly.

c) True Negative (TN): The outcome TN identifies that the anomaly detection model gave a negative result for detection of anomalous behaviour and in reality its true and the behaviour was not anomalous.

d) False Negative (FN): The outcome FN means that the anomaly model gave a negative detection of anomaly but in reality the behaviour was anomalous.

In an ideal anomaly detection model, the rate of TP and TN should be high, and the rate of FP and FN should be low.

2) Confusion Matrix & Accuracy: Confusion matrix (also known as error matrix) is a visual table, which is used by researchers to analyse the efficiency and accuracy of the any model. In anomaly detection mechanisms, it is used to carry out comparison between the predicted and actual class labels. E.g., in a 2-by-2 confusion matrix, one side would represent the actual/true values, and the other side will represent predicted values. The matrix is then filled according to the outcomes of model in comparison with the actual values. E.g., if the predicted value is Yes, and the actual outcome is also Yes, then the value of TP in the table is incremented. Similarly, if the value of prediction is No, and the actual value is also No, then the value of TN is increments. In this way, all values of TP, TN, FP, and FN are filled, and are then used to compare accuracy of model via following equation [37]:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

3) Recall, Precision, and F-Score: Recall, precision, and F-score are the factors which are being used by researchers to analyse outcome of an anomaly detection model. The first terminology among them is recall, which can be defined as number of correctly identified instances by a model. To be more detailed, it is the number of true positives divided by

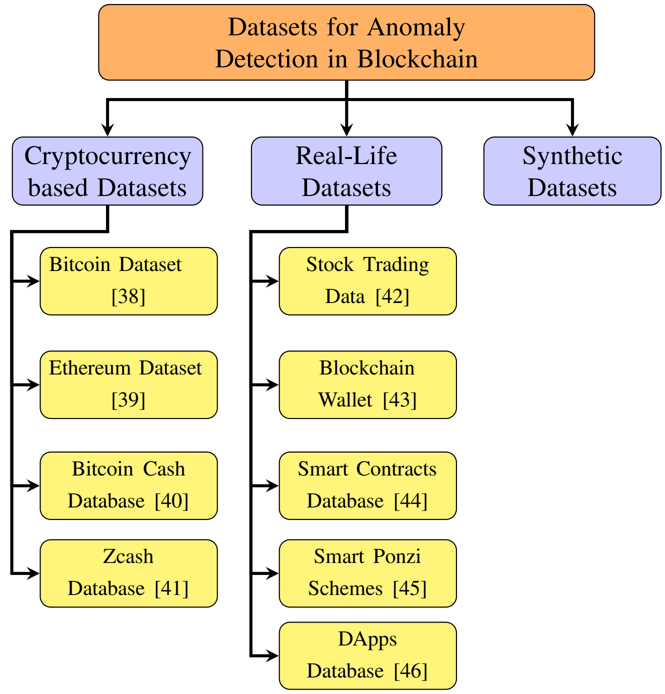

Fig. 3: Datasets for Anomaly Detection in Blockchain

the total number of actual positive instances. Precision can be termed as a counterpart of recall, as the precision is the ratio between total number of correct returned results and accumulative sum of positively identified results including false positives. F-Score, also known as F-1 Score mixes the property of precision and recall in a harmonic manner, so that the model can be evaluated in the best manner. The higher F-Score means the credibility of model to give good results is high. A detailed discussion about these parameters is out of scope of this article, readers interested in studying these parameters can study the interesting article by Bhuyan *et al.* [37].

C. Key Requirements for Anomaly Detection Mechanisms Design

Usually, behaviour of an anomalous participant varies noticeably as compared to other legitimate participants, therefore, they gets identified by anomaly detection models. However, in order to get efficient results, certain key requirements needs to be considered while designing anomaly detection models for blockchain networks. In this section, we discuss these requirements to give readers an overall viewpoint.

1) Data Collection Requirement: One of the major requirement for efficient functioning of any anomaly detection model is to have an adequate amount of data for analysis. Even this step of data collection is challenging in traditional anomaly detection environment, and in blockchain based anomaly detection scenarios, it becomes even more challenging due to various protecting mechanism in the way of blockchain data collection. For each type of blockchain, the data collection methods do vary, e.g., in case of public blockchain, the data is available to all participating nodes and one can carry out anomaly detection easily. However, in case of private or consortium setting, the data is not publicly available, and certain approvals are required before carrying out any processing over data. Furthermore, in all these types, the participating nodes are usually identified by pseudonyms, therefore, tracking the exact individual even after classifying it as an anomaly is sometimes very complicated due to lack of data about that individual. Some of the key datasets, which can be used to train models for future prediction of anomalies of blockchain network have been highlighted in Fig. 3.

2) Data Preprocessing Requirements: Raw data usually contains a lot of noise, therefore, preprocessing is a step in anomaly detection, in which collected data is modified and manipulated in order to reduce any noise and vulnerabilities from the data which can cause hurdle in detection of anomalies [47]. Some of the key steps involved in data preprocessing includes cleaning of data, transformation of data, selecting required features from data, reduction of data, and discretization of data. These steps are there to ensure that only the fine-grained data is fed to anomaly detection models, so that anomaly is detected as quick as possible. In majority of anomaly detection models, data preprocessing is considered as an essential step before feeding any data to anomaly detection models in order to enhance the detection accuracy and efficiency.

Similar pathway is also adopted in blockchain based anomaly detection models, in which a data collected via blockchain is nodes is preprocessed via pre-developed mechanisms, usually via preprocessing smart contracts [48], [49]. Similarly, in certain bytecode based anomaly detection models of blockchain, data is denoised using autoencoders [31]. Comparably in detection of anomalies in HYIP on blockchain, the preprocessing phase usually involve removal of transaction change part, alongside calculation and identification of patterns between transactions [50]. From the perspective of malicious account detection on Ethereum, the data is usually preprocessed in two steps, first via string comparison to identify duplicate addressed and then via filtering of EOA addressed and contract addresses [51]. Apart from the traditional preprocessing, certain real-time big data preprocessing tools and methods have also been developed till now for various applications, which modifies the steps involved in preprocessing in a way that it enhances the overall time and efficiency of preprocessing [52]. However, such works have not yet been carried out in the field of blockchain technology and there is a need for such integrations in future.

3) Choosing Appropriate Model for Anomaly Detection: Multiple methods to detect anomalies in networks have been developed by researchers, e.g., distance based models, classification based models, etc. Choosing Which models to choose for a specific type of anomaly is still a question that is being explored by the scientific community. Some people argue that distance and similarity based models, such as KNN, etc. are one of the viable ones, because they provide strong theoretical guarantees [47]. However, other works argue that traditional time-series based analysis can also be fruitful if used in an appropriate manner [53]. Nevertheless, this topic is debatable and according to us, a universal model cannot be used for all sort of anomalies and choosing of model purely depends upon the type of anomaly being targeted.

In order to provide our readers a detailed overview of anomaly detection techniques from blockchain point of view, we classify them into six sub-types on the basis of their working phenomenon. The detailed classification figure has been presented in Fig. 1. The first type in the classification are generative architectures, in which generative adversarial networks (GANs) are the most prominent ones. The use of generative architecture such as GANs for anomaly detection is recently being studied in different domains [54]. However, from the perspective of anomaly detection in blockchain, this field still require exploration. The next type of models are classification based models, which we believe are one of the most explored ones in blockchain scenarios, because of its diverse range of algorithms such as CNN, DNN, SVM, etc. The third and fourth type are clustering based and nearest neighbour based models, respectively, which are relatively simple but useful models that have been employed by researchers to study blockchain anomalies. The fifth one comprises of statistical and analytical modelling, where anomalies are studied on the basis of manual models developed according to the need and understanding of anomalies for a specific application. The last one are reinforcement learning based models, similar to first one, this field is also not much explored, but it has a lot of potential from perspective of anomaly detection in blockchain.

4) Computational Requirements: In order to detect anomalous behaviour in an efficient manner, the model needs to predict the anomalies within a specific time-frame, and that can only happen if the computational requirement to run an anomaly model matches with the required task. Certain models require high computational complexity, while on the other hand some models only need minimal computational efficiency. In blockchain anomaly detection, a large computational power is usually consumed during mining and consensus process. Therefore, in order to avoid overloading of machines, the on-chain anomaly detection models needs to require computational power alongside high accuracy for smooth functioning of the network. Various generic models have been designed to reduce computational cost of anomaly detection, however, in blockchain, this field is still progressing, and it has a lot of room in it for research.

5) Algorithm Design Requirement: Algorithmic complexity plays a key role while designing efficient algorithms for blockchain anomaly detection models. Anomaly algorithm design is also related with computational cost, e.g., if algorithm of an anomaly detection model is efficiently designed, then it will require less compactional cost to identify anomalies. The higher the algorithmic complexity, the more time and computational resources model will take to give desirable results. Plenty of ways are being developed by researchers to efficiently reduce the algorithmic complexity in order to achieve effective anomaly detection in blockchain environment.

6) Accuracy Requirements: Anomaly detection accuracy is another key element that cannot be ignored while developing anomaly detection models. This element even strengthens in case of blockchain anomaly because in blockchain decisions are irreversible and will always be there on the ledger. E.g., if anomaly detection predicts some specific node as an anomaly, and authorities take some action just on basis of that anomaly detection outcome, then this action will remain on blockchain forever. This is fine in case of an anomalous node, but in case of a false positive result, it will be a big challenge for authorities as they have identified a rational user as an anomaly, which can lead to disastrous outcomes. Therefore, it is important for a blockchain anomaly detection model to have high accuracy during prediction.

7) Privacy Requirements: Privacy aspect can never be underestimated while dealing with users' data for anomaly detection. Similar is the case with blockchain anomaly detection that if one wants to ensure users participating in the program, then privacy is one of the major concern for them. For example, no one want to share the information that they have transferred a specific amount of tokens to a specific person, however, when an anomaly detection model collects data, there are vey high chances that this information gets collected, which can cause a huge privacy issue for the individual in future. Therefore, while designing anomaly detection models, researchers try to ensure users that their privacy is intact, and it is not at the verge of leakage.

IV. Anomaly Detection in Data Layer of Blockchain

In this section, we classify the works that have highlighted these anomalies and worked over their efficient prediction from perspective of data layer (cf. Section II-C for details).

A. Bitcoin Fraud Detection

Discussing about data, one cannot undermine the data available on Bitcoin platform, which is the highest ranked cryptocurrency till now according to its market value. The price of Bitcoin has increased dramatically in the past few years, according to a report by Investopedia, the net worth of circulating Bitcoin is more than $600 billion in total as of May 2021 [67]. On one hand, Bitcoin provides a trusted and secure medium of financial transactions, but on the other hand due to is pseudonymity it also provides a safe passage for exchange or purchase of illegal assets, services, and good. Therefore, it is important to highlight and take appropriate action against such activities in a timely manner. Anomalies over Bitcoin network can be traced and tracked back at different blockchain layers, and data layer has a significant importance among them, because it allows participants to view the complete blockchain ledger, which makes it easy to identify particular anomalies of the network.

One such work has been carried out by Battista *et al.* in [55], where the authors proposed a system named as 'BitConeView' for analysis of Bitcoin transaction in a visual manner. The aim of the work is to carry out deanonymization of transaction flow by developing a visually analysable system. *BitCoveView* allows analysing users to track the sources, flow, and patterns of Bitcoin transactions in a detailed manner. One of the critical use case of these type of systems is to detect fraudulent and money laundering transactions, which authors investigated

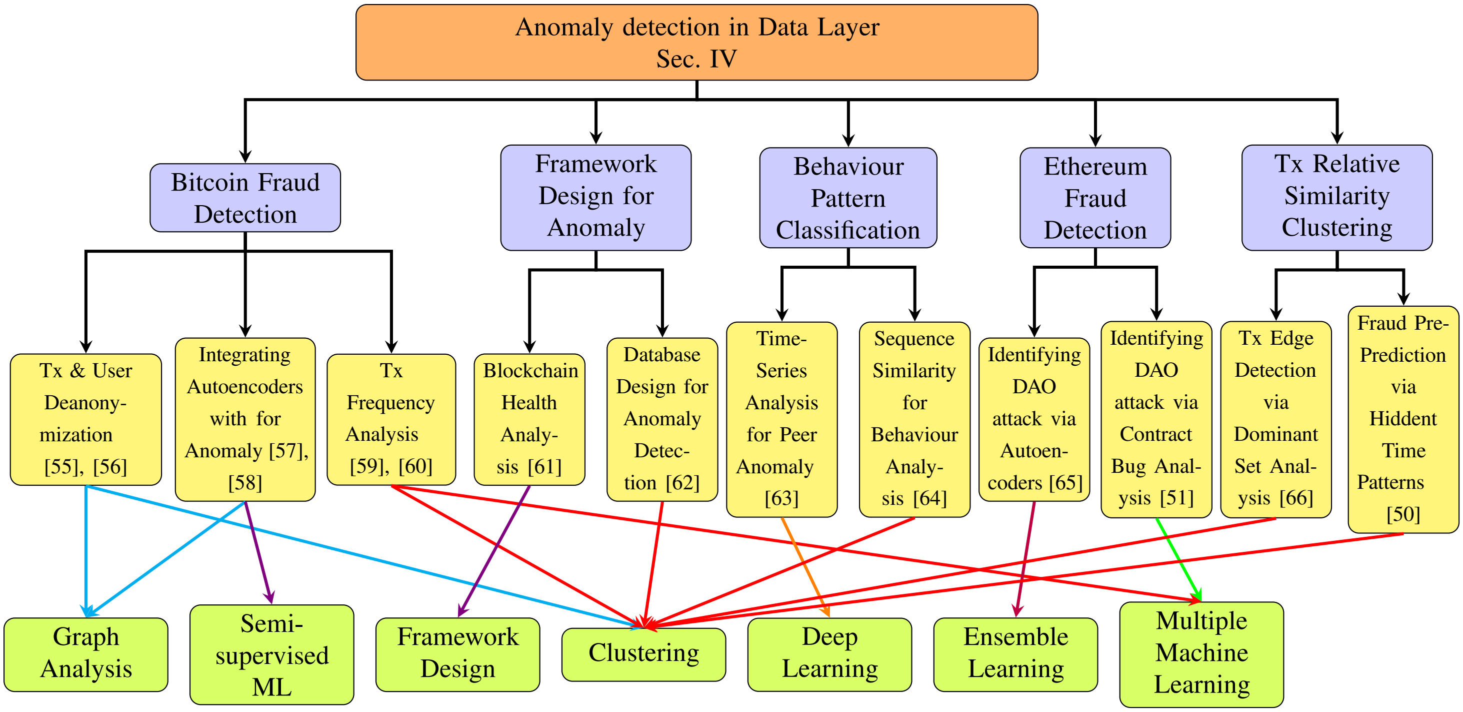
Anomaly detection in Data Layer
Sec. IV
Bitcoin Fraud Detection
Framework Design for Anomaly
Behaviour Pattern Classification
Ethereum Fraud Detection
Tx Relative Similarity Clustering
Tx & User Deanonymization [55], [56]
Integrating Autoencoders with for Anomaly [57], [58]
Tx Frequency Analysis [59], [60]
Blockchain Health Analysis [61]
Database Design for Anomaly Detection [62]
Time-Series Analysis for Peer Anomaly [63]
Sequence Similarity for Behaviour Analysis [64]
Identifying DAO attack via Autoencoders [65]
Identifying DAO attack via Contract Bug Analysis [51]
Tx Edge Detection via Dominant Set Analysis [66]
Fraud Pre-Prediction via Hiddent Time Patterns [50]
Graph Analysis
Semi-supervised ML
Framework Design
Clustering
Deep Learning
Ensemble Learning
Multiple Machine Learning

Fig. 4: Classification of Blockchain Anomalies from Perspective of Data Layer

deeply and claimed that their proposed model can help in detecting money laundering pattern in an efficient manner. Another similar article, that focuses of deanonymization and identification of anomalous users in bitcoin network has been carried out by Shao *et al.* in [56]. The article proposed a novel mapping based system which learns on the basis of address similarity from perspective of a compact Euclidean space. The work identifies k-similar addresses and then use the proposed model to identify presence of an anomalous participant.

From perspective of blockchain feature analysis, a unique work has been presented by Nan and Tao in [57]. The authors focused mainly on detecting mixing and de-mixing services for Bitcoin cryptocurrency. By using the features of graph embedding, authors proposed a mixing identification model in which one can figure out services on basis of their specified features. They further tested the model on real Bitcoin datasets to show the effectiveness, and the simulation results demonstrate that the proposed model achieves a good performance secure for outlier detection. One more considerable work from perspective of unseen patterns detection from blockchain network traffic have been presented by Kim *et al.* in [58]. The work is slightly different from traditional works, as it highlights the use of network traffic rather than stored ledger data. In the proposed model, authors developed an engine which collects multi-dimensional data stream in an organized and periodic manner. The collected data is then fed to a semi-supervised learning model, which detects novel patterns from the blockchain data. The work further highlights that they introduced a profiling-based engine for efficient anomaly detection, which is implemented over autoencoder. The presented model is further tested and compared with other similar models, such as DNN, LR, GB, OC-SVM, and RF, and it can be visualized that it outperformed other models in terms of training time and detection.

Another critical work to identify high yielding programs for Bitcoin investments has been presented by Toyoda *et al.* in [66]. Authors devised certain Bitcoin features and then ranked the transactions on the basis of these features to identify specific actions. To elaborate it further, authors distributed the Bitcoin payback into different distribution classes and identified that whether the payback amount is from a high yielding investment program, or it belongs to some other category, such as donation, exchange, mining pool, faucet, etc. Similar to identification of high yielding investment programs via analysing Bitcoin transactions, another article targeting analysis of transaction history for address classification has been carried out by Lin *et al.* in [60]. The article works over proposing of novel features to develop abnormality detection classification models for Bitcoin. Authors further used these features to carry out training of supervise machine learning models, which are then used to carry out prediction and evaluation of anomalous Bitcoin addresses.

B. *Generic Framework Design for Blockchain Anomaly*

Since, blockchain is a well-applied field and now it has application in almost every aspect of life ranging from healthcare to smart grid. Therefore, apart from developing models just for Bitcoin transactions, it is also important to design models to check health and anomaly of generic blockchain models as well. One such work from perspective of designing of a visualization tool for anomaly detection in blockchain based IoT systems has been presented by Song *et al.* in [61]. The proposed framework has two major aims, determination of health and detection of anomaly in blockchain network.

TABLE II
ANOMALY DETECTION IN DATA LAYER.

Domain	Ref No.	Contribution	Detected Anomaly	Anomaly Factors	Blockchain Type	Platform Language	Applications	Dataset	Complexity
Bitcoin Fraud Detection	[55]	Developed a system for visual analysis of Bitcoin Flow	• Malicious Tx • Pure & impure circulated money	• budget • Purity • Transfer Analysis	Public	Python	• Cryptocurrency	Bitcoin Database	–
	[56]	Address similarity mapping via Euclidean space	• Anomalous Bitcoin Users	• K-similar addresses	Public	N/S	• Cryptocurrency	Bitcoin Database	–
	[57]	Feature based identification of Bitcoin mixing-demixing	• Graph intermediate point for mixing services	• Tx graph reconstruction • Outlier & clustering	Public	N/S	• Cryptocurrency	Bitcoin Database	$O(n^2)$
	[58]	Profiling based anomalous pattern detection from multi-dimensional data	• Anomalous Tx Patterns	• Network traffic statistics • User and Tx profiling	Public	Python	• Cryptocurrency	Bitcoin Database	$O(n^2)$
	[59]	Tx pattern analysis for anomalous activity in HYIP	• Uneven payback rate • Uneven Tx frequency	• Address clustering • Feature gain	Public	R	• Cryptocurrency	Bitcoin Database	–
	[60]	High order Tx moment based anomaly detection	• Malicious Tx records	• Tx moments • Tx history summary	Public	Python	• Cryptocurrency • Crowdsensisng • Stocks	Stock Trading Data	–
Framework Design	[61]	Analysing Tx & block interval to measure healthiness & Anomaly of IoT blockchain	• False IoT data storage	• Block No & Tx interval	Public	N/S	• IoT	Real-Time IoT Data	–
	[62]	A scalable data analysis tool design for Blockchain via MySQL and MongoDB Databases	• Uneven Tx rate	• Tx fee • Address tags	Public	Multiple	• Cryptocurrency	Bitcoin, Ethereum	–
Behaviour Pattern Classification	[63]	Classified peers of blockchain w.r.t their behaviour via Deep Learning	• Non-similar peers	• Batch size • Class label prediction	Public	N/S	• Cryptocurrency	Bitcoin Database	–
	[64]	Clustering based behaviour analysis for blockchain nodes	• Anomalous behaviour sequences	• Sequence similarity	Public/ Private	N/S	• Cryptocurrency • Stocks • IoT	Stock Trading Data	–
Ethereum Fraud Detection	[65]	Strengthening encoder-decoder model against DAO attacks	• Decentralized autonomous organization	• Block size • Average gas usage	Public	N/S	• Cryptocurrency • Non-trusted organizations	Real & Synthetic Eth Data	–
	[51]	Malicious Tx behaviour detection via supervised learning	• Malicious Ethereum nodes	• Tx gas analysis • Tx timestamp	Public	Python	• Cryptocurrency	Ethereum Data	–
Tx Relative Similarity Clustering	[66]	Identified common behavioural nodes via dominant set analysis	• Uneven Tx behaviour	• Tx edges • cluster edges • similar dominant set	Public	MATLAB	• Cryptocurrency • Crowdsensing	N/S	–
	[50]	Identifying hidden time patterns from Blockchain Tx	• Anomalous Tx • Anomalous behaviour nodes	• Tx logs • Levenshtein distance	Public	N/S	• Cryptocurrency	N/S	–

From the viewpoint of health classification of blockchain, authors analysed height, number of transactions, and generation interval of each block in the network. Similarly, to identify malicious activity on blockchain, authors used the data of specific number of events alongside IoT data statistics, which helped visualize and identify a prospective anomalous node. A more generalized work towards development of anomaly detection tool for generic blockchain network has been presented by authors in [62]. The authors first discuss the generic model of blockchain and then highlighted and exposed certain anomalies, such as anomalous metadata, transaction fees, and address tags. Afterwards, authors implemented and validated their framework on Ethereum and Bitcoin data, which are two major blockchain models nowadays. The work developed APIs and used MongoDB and MySQL databases to evaluate their claims.

C. Behaviour Pattern Classification

A blockchain network usually comprises of a large number of peer participating nodes, which are linked with each other in a distributed decentralized manner. This number is even more abundant in public blockchains, where anyone from any part of the world can join with no or very minimal verification. Thus, ensuring legitimacy in this large group of nodes is fairly challenging, as it is hard to classify if some nodes starts misbehaving. Classifying a node as an anomalous node takes a huge amount of effort and time, and still then, the results are not 100% accurate if one uses traditional anomaly detection approaches on blockchain network [68]. Therefore, research works have highlighted that one needs to check and classify behaviour of each participating node in order to get deeper insights and prediction about anomalous blockchain nodes.

A pioneering work in the field of peer behaviour classification has been carried out by Tang *et al.* in [63]. The authors first developed a strong motivation that why traditional anomaly detection approaches such as decision tree & SVM are not reliable and do not produce satisfactory outcome. Afterwards, the authors propose their own time series analysis based deep learning behaviour classification approach and named it as *PeerClassifier*. From the experimental analysis and evaluation, authors demonstrated that their proposed approach shows significant improvements in accuracy as compared to other traditional learning approaches. Another similar work from perspective of clustering for anomalous behaviour classification has been carried out by Huang *et al.* in [64]. Authors first work over evaluating the similarity list among blockchain peers and then carried out peer identification according to distance among them. Authors further evaluated their work from perspective of precision and compared it with classical approaches to show the improvements. An evaluation of both of the above works show that both clustering and deep learning models provide significant improvement in anomaly detection as compared to traditional approaches. However, this field of behaviour pattern classification is not much discussed in research, and it has a lot of research potential for future works.

D. Ethereum Fraud Detection

Ethereum is the second largest blockchain platform after Bitcoin, but Ethereum is different from Bitcoin as in Ethereum, users can carry out deployment of decentralized applications alongside doing cryptocurrency transactions [69]. Furthermore, in order to deploy these decentralized applications, Ethereum provides its users the facility of decentralized smart contracts, which can be termed as piece of code executed in a decentralized manner [70]. Due to these large number of features and benefits, Ethereum is a big attraction for anomalous peers, as they continuously try to take unfair advantage of these features. This discussion of finding anomalies in Ethereum network can be divided into two major categories, one from perspective of data layer and other from perspective of contract layer. A detailed discussion about smart contract and their anomalies will be given in Section VII. However, in this section, we discuss the data layer aspect of these anomalous behaviours in Ethereum blockchain network.

The first work discussing the detection of vulnerabilities in Ethereum blockchain network by the use of ensemble deep learning have been carried out by authors in [65]. Authors work over strengthening encoder-decoder model for any prospective attack. Authors did so by applying phenomenon of learning and aggregation iteratively at multiple instances, in order to carry out computation of any prospective outlier for every observed reading. Although the work did not provide in-depth theoretical analysis, but the evaluation results show that the proposed model predicts DAO attack in an efficient manner. The second work from perspective of detection of malicious Ethereum accounts via supervised learning approach has been carried out by Kumar *et al.* in [51]. The major focus of the work is to understand behaviour of transactions among Ethereum accounts. Authors further classified the Ethereum network into two subtypes named as smart contract accounts and externally owned accounts. After this classification, authors studied the anomalies in both of these types using supervised machine learning approaches such as random forest, decision free, K-NN, etc. From the evaluated outcomes, it can be seen that the proposed strategy efficiently helps in detection of anomalies in the given conditions.

E. Tx Relative Similarity Clustering

Apart from traditional anomalies, it is also important to study a chain of events in blockchain transactions that has led to a harmful catastrophe. In order to do so, analysing malicious transactions and time-stamps play a very critical role. For instance, by deeply studying behaviour of a catastrophic incident, one can carry out predictions of reoccurrence of a similar or even worse incident in future. This field of prediction via hidden timestamp and malicious transactions is not well-studied in the literature, however, two critical works have been carried out in this domain so far.

One of the work, which focuses over carrying out in-depth transactional analysis by the usage of dominant set analysis have been performed by Awan and Cortesi in [66]. In the work, authors emphasized that learning the behaviour of transaction for each node is critical in identifying and predicting any current and prospective anomaly. Therefore, in order to carry out efficient behaviour learning, authors proposed a dominant set approach which categorizes each transaction in the blockchain network according to the most relevant set. The aim of the proposed model was to achieve high clustering accuracy which is demonstrated in the experimental outcomes of the article. Another innovative work that emphasized over discovering hidden time patterns via clustering in a decentralized blockchain network has been presented by authors in [50]. The major focus of the article is to carry out future predictions via analysing current timestamp patterns in the transactions. Authors first clustered all transactions on the basis of allocated patterns, and afterwards worked over detection anomalous behaviours by observing various parameters, such as distance, etc.

F. Summary and Insights

The role of data layer in successful functioning of blockchain cannot be undermined as it acts as a backbone of blockchain from the perspective of handling and securing data records for blockchain networks. However, on the other hand, the anomalies in the data layer cannot be ignored as well, because they can cause catastrophic consequences otherwise. The anomalies over the data layer can be divided into five subtypes, in which a major proportion is occupied by the anomalies from perspective of Bitcoin and Ethereum fraud detection. The remaining of anomalies in data layer are oriented finding patterns of anomalies behaviours and transactions carried out by blockchain nodes, which have been recorder over blockchain ledger. It is important to identify such anomalies and their corresponding user accounts, so that one can restrict such accounts from carrying out such acts in the future transactions.

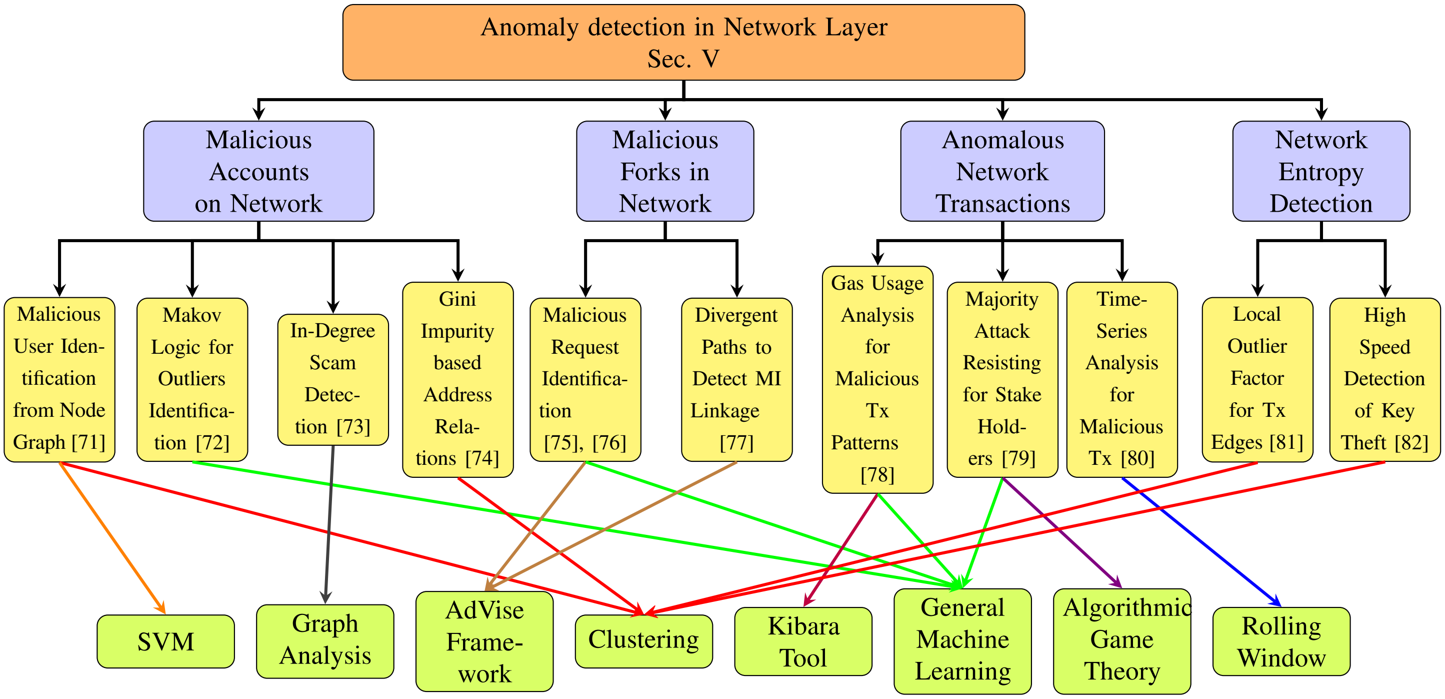
Anomaly detection in Network Layer
Sec. V
Malicious Accounts on Network
Malicious Forks in Network
Anomalous Network Transactions
Network Entropy Detection
Malicious User Identification from Node Graph [71]
Makov Logic for Outliers Identification [72]
In-Degree Scam Detection [73]
Gini Impurity based Address Relations [74]
Malicious Request Identification [75], [76]
Divergent Paths to Detect MI Linkage [77]
Gas Usage Analysis for Malicious Tx Patterns [78]
Majority Attack Resisting for Stake Holders [79]
Time-Series Analysis for Malicious Tx [80]
Local Outlier Factor for Tx Edges [81]
High Speed Detection of Key Theft [82]
SVM
Graph Analysis
AdVise Framework
Clustering
Kibara Tool
General Machine Learning
Algorithmic Game Theory
Rolling Window

Fig. 5: Classification of Blockchain Anomalies from Perspective of Network Layer

V. Anomaly Detection in Network Layer of Blockchain

In this section, we provide an in-depth discussion about the functionalities, limitation, and comparative analysis of anomaly detection models from perspective of the network layer of blockchain (cf. Section II-C for details).

A. *Malicious Accounts on Network*

In a decentralized blockchain network, one of the prime focus for anomalous peers on the network is to mask their identity so that they become untraceable. Anomalous peers try to hide their identities by taking unfair advantages of loopholes in the network. This category of masking the identify while using benefits of the network comes under adversarial activities over the blockchain network layer, therefore, in this section, we first discuss the works which identified deanonymizing the identities of these malicious accounts. One of the critical work in this regard has been carried out by Phan and Lee in [71]. The work focused over analysing behaviours of users on the Bitcoin network by analysing network graphs via different unsupervised learning mechanisms. In order to carry out unsupervised learning on users and transactions graphs, authors used three renowned mechanisms named as support vector machine (SVM), k-Means clustering, and Mahalanobis distance. From the evaluation outcome, authors were able identify few cases of fraud and theft accordingly. Another detailed work which focuses over detection of malicious behaviour on the network of Bitcoin cryptocurrency has been presented in the form of thesis by Frank Jobse in [72]. The work first discussed about suspicious patterns and fraud detection in Bitcoin network, and afterwards, it provides detailed discussion about dataset, data analysis, and sampling techniques used in the evaluation. After that the author presented discussion about Markov logic network its usage in the anomaly detection methodology, and finally the work evaluated the proposed model and compared the work with baseline methods.

One more work from perspective of detection of artificial and strange behaving nodes via user graphs in Bitcoin has been performed by Maesa *et al.* in [73]. The article majorly emphasizes over the outliers in the category of indegree distribution of frequency and remarkably high diameters. Article further discussed the formation of various chain transactions by providing in-depth discussion about various transaction types, such as *BPS, GPS, PS*- transactions. Afterwards, the article evaluated the economic meaning of these transactions and related the anomalies to these transaction types for successful identification of malicious nodes. The final work in this domain of malicious account detection on network layer of blockchain has bene carried out by Chang and Svetinovic in [74]. Authors worked over analysing different transaction patterns with a goal to cluster the addresses with similar ownership information. In order to do so, authors developed a clustering approach and clustered all transactions on the network into five different patterns such as peeling transactions, relay transactions, etc. Another novel this that authors did is that they used Gini impurity measure to evaluate the outcome of the proposed clustering model. Authors further compared the distributions with normal distribution and after applying the proposed model to carry out comparative analysis for the work.

B. *Malicious Forks in Network*

The simplest definition of fork can be termed as disagreement on choosing the best way forward for the blockchain

TABLE III
ANOMALY DETECTION IN NETWORK LAYER.

Domain	Ref No.	Contribution	Detected Anomaly	Anomaly Factors	Blockchain Type	Platform Language	Applications	Dataset	Complexity
Malicious Accounts on Network	[71]	Evaluated three unsupervised learning models for Bitcoin anomaly	• Malicious users • Malicious Tx	• Graph overlapping • Out-degree	Public	N/S	• Cryptocurrency	Bitcoin Database	–
	[72]	Random Forest algorithm based prediction for ground truth cases	• Fraudulent nodes	• Precision & Recall of Tx users data	Public	Python	• Cryptocurrency • Markov Logic Networks	Bitcoin Database	–
	[73]	Exploited structural properties of graph to find unusual patterns	• Artificial Tx patterns	• Common & Uncommon Output Amount • Tx Frequency	Public	N/S	• Cryptocurrency	Bitcoin Database	–
	[74]	Analyzed Bitcoin Tx patterns via clustering of ownership records	• Fraudulent owner clusters	• Gini Impurity Index	Public	Blockseer	• Cryptocurrency	Bitcoin	–
Malicious Forks in Network	[75]	Used meta-data of blockchain systems to tackle eclipse attacks	• Malicious requests • Malicious forks	• Pattern requests	Public	N/S	• Cryptocurrency • IoT	N/S	–
	[76]	Developed a novel Blockchain anomaly detection system	• Malicious code • Malicious requests	• Bandwidth overhead • Request Patterns	Public	N/S	• Cryptocurrency • IoT	Bitcoin Dataset	$O(k)$
	[77]	Link-mining tool based anomaly detection for IoT	• Malicious forks	• Mutual Information	Public	N/S	• Cryptocurrency • IoT	N/S	–
Anomalous Network Transactions	[78]	Using visualized features to detect anomalous gas spent	• Malicious Tx	• Tx throughput • Gas usage	Public/ Private	Node.js Python	• Cryptocurrency • DApps	Ethereum	–
	[79]	Stakeholder activity monitoring via software agents	• Malicious nodes • Double spending	• Tx Payoff	Public	N/S	• Cryptocurrency	Bitcoin Database	–
	[80]	Personalized detection of anomaly via automated Tx signing	• Malicious Tx	• Tx time-frame • Tx frequency	Public	Python	• Cryptocurrency	Ethereum	–
Malicious Forks in Network	[81]	Local outlier factor based clustering for anomaly detection	• Suspicious Tx • Suspicious users	• Tx Edges	Public	N/S	• Cryptocurrency	Bitcoin Dataset	–
	[82]	High-Speed anomaly detection for blockchain using In-GPU cache	• Suspicious Tx	• Abnormal execution time • Avg withdrawal & deposit	Public	CUDA	• Cryptocurrency	Bitcoin Dataset	–

network, this disagreement usually occurs between multiple miners, which control the computational power of blockchain network [83]. As a fork results in splitting of blockchain into two separate chains, therefore, there is a strong possibility that a fork can be carried out for both advantageous and malicious purposes [84]. Some forks can be good for the blockchain network, e.g., division of one organization into two independent organization. However, contrary to this, some forks can be forms as a result of purely malicious and adversarial practices, such as selfish mining, etc [85]. Therefore, it is important to carry out timely prediction of formation of these forks in order to save the network from collapsing.

Certain number of works have been carried out in efficient identification and prediction of these forks, and the first work in this regard has been presented by Pontecorvi *et al.* in [75]. Authors developed a postulate that malicious requests in the network leads to formation of malicious forks and it also paves path for prospective attacks in the network. Therefore, in order to efficiently eradicate these catastrophic conditions, it is important to detect and overcome these malicious requests. In order to do so, authors developed a malicious activity detection tool and named it as AdvISE. The proposed collects and analyses data of blockchain networks and highlights potential adversarial requests for timely action. The extension and further implementation and investigation of this idea has been carried out authors in [76], in which authors developed a thorough anomaly detection and fork leveraging tool for blockchain named it a 'BAD'. Authors developed a complete framework to tackle anomalies at massive level, and also made the proposed system resilient from eclipse attack. Afterwards, authors implemented a thorough testbed, in which they implemented the complete network by using two types of nodes named as full nodes and client nodes. From the evaluation, authors ensured that the proposed model detects anomaly efficiently alongside tackling eclipse attack. Another interesting work focusing over the use of link-mining tool on the basis of blockchain network anomaly detection has been presented by Agure *et al.* in [77]. The aim of the work is to collect blockchain meta-data in the form of network forks, which are further used to figure out prospective anomalous paths in the network. A critical parameter, which is used to aid the experimentation is mutual information (MI), which is used as a measure to ensure efficiency of the proposed model.

C. Anomalous Tx on Network

In the previous section (Sec. IV), we discuss that how anomalous transactions can be identified from the stored data records on data layer. However, research works have indicated that these transactions can be filtered even before recording them to data layer of blockchain. In this section, we discuss the network layer aspects and detection of these anomalous transactions. The first work in this regard has been carried out by Bogner in [78]. The work introduces an online solution based on machine learning for optimal visualization of anomalous transactions on the network. The goal of the experimentation is to design a user friendly visualization tool which will be easy enough to be used by non-technical human operators to identify prospective anomalies in the network transactions. In order to classify a transaction as an anomalous transaction, author categorized the transactions on the basis of gas used for them. The aspect of gas usage is a pure Ethereum terminology, which can be linked as computational effort to complete the transaction operation [86]. A detailed discussion about gas and its consumption is out of scope of this article, interested readers can study an impressive article from Bistarelli *et al.* over the similar topic [87]. Moving back to the detection of anomalies in the network transaction, Bogner built the prototype around Elastic Stack via JSON and developed the frontend using Kibana. The work further performed evaluation experimentation on Ethereum transactions and identified prospective anomalous transactions on the basis of gas usage. A pioneering work which aims to detect malicious transactions intended for the purpose of majority attack has been carried out by Dey in [79]. Author first developed the motivation that in consortium blockchain, the chances of majority-attack is far greater as compared to public networks because any collusion among governing companies can result in a majority attack. Afterwards, the author discussed that the malicious transactions intended for the purpose of majority attack can be detected and prevented timely if appropriate measures are taken. In order to detect and prevent such malicious transactions, author used game-theoretic supervised learning model, which can detect the legitimacy of a transaction and stakeholder on the basis of past transactions. Another pioneering work using machine learning for automated transaction signing to ensure efficient anomaly detection on blockchain network has been carried out by Podgorelec *et al.* in [80]. Authors first emphasized that digital signing of transaction takes time, and that is the prime reason that blockchain is not being integrated in time-critical applications. Afterwards, authors work over proposing an automated and decentralized digital signing framework on the basis of machine learning, which according to the claim not only will make the blockchain efficient but will also detect anomalous transactions at the time of digital signature via time-series machine learning analysis. While evaluating the framework, authors carried out a comparison between the proposed framework and the original process and demonstrated that their proposed framework optimizes blockchain efficiency and anomaly detection.

D. Network Entropy Detection

In fact, this section V as a whole section discusses anomalies and their detection in network layer of blockchain. But works discussed prior to this subsection provides information about analysis of network for some particular issue, such as malicious forks, transactions, accounts, etc. However, in this particular subsection, we discuss that how we can make the whole network secure from generic anomalies and what are the works that have been carried out in this domain so far. One pioneering work by Pham and Lee [81] provides a thorough analysis about integration of anomaly detection in network layer of Bitcoin cryptocurrency. First of all, authors developed a methodology for data collection from Bitcoin network, in which they classified different types of data streams in user and truncation graph. Afterwards, authors developed a k-means clustering model, which uses six features from user-node and three features from transaction node and cluster them accordingly. Afterwards, authors work over identification of anomalies, for which, they worked from perspective of local power degree, outlier factor, and densification laws. Another novel work from perspective of accelerating the process of anomaly detection in for blockchain network has been carried out by authors in [82]. Authors first developed motivation about their work by stating the issues which can be caused due to a malicious transaction if it gets recorded on a tamper proof ledger. Thus, in order to eradicate these issues, authors mentioned that high-speed anomaly detection at the network layer is required, so that one stops malicious transactions from being recorded on the ledger. To facilitate this cause, authors developed a model which uses k-means algorithm to detect anomalous transactions, however, to accelerate the process, authors propose a model which carry out both abnormality detection and feature extraction in GPU memory. The proposed model is then evaluated and compared with traditional models, which showed that the proposed model 37.1 times quicker than the traditional CPU based processing model. TO demonstrate it further authors compared it with traditional GPU based model, which does not carry out feature extraction in GPU, the results showed that the proposed model is 16.1 time speedier than the traditional GPU based model.

E. Summary and Insights

Network layer on blockchain is responsible to carry out activities related to communication and information delivery over the blockchain network. Since, this layer is establishing communication between multiple nodes and is ensuring the legitimacy of transactions and data being transferred, thus the anomalies and frauds for this layer are pretty disastrous and needs strong consideration. The anomalies over the network layer of blockchain can be divided into four subtypes on the basis of their impact. The most prominent type is malicious accounts over the networks, where anomalous users try to pretend as legitimate ones. The next type include formation of malicious forks over the network which is done via either carrying out malicious requests over the network or via making divergent paths. The next two types constitute of carrying out anomalous transactions over the network and to carrying out anomalous behaviours over the network, such as key theft, etc. Irrespective of the type of anomaly, it is important to highlight that as these anomalies are usually being done via communication link, thus, they can be traced and stopped before causing catastrophe, if proper actions are taken.

VI. Anomaly Detection in Incentive Layer of Blockchain

In this section, we provide a thorough literature review from perspective of anomaly detection in incentive/currency layer of blockchain (cf. Section II-C for details).

A. Bitcoin Fraud Detection

Nevertheless, Bitcoin is the most hyped and valuable cryptocurrency so far. Therefore, it will not be wrong to say that whenever one is dealing with anomalies in incentive layer of blockchain, the aspect of Bitcoin anomaly detection cannot be ignored at all. Therefore, in order to discuss anomaly in incentive layer, we initiated our discussion from anomalies highlighted by researchers in the context of Bitcoin currency. An initial work as a part of course project has been carried out by Hirshman *et al.* in [88]. The focus of the article is to define and figure out atypical transaction patterns in Bitcoin currency. In order to do so, authors performed relational checks on Bitcoin data in order to figure out the roots of coin mixing for any anomalous transaction. In order to do so, authors used K-means clustering model and developed and identified different clusters on the basis of degree variance and hub count. Finally, authors examined certain real-time splits to figure out the level of anomaly from various coin mixing services alongside identifying certain intermediate addresses involved in the malicious transaction.

A very interesting work from perspective of anomaly detection in Bitcoin network in the existence of label scarcity has been carried out by Lorenz *et al.* in [89]. The focus of the article is to basically detect money laundering patterns in cryptocurrencies by specially focusing over Bitcoin. Authors first highlighted that traditional unsupervised money laundering detection models are not good enough for Bitcoin network, and therefore, authors designed a supervised learning models to detect illicit laundering pattern in the network. In order to evaluate the proposed model, authors worked over reporting unlawful F1-score in a unit time-step performed during the test. The reported scores are then used to identify anomalous users in order to take action against them. Another similar work that focuses over the usage of global and local outliers for the identification of Bitcoin fraud has been carried out by Monamo *et al.* in [90]. The authors first highlighted the issue that lack of class labels in Bitcoin network is one of the root cause due to which it is hard to figure financial anomalies in the network. Afterwards, authors discussed fraud in Bitcoin network from both global and local perspective. Then in order to identify the anomalies, authors highlighted the use of both unsupervised and supervised models for the identification of global and local outliers. For unsupervised models, authors worked over $k-means$ and $kd-trees$ clustering, in which they identified that the clustering mode of '8' gives the optimal result. Similarly, for supervised learning models, authors used GLM logistic regression, boosted logistic regression, and random forest. Authors further emphasized the use of supervised learning models on the basis of findings and the detection accuracy of these models.

Till now, we discuss the use of outliers and similar other patterns, but a very different work from the perspective of use of Hypergraph for malicious user identification of Bitcoin has been carried out by authors in [91]. The article focuses over identification of specific exchange patterns of Bitcoin with respect to its spending and acquisition. To study it further, authors work over building a classification model which discriminate various feature features and the major focus was to identify the root of a malicious address, which means verification of an address that whether it is owned by a specific exchange or not. The basic reason behind designing and analysing of 2-motif hypergraph is to figure out hidden pattern via learning models. To evaluate it further, authors used five learning models named as linear SVM, perceptron, random forest, logistic regression, and AdaBoost. Authors compared these models on the basis of their precision, recall, and F1 score. A final work that evaluates and proposes the anomalous aspects in Bitcoin wallets have been carried out by Zambre and Shah in [92]. The developed project aims to identify the malicious users and entities who are targeting vulnerable wallets and accounts of Bitcoin users with an intention to compromise them for illicit purpose. The article gave examples of certain robberies and thefts that have been carried out over Bitcoin network so far, and afterwards, authors used k-means clustering for malicious user identification. In order to get efficient results for this k-means clustering, authors first extracted 21 available features from available Bitcoin data and afterwards evaluated and categorised users on the basis of occurrence frequency. Authors mentioned that they were able to detect the illicit behaviour with 76.5 percent accuracy.

Despite of these works in the domain of Bitcoin anomaly detection in incentive layer, it is important to mention that this domain still lacks a lot and there is a huge need for more research in order to make the cryptocurrency more secure and trustworthy for future users.

B. Malicious Blockchain Accounts

From the perspective of general blockchain networks, it is equally important to identify malicious accounts on the network especially the malicious activities of these accounts over the incentive layer. Because if a malicious account is able to compromise incentive layer, then it can lead to catastrophic outcomes for the whole network. Since blockchain is immutable, therefore, it is also equally important to detect the malicious transactions before updating them on the ledger, and for this, we need highly efficient models, that scrutinize transactions at a high pace. One such work has been carried out by Morishima in [93]. The work basically revolves around use of GPUs to speed up anomaly detection process in blockchain network. In the article, authors first used the concept of fixed size subgraphs, which centric towards blockchain users in order to develop an anomaly detection model. However, use

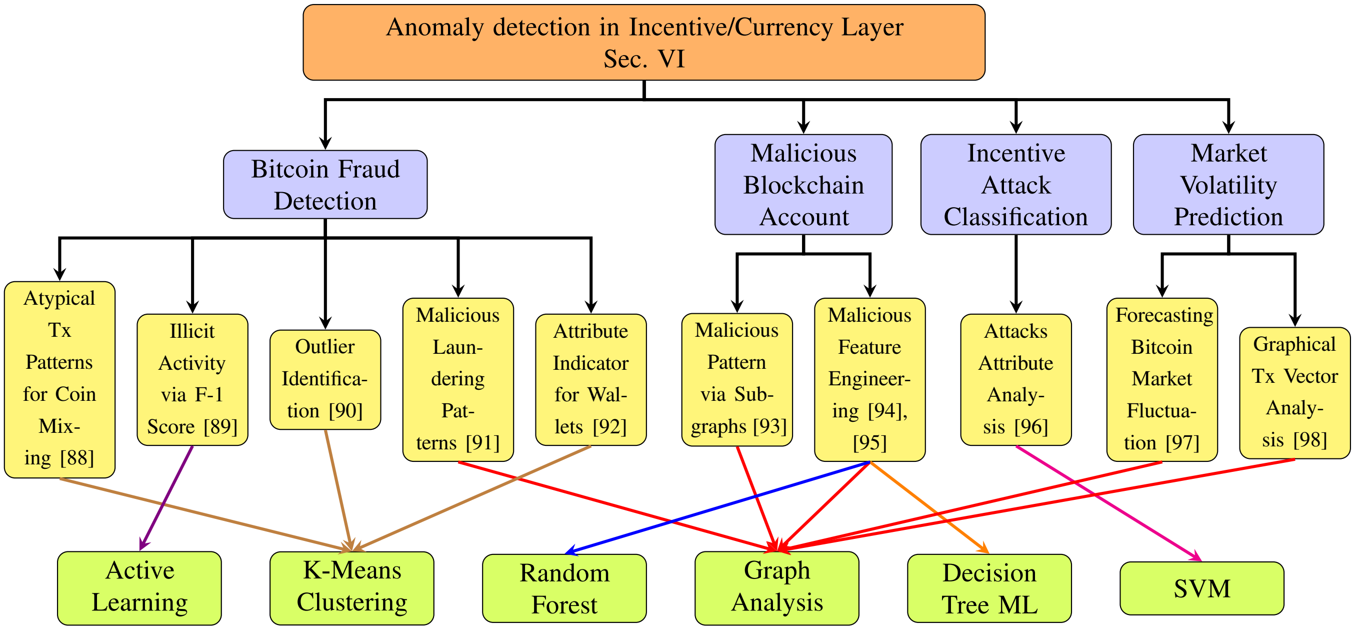
Anomaly detection in Incentive/Currency Layer
Sec. VI
Bitcoin Fraud Detection
Malicious Blockchain Account
Incentive Attack Classification
Market Volatility Prediction
Atypical Tx Patterns for Coin Mixing [88]
Illicit Activity via F-1 Score [89]
Outlier Identification [90]
Malicious Laundering Patterns [91]
Attribute Indicator for Wallets [92]
Malicious Pattern via Subgraphs [93]
Malicious Feature Engineering [94], [95]
Attacks Attribute Analysis [96]
Forecasting Bitcoin Market Fluctuation [97]
Graphical Tx Vector Analysis [98]
Active Learning
K-Means Clustering
Random Forest
Graph Analysis
Decision Tree ML
SVM

Fig. 6: Classification of Blockchain Anomalies from Perspective of Incentive Layer

of these types of subgraphs normally result in increase in the total time of execution for any model. Therefore, in order to overcome this execution time issues, authors work over proposing GPU oriented structural graphs which speed up the execution and detection process for a timely action. Authors further evaluated their proposed model over 300 million transactions and claimed that their proposed model provides 195 times faster execution time as compared to traditional methods. Similarly, from perspective of detection accuracy, authors claimed that their true positive rate is substantially larger as compared to traditional anomaly detection models due to the use of GPU and developed subgraphs.

Apart from attacks over users' identities, attackers and adversaries also try to play with different transactional features in order to compromise accounts so that they can steal critical assets or cryptocurrencies. Therefore, playing with different features to identify anomalies in blockchain is a critical aspect which needs more consideration. Till now, two critical works [94], [95], have been carried out in this domain so far, one from perspective of malicious features for compromised wallets identification and other from the point of view of top rated feature engineering for timely user anomaly identification. Nevertheless, certain other works also highlighted feature extraction in different aspects, but these two work purely focused over feature identification in order to identify malicious accounts falling under incentive layer, therefore, we are describing them here in detail. The first work in the domain has been carried out by Baek *et al.* in [94]. The article mainly focuses over investigating Binance platform, which is one of the most commonly used cryptocurrency platform nowadays. Authors evaluated more than 38,000 wallets in order to identify transactions for malicious purposes. In order to enhance the detection, authors worked over feature engineering, in which they identified and used the most suitable features for unsupervised learning model, such as random forest. The work further advocates labelling of the flagged cryptocurrency wallets and transactions for future transactions. In this way, it will become easier to detect any malicious activity from the flagged accounts in future and by this way one can take timely action and can prevent some catastrophe in future.

The other work in the domain of explicitly feature engineering for detection of malicious accounts have been carried out by Farrugia *et al.* in [95]. The basic motivation of the work is to figure out the top features which have the largest impact on the outcome of anomaly detection model. Therefore, after thorough examination and evaluation authors identified 'Min received value', 'time difference', and 'total balance' as the three most influential features specifically for Ethereum blockchain. Alongside doing this feature engineering, authors also work over proposing an effective method to detect the malicious accounts, for which authors used XGBoost, which basically is a method of ensemble machine learning via decision tree. Considering this discussion about malicious accounts detection on incentive layer, it can be concluded that researchers are deeply exploring this domain, however, a large number of prospects needs to be identified till now.

C. Incentive Attack Classification

Incentive layer is prone to many attacks ranging from double spending to DDoS attack, etc. However, majority of works discussed above focused majorly over either identification of a particular attack, or identification of some sort of anomalous behaviour in the blockchain model. However, considering the diverse range of attacks, it is equally important for an anomaly detection model to pin point the type of attack which is being carried out in the network. From the perspective of incentive layer, one such incredible work has been carried out by Sayadi *et al.* in [96]. The work focuses over using two separate machine learning models to first detect the anomalies

TABLE IV
ANOMALY DETECTION IN INCENTIVE LAYER.

Domain	Ref No.	Contribution	Detected Anomaly	Anomaly Factors	Blockchain Type	Platform Language	Applications	Dataset	Compl-exity
Bitcoin Fraud	[88]	Clustering hub based coin mixing detection	• Atypical Tx patterns	• Hub count • Tx value variance	Public	N/S	• Cryptocurrency	Bitcoin	–
	[89]	Detected money laundering via supervised learning	• Illicit Tx	• F-1 Score	Public	Python	• Cryptocurrency	Bitcoin	–
	[90]	Detecting global & local frauds of Bitcoins via supervised and unsupervised learning	• Malicious Tx groups	• Account inputs • Account outputs	Public	N/S	• Cryptocurrency	Bitcoin	–
	[91]	Identifying laundering patterns from Hypergraph with high accuracy	• Malicious Tx patterns	• Exchange addresses	Public	Python	• Cryptocurrency	Bitcoin	–
	[92]	Analyzed Bitcoin data to detect occurred fraud	• Compromised wallets	• Tx frequency • User occurrence frequency	Public	N/S	• Cryptocurrency	Bitcoin	–
Malicious Blockchain Accounts	[93]	GPU based high speed anomaly detection from user-centric subgraphs	• Abnormal Tx	• Tx Edges	Public	CUDA	• Cryptocurrency	Ethereum	–
	[94]	Random Forest based malicious node detection	• Tx with discernible purpose	• Wallet labelling	Public	Python	• Cryptocurrency	Bitcoin	–
	[95]	Using XGBoost classifier to detect malicious features	• Illicit accounts	• Avg Tx value • Received & sent values	Public	Python	• Cryptocurrency	Ethereum	–
Incentive Attack Classifica-tion	[96]	Used one-class SVM and K-Means on electronic Tx data	• Malicious Tx	• Tx number, address, & volume	Public	Python	• Cryptocurrency	Bitcoin	–
Market Volatility Prediction	[97]	Predicting volatility & return in Bitcoin market via network theory	• Price variance	• Market In-Out Tx • Impulse response	Public	N/S	• Cryptocurrency	Bitcoin	–
	[98]	Identified mining based market manipulation of Bitcoin	• Fluctuating exchange rate • Abnormal Tx patterns	• Account BTC production	Public	N/S	• Cryptocurrency	Bitcoin	–

and then the second to classify it up further. In order to detect outliers in the transactions of blockchain, authors used one class SVM also known as OCSVM, which basically separates novelty outliers on the basis of hyperplane distance among the transactions. Afterwards, the model basically labels them and feed to the next classification model, for which authors used K-means clustering. This K-means clustering is basically a further extension via which authors picked and classified the anomalies into different types of attacks. From the selected anomalies, authors were able to identify the presence of double spending, DDoS, and 51% vulnerability from the identified labels.

It is important to highlight that till now very few works focused over classification of attacks on blockchain network, and only one work emphasized it thoroughly purely from the perspective of incentivization and incentive layer. Therefore, it will not be wrong to say that this direction still has huge potential, and a large amount of research needs to be carried out in which researchers are required to develop such models which can be used to predict and pinpoint the exact attack being carried on the network, especially on incentive layer.

D. Market Volatility Prediction

While talking about incentive or currency layer of blockchain, the aspect of market control, manipulation, and volatility cannot be ignored because it is one of the key aspect over which a huge amount of investment depends upon [99]. Nevertheless, cryptocurrencies, especially Bitcoin and Ethereum have a large amount of market capitalization and this capitalization is continuously increasing. E.g., the total market capital of Bitcoin in 2017 was $18billion, whichincreasedto$599 billion in 2018 [100]. Considering these aspects, it is important to have answers to know more about these cryptocurrencies, especially about market volatility from an anomalous manipulation viewpoint. E.g., is it important to figure out how the prices of these cryptocurrencies evolve and vary? How are these financial markets stabilized? Are there any spillovers in the market? Knowing the answer to these and many other similar questions can help us make better predictions.

From a technological point of view, it is important to identify and predict the occurrence of anomalous factors which can cause a huge market volatility. It is also important to predict the occurrence of a major surge in the network because it can also be due to some adversarial attack on the network, which can further lead to disastrous outcomes. One such to predict market volatility and return from via Bitcoin price movements and transactions movement have been carried out by authors in [97]. The work highlights that predicting Bitcoin market is not fairly simple as plenty of complex aspects are associated with this. In order to carry out efficient prediction, authors developed relationship between the market's volatility and the complexity measures. For complexity measures, authors used the concept of transactions connectivity with regard to number of roles, alongside this, authors used the measures from information theoretic perspective as well. Afterwards, these measures were fed into a prediction model, which first characterises the joint behaviour via vector autoregression and afterwards carry out selection and model estimation accordingly. Another interesting work which focuses over identification of market manipulation of Bitcoin due to adversarial and anomalous identities have been carried out by Chen *et al.* in [98]. To train and develop the market manipulation model, authors picked a previous database of transaction leakage and organized the transactions into three sub-graphs. Afterwards, authors worked over identification of influence of each account on the fluctuation of market in order to identify the most influenced accounts, and they carried out this experimentation with the help of singular value decomposition. From SVD, the authors were able to identify certain base accounts and network, which had a direct relation with the volatility of the network. Similarly, authors were able to identify the types of abnormal transactions which can be carried out between malicious users, e.g., unidirectional, self-loop, bi-direction, polygon, triangle, and star transactions. From the given work, one can efficiently detect the presence of any anomalous factor which can cause market manipulation in the near future.

While discussing market manipulation and volatility in the context of blockchain anomaly detection, it is important to mention that in this section, we only consider works which discuss these aspects from a technological viewpoint. Contrarily, there are plenty of other works, which purely focus over economical or financial viewpoint, therefore, we did not include these articles in our discussion because they were out of scope of this article. Interested readers can study more about economic growth, volatility, and manipulation of cryptocurrencies in the interesting article written by Bariviera and Sola [101].

E. Summary and Insights

Incentive layer is the major driving force in blockchain technology, which motivates participating nodes to take part in mining and other relevant processes. Nevertheless, it is a driving force because of the incentivization, but for anomalous peers, its also one of the most critical layer to target, because they can get direct benefit from this layer in terms of incentives, tokens, etc. Majority of attacks and anomalies over this layer comprise of frauds among cryptocurrencies, such as Bitcoin, Ethereum, etc. Apart from cryptocurrency frauds, the second critical anomaly types is carrying out fluctuations in the market, which can cause huge rise and drop among the shares and trading values of assets and currencies over the blockchain network. Another significant direction towards working over anomaly detection from the incentive layer perspective is to identify malicious accounts of the network carrying out such anomalous activities and flag or ban such accounts in order to prevent them for carrying out fraud or market instability.

VII. Anomaly Detection in Contract Layer of Blockchain

In this section, we provide a detailed review of existing works from the perspective of anomaly detection in contract layer of blockchain (cf. Section II-C for details).

A. Paxos Anomaly in Blockchain

As the name suggests, in this section we will be discussing an anomaly related with dependent transfers, which is categorized under the name of famous consensus protocol 'Paxos'. In order to understand this anomaly a bit further, it is important to understand two major concepts, one is the concept of Paxos consensus and the other is termination of a consensus model in a decentralized blockchain environment. From the perspective of Paxos, it can formally be defined as a family of selected protocols which can be used to reach a consensus in an unreliable processors network [110]. From the perspective of a decentralized network as that of blockchain, the Paxos anomaly originates due to the difficulty in implementation of requests which are applicable only on basis of some conditional guarantees, which will be discussed in detail later in this section. Before moving to the technical discussion about Paxos anomaly in blockchain transactions, it is also important to discuss the concept of finality and termination of blockchain consensus. In a blockchain setting, a consensus is considered terminated deterministically if from the chain structure it can be determined that a new block has been decided, which can also be termed that the transactions inside a block has been verified and commitment had been made over these transactions. This determination and commitment is an integral part of cryptocurrency and other decentralized network because if the determination has not been established, adversaries can take unfair advantage of it by doing attacks, such as double spending, etc.

Moving towards discussion of similarities between Paxos anomaly and blockchain anomaly, it is important to highlight a critical and thorough work carried out by Natoli and Gramoli in [102]. Authors provided a through analysis that how the asynchronous nature of blockchain and message delays can cause a major issue in termination of consensus, which can then lead to start of two simultaneous chains where both the miners agrees simultaneously over their own 'k' set of blocks. Authors further evaluated that this problem accelerates and becomes more catastrophic in case of dependent transactions especially in case of a private blockchain, which can further lead to double spending attack and uncommitting of transactions. Authors further developed a complete model to study this effect in which they evaluated by automating the anomaly reproduction in the decentralized network at different difficulties involved in mining the block. Through the evaluation, authors shows the hazardous effects which can be caused if proper actions are not taken. Finally, authors discussed a

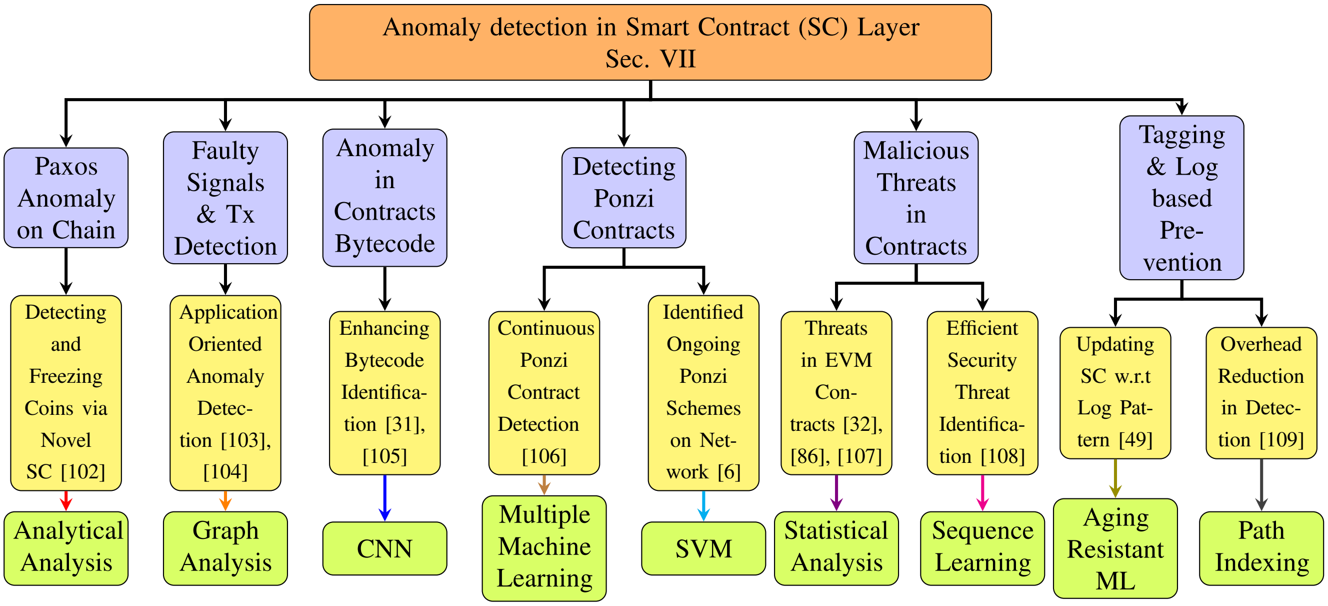
Anomaly detection in Smart Contract (SC) Layer
Sec. VII
Paxos Anomaly on Chain
Faulty Signals & Tx Detection
Anomaly in Contracts Bytecode
Detecting Ponzi Contracts
Malicious Threats in Contracts
Tagging & Log based Prevention
Detecting and Freezing Coins via Novel SC [102]
Application Oriented Anomaly Detection [103], [104]
Enhancing Bytecode Identification [31], [105]
Continuous Ponzi Contract Detection [106]
Identified Ongoing Ponzi Schemes on Network [6]
Threats in EVM Contracts [32], [86], [107]
Efficient Security Threat Identification [108]
Updating SC w.r.t Log Pattern [49]
Overhead Reduction in Detection [109]
Analytical Analysis
Graph Analysis
CNN
Multiple Machine Learning
SVM
Statistical Analysis
Sequence Learning
Aging Resistant ML
Path Indexing

Fig. 7: Classification of Blockchain Anomalies from Perspective of Contract Layer

prospective smart contract based solution in which authors provided certain examples showing the specific conditions and statements which can be added in a smart contract to overcome the occurrence of this anomaly, e.g., detecting and freezing of coins at the time of need. Authors further emphasized that modern blockchain networks, such as Ethereum, etc should develop more secure smart contracts so that the possibility of such anomalies can be eradicated completely.

B. *Faulty Signals & Tx Detection*

Researchers are actively working in development of modern blockchain systems, which are paving paths towards development of 'Blockchain 3.0' [111]. One such example is the development of state of the art blockchain based Internet of Things (IoT) applications (such as Industrial IoT and supply chain), which are usually autonomous and can perform duties as assigned and programmed. However, maintaining the appropriateness of data and transactions from these applications is vital because important decisions are being taken on the basis of reported data. Contrarily, in case if arises any fault arises or some adversary tries to corrupt the data coming from the application nodes, it is way too important to identify it as an anomaly before it gets recorded on the immutable blockchain network.

In order to overcome this anomaly detection issue, authors in [103] worked over proposing a deterministic smart contract based anomaly detection in blockchain based IoT networks. Authors work over development of a collaborative learning approach in which authors used the functionalities of probabilistic dictionary learning to figure out existence of a particular faulty singals anomaly in the blockchain based IoT network. To carry out the work, authors first formulated the problem by developing anomaly score on the basis of clients, network participants, and used dataset. Afterwards, authors developed the algorithm for density estimation in the blockchain network in a collaborative manner, which further led to the development of protocol which is used to update the parameters data. From the evaluated models, authors claimed that the proposed model identified anomaly in a more accurate manner as compared to the previous work. Another similar work from perspective of anomaly detection over supply chain network operating over Hyperledger Sawtooth have been carried out by Oh *et al.* in [104]. The work focuses over identification of anomalies via graph analysis, which is done with the help of a smart contract. The smart contract is designed in such a manner that it identifies that whether order of a transaction is correct, and it has all the necessary linkage with its predecessors. If the transaction obeys all the conditions, then it is recorder over the ledger, elsewise, it is flagged as an anomalous transaction. Authors successfully implemented this anomaly detection notion over Hyperledger Sawtooth in the form of an additional layer, which can be integrated with blockchain platforms to ensure the capability of anomaly detection.

C. *Anomaly in Contracts Bytecode*

With the development of Ethereum smart contracts, a plethora of opportunities and future directions which require the integration of decentralized services for their efficient functionalities can visualized. Since, smart contracts is a deterministic piece of code which cannot be stopped once it start execution on the blockchain network, therefore, it is equally important to ensure that the outcome of a particular smart contract is in the favour of blockchain network, and it will not cause any catastrophe. Therefore, it is important to figure out the anomalies and hazardous elements in a blockchain smart contract before its deployment on the network. Since, Ethereum intiatied this concept of smart contracts in its DApps, thus, nowadays the majority of everyday applications and research we see usually use smart contracts deployed on Ethereum. In Ethereum, the smart contracts are usually

TABLE V
ANOMALY DETECTION IN CONTRACT LAYER.

Domain	Ref No.	Contribution	Detected Anomaly	Anomaly Factors	Blockchain Type	Platform Language	Applications	Dataset	Complexity
Paxos Anomaly in Blockchain	[102]	Highlighted a complex immutability related anomaly & proposed coin freezing.	• Malicious risky Tx • Uncommitting Tx	• Swap frequency • Tx movement	Public	EVM	• Digital Assets • Cryptocurrency	N/S	–
Faulty Signal and Tx Detection	[103]	Graphical lasso & collaborative dictionary learning based anomaly detection in industrial data.	• Faulty data files	• Logarithmic loss • Sample accuracy	Private (Permissioned)	Hyperledger Fabric	• Digital Assets • Industrial Data Storage	Real-world Data	–
	[104]	Graph analysis based anomalous Tx detection from supply chain data.	• Faulty unordered transactions	• Tx life cycle • Tx order	Private (Permissioned)	Hyperledger Sawtooth	• Digital Assets • Supply Chain	Real-world Data	–
Anomaly in Contracts Bytecode	[105]	Converted bytecode to RGB for efficient anomaly extraction.	• Compiler bugs in contracts	• Malicious smart contracts	Public	Solidity	• Cryptocurrency	Ethereum Contracts	–
	[31]	Identified Malicious Smart Contracts by Assigning Labels on basis of bytecode	• Anomalous bytecodes	• Malicious smart contracts	Public	Solidity	• Cryptocurrency	Ethereum Contracts	–
Detecting Ponzi Contracts	[106]	Developed model to predict Ponzi smart contracts from day zero.	• Fraudulent smart contract	• Precision • Recall • F-1 Score	Public	EVM	• Cryptocurrency	Ethereum Data	–
	[6]	Developed model to predict scamming smart contracts in Ethereum.	• Malicious Tx • Malicious SC body	• Ponzi detection count	Public	EVM	• Digital Asset • Cryptocurrency	Real-world Data	–
Malicious Threats in SC	[32]	Developed a taxonomy of honeypots of Ethereum smart contracts.	• Balance disorder • Inheritance disorder	• Hidden traps in SC	Public	Python	• Cryptocurrency	Ethereum SC data	–
	[86]	Identified 20 Defects in Ethereum smart contracts.	• Security, Availability, Performance, Maintainability, and Re-usability Defects	• Hidden defects in SC	Public	N/A	• Cryptocurrency	Ethereum SC data	–
	[107]	Linked Smart Contract Defects to Prospective Unwanted Behaviour.	• Contract dependencies	• Hidden attacks in SC	Public	Solidity	• Cryptocurrency	Ethereum SC data	–
	[108]	Sequentially learning smart contracts to find weaknesses.	• New attack trends	• Precision • Recall • F-1 Score	Public	EVM	• Cryptocurrency	Ethereum Data	–
Tagging & Log based Anomaly Prevention	[49]	Developed a self-adaptive model to detect log anomaly in smart contracts.	• Malicious data storage	• Time complexity	Public	N/S	• Cryptocurrency • Digital assets	Real-world Data	–
	[109]	Detected & prevented abnormal control flow in Ethereum smart contracts.	• Control paths	• Gas consumption • Overhead	Public	EVM	• Digital assets • Cryptocurrency	Multiple datasets	–

written in Solidity language, which is further compiled to EVM bytecode at the time of deployment on blockchain. Therefore, the key time to pick a malicious smart contracts is to identify adversaries hidden in the bytecode at the time of deployment.

One such work towards identification of anomalies in bytecode of Ethereum smart contracts has been carried out by Huang in [105]. Unlike other similar works, Huang did not focus primarily over extraction of novel features for efficient identification, instead, the major focus of the article is to reduce the overall labour cost associated with identification of anomalies in the Ethereum bytecode. In order to do so, authors first work over translation solidity bytecode into an RGB code, which is further used to develop an encoded image of the fixed-size. The RGB image is then fed to CNN for training, which automatically extract features, carry out learning, and then carry out detection of bugs of compiler at the time of execution of smart contract. In this way the proposed work is able to identify bugs in a more cost effective and efficient manner as compared to previous works. Another critical work to enhance bug prediction accuracy for smart contracts have been carried out by Kim *et al.* in [31]. Authors worked over analysing smart contract bytecode in order to categorize and attribute them in the form of tags for swift identification. In order to do so, the authors used learning model comprising of five different stages ranging from pre-training stage to inference stage. In order to evaluate the proposed methodology, authors used code examples EtherScan and Google BigQuery datasets. From the outcome results, it can be seen that the authors were able to successfully classify smart contracts bytecode on the basis of attribute present in them.

D. Detecting Ponzi Contracts

Since the advent of blockchain and cryptocurrencies, adversaries are continuously trying to take unusual and illegal advantages of certain hidden functionalise of it which common people are not aware of. One such type of fraudulent model is development of Ponzi scheme on the decentralized blockchain network, which not only effects an individual, but it also effects the economy in a deeper level [112]. Due to such extreme outcomes, some countries are not even allowing cryptocurrencies, or they have strong scrutiny of activities being carried out on the cryptocurrency network. Recent studies have also approved and highlighted such activities due to whom the participating users have lost millions of dollars in cryptocurrencies. Formally, a Ponzi scheme can be regarded as a malicious investment scheme via which the revenue for old investors is generated via investment from new participants and as a whole the company or scheme is not generating any external revenue [113]. Similarly, in the decentralized blockchain network, these schemes try to fool participants via using dissimilar smart contracts which seems legit and convincing, these contracts are also known as *Ponzi scheme contracts*. Therefore, considering the prospective catastrophic outcomes, it is a dire need to develop such models which predicts the existence and execution of such Ponzi smart contracts in the network, so that appropriate action could be taken against them within time. One such work towards development of data mining models to develop Ponzi smart contracts on Ethereum blockchain has been carried out by Jung *et al.* in [106]. Authors first highlighted the functioning and basic methods being used in Ponzi smart contracts by specifically focusing over Ethereum, and afterwards authors worked over building a dataset of Ponzi contracts on the network. Then the authors used these malicious Ponzi contracts to pick out specific features, for which authors used the transactions and the compiled code on the network for these malicious contracts. Then authors works over development of classification model, which efficiently predicts the presence or absence of malicious and Ponzi factors in a smart contract. Authors carried out evaluation of their proposed model for 250 days and the outcome results identified that it predicts malicious contracts in an efficient manner. Another similar work focusing over, exploitation of Ethereum blockchain to identify Ponzi contracts have been carried out by Chen *et al.* in [6]. In order to diversify their search, authors first manually picked 200 Ponzi smart contracts by analysing around 3,000 available Ethereum contracts. After that, authors extracted two malicious feature son the basis of operation codes and history of transactions. Afterwards, authors used data mining tools to develop the complete model which classified each new smart contract as Ponzi or safe. From the analysis, authors highlighted that more than 500 Ponzi schemes are currently being operated on the blockchain network.

It is important to highlight that the majority of the work towards development of Ponzi smart contracts has been carried out from Ethereum perspective because they are the first ones to introduce the feature of smart contract in blockchain, therefore, they are most vulnerable one. However, these Ponzi schemes are not just limited to Ethereum, as they are spreading to other cryptocurrencies and blockchains as well. Therefore, there is a dire need to develop such detection models, which accurately detect such Ponzi smart contracts before occurrence of any catastrophe.

E. Malicious Threats in Smart Contracts

Every smart contract being executed in blockchain network has its own dependencies and can affect the blockchain network in its own way. Similarly, once executed, it is impossible to stop the functioning of smart contract. Therefore, certain time adversaries take unfair advantage of this feature and try to add certain malicious threats and honeypots in smart contracts which can cause a serious harm to the network or individual. Therefore, destruction and timely identification of such smart contracts is mandatory to keep the network safe from adversaries. One such work towards identification of malicious honeypots on Ethereum smart contracts have been carried out by authors in [32]. The authors developed a tool for honeypot identification and named the tool as 'HoneyBadger'. To elaborate their concept a bit further, authors proposed a formal definition of honeypots, in which they described that a honeypot in a specific type of smart contract which tricks users to give their funds to attacker in the exchange of some leaked arbitrary funds. In order to attract audience, the attacker firs deploys a smart contract which seems to be giving funds to the executer. Then, the victim falls in prey in the greed of getting more funds, and thus he/she transfers the required sum to the attacker. Finally, the attacker withdraw both the funds and the original funds, and the victim is left with nothing in hand. Through visual examples, authors explained the severity of the situation and thus to overcome this, authors developed a complete taxonomy of such honeypots which are currently running over the Ethereum network. The authors also carried out an extensive analysis of such honeypots on the basis of their sub-components, such as various disorders and overflows. Another critical work working over identification of critical defects in the smart contracts of Ethereum has bene carried out by Chen *et al.* in [86]. Authors developed the motivation of their work by discussing that certain smart contracts can have defects, and some severe defects can deeply affect the functioning of the whole network and can impact the whole chain. Afterwards, authors identified defects in contracts by analysing gas consumption, keywords filtering, open card sorting, and similar other features. In this way, authors were able to successfully identify 20 critical defects which can cause severe issues in the network. The authors further classified these contracts to five sub-types named as security defects, availability defects, performance defects, maintainability defects, reusability defects. Then authors worked over collection of partitioners' perspective over the identified defects to figure out impact of the malicious types. In this way, authors were able to label 05 critical impacts that can be caused as a result of execution of these malicious contracts. As an extension of this work, authors propose a complete tool and named it as DefectChecker [107]. The proposed tool can detect 08 critical defects in the malicious contracts which could have caused abnormal and unwanted behaviour. Afterwards, worked over using the tool for the identification of level of impact. From the outcomes and experimental results, it can be seen that the proposed model can predict the given contracts with 88.8% F-Score. In this way, the authors were able to conclude that out of 1,65,621 analysed smart contracts, 25,815 had at least one identified as defected.

Till now, the works used analytical and statistical modelling and analysis to identify prospective threats in the smart contracts. However, a detailed model using sequence learning to carry out similar work has been presented by Tann *et al.* in [108]. In order to make the identification effective, authors worked over using long short term memory (LSTM machine learning model. For which, authors first classified the threats of smart contracts and then sequentially modelled them on the basis of opcode sequence. Then after labelling the data through *Maian*, authors used supervised learning to predict the smart contracts having critical threats. In this way, authors were able to identify threats with 99.57% test accuracy.

F. Tagging & Log based Anomaly Prevention

The integration of smart contract technology with decentralized blockchain network has initiated a new era of decentralized on-chain agreement. Due to this initiation a large number of applications are now being developed which utilized the tremendous advantages by smart contracts. However, due to the wider acceptance of smart contract, certain problems have also started rising, as discussed in the earlier sections. Apart from the abovementioned methods. another way derived by researchers to evaluate and overcome the hazardous outcomes of adversarial smart contract and inputs by users is to use tagging and log systems. By using such systems one can try to mitigate and overcome the hazardous outcomes by adversaries, because via this one can detect anomalies before recording their outcomes to blockchain ledger.

One such work using log systems for smart contracts to identify prospective anomalies have been carried out by Shao *et al.* in [49]. The work first proposed a thorough analysis regarding usage of log analysis and storage from perspective of smart contract execution. Afterwards, the authors proposed LSC architecture, via which authors proposed a complete framework which can detect anomalies by users with the help of efficient smart contracts. The protocol works over learning and analysing logs on the basis of aging-resistant machine learning models. Afterwards, the learnt output results which can also be used to as models of anomaly detection are forwarded to executable smart contracts in order to identify presence of anomalies in the network. The developed smart contract also keeps on updating on the basis of new available information in order to ensure the novelty and security of smart contract against vulnerabilities. Another interesting work towards usage of tagging system to defends smart contracts of Ethereum have been carried out by Wang *et al.* in [109]. The major goal of the work is to prevent execution of malicious smart contracts alongside enhancing the overhead of detection. In this way, authors can ensure that all nodes, even with a small exeution power will be able to run the contract without worrying about the overhead. From the empirical analysis, authors showed that their proposed model can effectively safeguard against 11 specific errors and attacks such as logic error, superficial randomness, abnormal control flow, etc. Authors further evaluated their proposed to identify that whether the proposed model is practical or not, and from the experimental results, it can be concluded that the proposed ContractGuard model only causes an additional 28.27% runtime overhead and 36.14% deployment overhead.

G. Summary and Insights

Contract layer is relatively a more technical and considerably a new layer in blockchain, which got famous in the second era of blockchain named as blockchain 2.0 when Ethereum platform provided its users the functionality of developing DApps. A large number of contracts in a well-established blockchain network are pre-developed and do not contain any bugs, however, the malicious participants in the network always try to find out loopholes by any means and smart contracts are their recent targets because its hard for a non-technical person to identify bugs and honeypots in the smart contracts. The woks from the perspective of detection of anomalies in this specific layer is divided into multiple types ranging from identification of contracts restricting dependent transactions to highlighting faulty signals being transmitted via deployment of a smart contract. However, the most prominent works in the anomaly detection over this layer have been carried out from perspective of detection of Ponzi schemes and detection of critical security threats in the contracts, such as hacking, etc.

VIII. Challenges & Future Research Directions

In this section, we highlight five most prominent challenges that the field of anomaly detection in blockchain is currently facing alongside discussing their prospective future directions.

A. Privacy Preserving Anomaly Detection in Blockchain

1) Key Challenge: From our analysis of blockchain based anomaly detection models, we observed that none of the work discussed integration of privacy preservation in their works. Nevertheless, blockchain works over the phenomenon of a decentralized ledger and every node has a copy, therefore, it has got a lot of privacy issues that researchers are tackling [22]. Similarly, from the perspective of anomaly detection in blockchain, this issue doubles because one need to analyse even the deep details of each transaction happening over the network in order to identify any anomalous behaviour. However, its highly unwilling that blockchain participants share their complete data unless they have been provided with complete privacy guarantee.

2) Future Directions: Considering the nature of privacy requirement in blockchain based anomaly detection, it will not be wrong to say that there is a dire need to work over this issue. In order to overcome such issue, researchers can work over integration of modern privacy preservation strategies, such as differential privacy [114], zero knowledge proofs [115], etc.

with anomaly detection models of blockchain. In this way, researchers will be able to provide blockchain users with a safe and secure platform via which they will be able to prevent prospective anomalies without the risk of losing their private data. It is important to highlight that each of the privacy preservation model comes with a take away, e.g., while employing differential privacy, one have to deal with a trade-off between utility and privacy, same goes with other privacy preserving models. Therefore, the prospective models which show minimum effect over the utility and privacy of system will be a key contribution in this domain of privacy preserving anomaly detection in blockchain technology.

B. Integrating Federated Learning with Blockchain Anomaly Detection

1) Key Challenge: In the recent years, researchers worked over integration of various machine learning based anomaly detection for blockchain technology including CNN, SVM, etc. However, as per our observation, none of the work have integrated federated learning for blockchain anomaly detection. Nevertheless, blockchain is a decentralized model and the basic phenomenon of federated learning is also leading in a decentralized manner instead of a centralized server. Therefore, these two technologies perfectly fit with each other from the perspective of framework. Similarly, certain work have also identified the effectiveness of federated learning in anomaly detection of IoT and similar technologies [116], [117]. Now, the need is to develop such federated learning based anomaly detection models which comply with the nature of blockchain technology.

2) Future Directions: Integrating federated learning with blockchain anomaly detection has two fold advantages. One from the perspective of security and trust in the network, and other from the perspective of reducing of computational overhead and data storage. From the first point of view, federated learning already has decentralized nature, therefore, anomaly detection models do not have to collect huge amount of data in centralized servers, which will enhance and prevail a sense of trust in the network and blockchain users will be able to trust such network which is not gathering deep detail of their transactions. In this way, a more secure and trustworthy blockchain network can be established which will also be resilient to anomaly attacks.

Similarly, from the second view point, it is important to mention that detection overhead and data storing in a centralized database are the two critical issues which usual anomaly detection models face. However, if an anomaly detection model start working in a decentralized manner, then these major issues can be reduced to a negligible amount. E.g., by integrating federated learning based anomaly detection in malicious contract detection model, the network moderators will be able to identify the execution of an anomalous contract before it goes to the other network participants in a decentralized manner. Therefore, we believe that this integration of blockchain and federated learning based anomaly detection can provide tremendous benefits to blockchain community.

C. Integrating Novel Anomaly Detection Models for Blockchain Anomalies

1) Key Challenge: Machine/deep learning is a well-diversified field and a large number of models are being developed by researchers every other day. Some of these models can also be used to carry out anomaly detection and even some outperforms traditional anomaly detection models from perspective of accuracy [118]. From the perspective of anomaly detection in blockchain technology, researchers are continuously trying to use state-of-the-art models for anomaly detection by modifying them according to blockchain scenarios (See Fig. 1). However, this direction needs a fair amount of pace because majority of technical works carried out so far from perspective of blockchain anomaly detection used traditional anomaly detection models, which use usually based on classification, clustering based, or statistical & analytical modelling. However, the usage of other new models, such as generative architecture and reinforcement learning based models have not been carried out in the literature till now. Therefore, there is a dire need to explore such models from perspective of blockchain.

2) Future Direction: After careful analysis of anomaly detection taxonomy and current technical works in the field of anomaly detection of blockchain, it can be mentioned that there is a strong need for integration of modern anomaly detection models with blockchain. For instance, very minimal works suggested the use of generative adversarial networks (GANs) to detect anomaly in blockchain. However, this field of GANs for detecting anomalies is well established in multiple other domains, such as anomaly detection medical/clinical records [54], [119]. Similarly, the use of reinforcement learning is also not well-explored from blockchain anomaly detection perspective. Contrarily, this field of anomaly detection via reinforcement learning is also getting a lot of attention especially in the areas, where partially labelled datasets are available [120]. Therefore, considering recent developments in anomaly detection models, we believe that integrating state-of-the-art models with blockchain technology can produce fruitful outcomes and this direction needs extensive exploration.

D. Malicious Threats Identification in Modern Smart Contract Platforms

1) Key Challenge: Ethereum introduced the usage and functionality of smart contract in blockchain technology. However, now almost every new blockchain model has its own smart contracts for its functioning. One of the largest example after Ethereum is Hyperledger platform, which has its own diverse range of smart contracts especially focusing over enterprise functioning [121]. Similarly, all other blockchain platforms have their own personalized smart contract which facilitates their functions. However, if we analyse the integration of anomaly detection in blockchain smart contracts, majority of the work just focuses over anomaly detection in Ethereum smart contracts. No doubts, Ethereum was the first one to introduce smart contracts, therefore, there is a huge amount of literature over it. But now there is a need to work over identification of anomalies in smart contract from other technologies as well.

2) Future Directions: From our analysis regarding integration of anomaly detection blockchain for malicious smart contract identification, the majority of work we found only targets Ethereum based smart contracts. Considering the recent development of blockchain based technologies, we believe integration of anomaly detection with smart contracts of other technologies can be a key towards development of secure blockchains. For example, Hyperledger Fabric is one of the most viable alternative to Ethereum, however, a very minimal literature highlighting anomalous effects in Hyperledger Fabric can be found. Similarly, certain other blockchain platforms, such as Stellar, Waves, Nem, etc. have tremendous smart contract features, however, no work is available over identification of anomalous users and contracts for these technologies. It is important to highlight that while developing anomaly detection models for new blockchain models one key thing that needs to be kept in mind is their community standard. E.g., some communities might be willing to share significant amount of information towards the cause of anomaly detection. Contrarily, some communities might be stricter in sharing the data for development and execution of anomaly detection models. Therefore, while developing such models, the aspect of community requirements and standards needs to be taken into consideration.

E. Partially Observed Anomalies in Blockchain

1) Key Challenge: While dealing with new blockchain platforms and technologies, the data is limited and it is hard to categorise a behaviour as an anomalous behaviour due to lack of data. This aspect of unavailability of labelled data in supervised learning led to the formation of a phenomenon which is also known as partially observed anomalies [122]. Let us take the example of malicious smart contracts of a specific platform, in certain cases, apart from a huge number of unlabelled smart contracts, we only have a fewer number of labelled anomalous smart contracts with the help of basic learning or classification models. Thus, contrary to traditional supervised learning models, in which we provide a huge number of sample set for both positive and negative outcomes, we only have a small positive set. Therefore, supervised learning models cannot be applied directly to such scenarios. However, contrary to unsupervised learning model, we in addition have certain samples, which can aid in enhancement of prediction. Therefore, efficient models which overcome such issues are required for efficient blockchain anomaly identification.

2) Future Directions: From out point of view, this aspect of partially observed anomalies carried a critical stature, because blockchain is a new paradigm and a very large data regarding anomalies in the network is not as such available. Similarly, new blockchain models and applications are developing every day and each new application and technology has its own adversarial attacks and anomalies. Therefore, its very hard to collect a very large amount of data in a short time, but on the other hand a minimal amount of data can be collected as a result of statistical observation, which can be used for future detection. Therefore, a need to develop such models, which can provide fruitful results even in the presence of small labelled data are required. One such paradigm is positive and unlabelled (PU) learning, which is being discussed in certain other scenarios [123]. However, this specific learning is not yet integrated deeply in blockchain scenario. Therefore, we believe a research in the direction of learning from partially observed anomalies has a large scope in the blockchain network.

F. Development of Efficient Consensus Models for Anomaly Detection

1) Key Challenge: Since the advent of blockchain, a vast number of consensus mechanisms are being developed by researchers and experts to enhance the aspect of trust among peers and to overcome any prospective vulnerability [124]. As in consensus model, all nodes reach consensus over a unified transaction, similar to this, in case of a detected anomaly all nodes have to reach consensus that the nominated vulnerability is an anomaly. This becomes even more difficult when a vulnerability is not universally recognized or identified as an anomaly. E.g., for some nodes, an anomaly can just be a random behaviour but for other nodes, it could be a point of deep concern. Therefore, in such cases, reaching a consensus on a unified opinion becomes even more difficult.

2) Future Directions: It has been proven that the consensus carries an importance of backbone in blockchain technology, because due to this feature, blockchain nodes can reach and agree upon a unified claim. Similarly, in case of anomaly detection, this consensus needs to be finalized in a deterministic way so that none of the adversary take unusual advantage of it. However, till now, as per our knowledge, there is no specific consensus model which facilitates the early finality of consensus in case of an anomaly. Therefore, there is a strong need to develop such consensus models, which have specific features regarding detection of an anomaly. This can also be done by integrating some specific functionalities of anomaly detection in current running consensus models. E.g., a specific feature can be enabled if an anomaly is reported by a trustworthy mining node, or in case of an anomaly detection via some detection model. Similarly, certain penalty functions can be formulated and can be added in the existing consensus models, which reward or penalize the reporting of true or false anomalies. Nevertheless, this field of consensus modification in blockchain technology is pretty huge and it has a large gap especially for anomaly detection oriented consensus models which can be explored by researchers.

IX. Conclusion

Since the beginning of blockchain technology, it has attracted a tremendous amount of attention from both academia and industry due to its applicability in modern day applications. One of the prime reason behind this attention is the P2P architecture of blockchain, which makes it a secure, trustworthy, and truthful platform which is immutable and can be verified at the time of need. Even though, blockchain has such tremendous benefits, but it is also vulnerable to a huge number of attacks by adversaries, such as security, privacy, reliability, and performance attack, etc. Therefore, in order to keep these functionalities in the full running condition and to preserve any big catastrophe, it is important to identify any anomalous behaviour in the network within a limited time. In order to do so, anomaly detection techniques come into effect, which identify any anomalous behaviour in the network and report it for timely action. In this article, we work over providing a thorough survey of these anomaly detection models. Firstly, we provide a through discussion that how anomaly detection can help in enhancing the trust and security of blockchain network and its ongoing applications. Afterwards, we provide a detailed classification of blockchain anomalies alongside discussing evaluation matrices and key requirements for the development of anomaly detection models in the network. Afterwards, we provide a detailed in-depth analysis of existing anomaly detection works from perspective of four most prominent layers of blockchain technology. Finally, we provide a comprehensive discussion about certain challenges and future research directions which needs attention from the researchers working in the field of anomaly detection in blockchain technology.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Manubot, Tech. Rep., 2019.

[2] M. S. Ali, M. Vecchio, M. Pincheira, K. Dolui, F. Antonelli, and M. H. Rehmani, "Applications of blockchains in the internet of things: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1676–1717, 2018.

[3] M. Saad, J. Spaulding, L. Njilla, C. Kamhoua, S. Shetty, D. Nyang, and D. Mohaisen, "Exploring the attack surface of blockchain: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1977–2008, 2020.

[4] Y. Xiao, N. Zhang, W. Lou, and Y. T. Hou, "A survey of distributed consensus protocols for blockchain networks," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1432–1465, 2020.

[5] F. Dewanta and M. Mambo, "Bpt scheme: Establishing trusted vehicular fog computing service for rural area based on blockchain approach," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1752–1769, 2021.

[6] W. Chen, Z. Zheng, E. C.-H. Ngai, P. Zheng, and Y. Zhou, "Exploiting blockchain data to detect smart ponzi schemes on ethereum," *IEEE Access*, vol. 7, pp. 37 575–37 586, 2019.

[7] M. U. Hassan, M. H. Rehmani, and J. Chen, "Deal: Differentially private auction for blockchain-based microgrids energy trading," *IEEE Transactions on Services Computing*, vol. 13, no. 2, pp. 263–275, 2020.

[8] S. Zhang and J.-H. Lee, "Double-spending with a sybil attack in the bitcoin decentralized network," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5715–5722, 2019.

[9] M. Rahouti, K. Xiong, and N. Ghani, "Bitcoin concepts, threats, and machine-learning security solutions," *IEEE Access*, vol. 6, pp. 67 189–67 205, 2018.

[10] M. C. K. Khalilov and A. Levi, "A survey on anonymity and privacy in bitcoin-like digital cash systems," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2543–2585, 2018.

[11] W. Meng, E. W. Tischhauser, Q. Wang, Y. Wang, and J. Han, "When intrusion detection meets blockchain technology: A review," *IEEE Access*, vol. 6, pp. 10 179–10 188, 2018.

[12] T. Salman, M. Zolanvari, A. Erbad, R. Jain, and M. Samaka, "Security services using blockchains: A state of the art survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 858–880, 2019.

[13] M. U. Hassan, M. H. Rehmani, and J. Chen, "Privacy preservation in blockchain based iot systems: Integration issues, prospects, challenges, and future research directions," *Future Generation Computer Systems*, vol. 97, pp. 512–529, 2019.

[14] S. Omer, "Anomaly Detection in Blockchain," Master's thesis, Tampere University, 2019.

[15] M. Wu, K. Wang, X. Cai, S. Guo, M. Guo, and C. Rong, "A comprehensive survey of blockchain: From theory to iot applications and beyond," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8114–8154, 2019.

[16] M. Belotti, N. Božić, G. Pujolle, and S. Secci, "A vademecum on blockchain technologies: When, which, and how," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3796–3838, 2019.

[17] J. Liu and Z. Liu, "A survey on security verification of blockchain smart contracts," *IEEE Access*, vol. 7, pp. 77 894–77 904, 2019.

[18] J. Sengupta, S. Ruj, and S. D. Bit, "A comprehensive survey on attacks, security issues and blockchain solutions for iot and iiot," *Journal of Network and Computer Applications*, vol. 149, p. 102481, 2020.

[19] M. Saad, J. Spaulding, L. Njilla, C. Kamhoua, S. Shetty, D. H. Nyang, and D. Mohaisen, "Exploring the attack surface of blockchain: A comprehensive survey," *IEEE Communications Surveys & Tutorials, in Print*, 2020.

[20] I. Homoliak, S. Venugopalan, D. Reijsbergen, Q. Hum, R. Schumi, and P. Szalachowski, "The security reference architecture for blockchains: Towards a standardized model for studying vulnerabilities, threats, and defenses," *IEEE Communications Surveys Tutorials, in Print*, pp. 1–1, 2020.

[21] J. Li, C. Gu, F. Wei, and X. Chen, "A survey on blockchain anomaly detection using data mining techniques," in *Blockchain and Trustworthy Systems*, Z. Zheng, H.-N. Dai, M. Tang, and X. Chen, Eds. Singapore: Springer Singapore, 2020, pp. 491–504.

[22] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy in blockchain technology: A futuristic approach," *Journal of Parallel and Distributed Computing*, vol. 145, pp. 50–74, 2020.

[23] B. Bhushan, P. Sinha, K. M. Sagayam, and A. J, "Untangling blockchain technology: A survey on state of the art, security threats, privacy services, applications and future research directions," *Computers & Electrical Engineering*, p. 106897, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0045790620307497

[24] Y. Xiao, N. Zhang, W. Lou, and Y. T. Hou, "A survey of distributed consensus protocols for blockchain networks," *IEEE Communications Surveys Tutorials*, vol. 22, no. 2, pp. 1432–1465, 2020.

[25] V. Y. Kemmoe, W. Stone, J. Kim, D. Kim, and J. Son, "Recent advances in smart contracts: A technical overview and state of the art," *IEEE Access*, vol. 8, pp. 117 782–117 801, 2020.

[26] S. Bulusu, B. Kailkhura, B. Li, P. K. Varshney, and D. Song, "Anomalous example detection in deep learning: A survey," *IEEE Access*, vol. 8, pp. 132 330–132 347, 2020.

[27] C. O'Reilly, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Anomaly detection in wireless sensor networks in a non-stationary environment," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1413–1432, 2014.

[28] J. Xie, H. Tang, T. Huang, F. R. Yu, R. Xie, J. Liu, and Y. Liu, "A survey of blockchain technology applied to smart cities: Research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2794–2830, 2019.

[29] S. Rouhani and R. Deters, "Security, performance, and applications of smart contracts: A systematic survey," *IEEE Access*, vol. 7, pp. 50 759–50 779, 2019.

[30] R. Norvill, B. B. F. Pontiveros, R. State, I. Awan, and A. Cullen, "Automated labeling of unknown contracts in ethereum," in *26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2017, pp. 1–6.

[31] Y. Kim, D. Pak, and J. Lee, "Scanat: identification of bytecode-only smart contracts with multiple attribute tags," *IEEE Access*, vol. 7, pp. 98 669–98 683, 2019.

[32] C. F. Torres, M. Steichen *et al.*, "The art of the scam: Demystifying honeypots in ethereum smart contracts," in *28th {USENIX} Security Symposium ({USENIX} Security 19)*, 2019, pp. 1591–1607.

[33] R. Barone and D. Masciandaro, "Cryptocurrency or usury? crime and alternative money laundering techniques," *European Journal of Law and Economics*, vol. 47, no. 2, pp. 233–254, 2019.

[34] G. Ramezan and C. Leung, "Analysis of proof-of-work-based blockchains under an adaptive double-spend attack," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 7035–7045, 2020.

[35] Y. Liu, X. Liu, C. Tang, J. Wang, and L. Zhang, "Unlinkable coin mixing scheme for transaction privacy enhancement of bitcoin," *IEEE Access*, vol. 6, pp. 23 261–23 270, 2018.

[36] Y. Liu, R. Li, X. Liu, J. Wang, L. Zhang, C. Tang, and H. Kang, "An efficient method to enhance bitcoin wallet security," in *11th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, 2017, pp. 26–29.

[37] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," *IEEE communications surveys & tutorials*, vol. 16, no. 1, pp. 303–336, 2013.

[38] *Multiple Bitcoin Datasets* , *https* : //*data.world/datasets/bitcoin* [Last Accessed: 16 October 2021].

[39] *Ethereum Blockchain: Complete live historical Ethereum blockchain data (BigQuery)* , *https* : //*www.kaggle.com/bigquery/ethereum* − *blockchain* [Last Accessed: 16 October 2021].

[40] *Bitcoin Cash Blockchain: Live Bitcoin Cash historical blockchain data (BigQuery)*, *https* : //*www.kaggle.com/bigquery/crypto* − *bitcoin* − *cash* [Last Accessed: 16 October 2021].

[41] *Zcash Crypto Blockchain: Complete live Zcash historical Blockchain data (BigQuery)*, *https* : //*www.kaggle.com/bigquery/crypto* − *zcash* [Last Accessed: 16 October 2021].

[42] *Stocks Data by "Multi-class Bitcoin-enabled Service Identification-Based on Transaction History Summarization"*, *shorturl.at/drEIZ* [Last Accessed: 16 October 2021].

[43] *Rare Pepes Wallet*, *https* : //*data.world/fivethirtyeight/rare* − *pepes* [Last Accessed: 19 October 2021].

[44] *Smart Contract Attribute Dataset*, *https* : //*www.kaggle.com/xblock/smart* − *contract* − *attribute* − *dataset* [Last Accessed: 19 October 2021].

[45] *Smart Ponzi Scheme Labels*, *https* : //*www.kaggle.com/xblock/smart* − *ponzi* − *scheme* − *labels* [Last Accessed: 19 October 2021].

[46] *DApps Quality Characteristics Dataset*, *shorturl.at/oFMR8* [Last Accessed: 19 October 2021].

[47] D. J. Weller-Fahy, B. J. Borghetti, and A. A. Sodemann, "A survey of distance and similarity measures used within network intrusion anomaly detection," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 70–91, 2015.

[48] A. Kamisalic, R. Kramberger, and I. Fister, "Synergy of blockchain technology and data mining techniques for anomaly detection," *Applied Sciences*, vol. 11, no. 17, p. 7987, 2021.

[49] W. Shao, Z. Wang, X. Wang, K. Qiu, C. Jia, and C. Jiang, "Lsc: Online auto-update smart contracts for fortifying blockchain-based log systems," *Information Sciences*, vol. 512, pp. 506–517, 2020.

[50] A. Epishkina and S. Zapechnikov, "Discovering and clustering hidden time patterns in blockchain ledger," in *First International Early Research Career Enhancement School on Biologically Inspired Cognitive Architectures*. Springer, 2017, pp. 245–250.

[51] N. Kumar, A. Singh, A. Handa, and S. K. Shukla, "Detecting malicious accounts on the ethereum blockchain with supervised learning," in *International Symposium on Cyber Security Cryptography and Machine Learning*. Springer, 2020, pp. 94–109.

[52] S. Garcia, S. Ramirez-Gallego, J. Luengo, J. M. Bentez, and F. Herrera, "Big data preprocessing: methods and prospects," *Big Data Analytics*, vol. 1, no. 1, pp. 1–22, 2016.

[53] B. Al-Musawi, P. Branch, and G. Armitage, "BGP Anomaly Detection Techniques: A Survey," *IEEE Communications Surveys Tutorials*, vol. 19, no. 1, pp. 377–396, 2017.

[54] X. Han, X. Chen, and L.-P. Liu, "Gan ensemble for anomaly detection," *arXiv preprint arXiv:2012.07988*, 2020.

[55] G. Di Battista, V. Di Donato, M. Patrignani, M. Pizzonia, V. Roselli, and R. Tamassia, "Bitconeview: visualization of flows in the bitcoin transaction graph," in *IEEE Symposium on Visualization for Cyber Security (VizSec)*, 2015, pp. 1–8.

[56] W. Shao, H. Li, M. Chen, C. Jia, C. Liu, and Z. Wang, "Identifying bitcoin users using deep neural network," in *International Conference on Algorithms and Architectures for Parallel Processing*. Springer, 2018, pp. 178–192.

[57] L. Nan and D. Tao, "Bitcoin mixing detection using deep autoencoder," in *IEEE Third International Conference on Data Science in Cyberspace (DSC)*, 2018, pp. 280–287.

[58] J. Kim, M. Nakashima, W. Fan, S. Wuthier, X. Zhou, I. Kim, and S.-Y. Chang, "Anomaly detection based on traffic monitoring for secure blockchain networking," in *IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2021, pp. 1–9.

[59] K. Toyoda, T. Ohtsuki, and P. T. Mathiopoulos, "Identification of high yielding investment programs in bitcoin via transactions pattern analysis," in *GLOBECOM IEEE Global Communications Conference*, 2017, pp. 1–6.

[60] Y.-J. Lin, P.-W. Wu, C.-H. Hsu, I.-P. Tu, and S.-w. Liao, "An evaluation of bitcoin address classification based on transaction history summarization," in *IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2019, pp. 302–310.

[61] J. Song, J. Nang, and J. Jang, "Design of anomaly detection and visualization tool for iot blockchain," in *International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2018, pp. 1464–1465.

[62] M. Bartoletti, S. Lande, L. Pompianu, and A. Bracciali, "A general framework for blockchain analytics," in *Proceedings of the 1st Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers*, 2017, pp. 1–6.

[63] H. Tang, Y. Jiao, B. Huang, C. Lin, S. Goyal, and B. Wang, "Learning to classify blockchain peers according to their behavior sequences," *IEEE Access*, vol. 6, pp. 71 208–71 215, 2018.

[64] B. Huang, Z. Liu, J. Chen, A. Liu, Q. Liu, and Q. He, "Behavior pattern clustering in blockchain networks," *Multimedia Tools and Applications*, vol. 76, no. 19, pp. 20 099–20 110, 2017.

[65] F. Scicchitano, A. Liguori, M. Guarascio, E. Ritacco, and G. Manco, "Deep autoencoder ensembles for anomaly detection on blockchain," in *International Symposium on Methodologies for Intelligent Systems*. Springer, 2020, pp. 448–456.

[66] M. K. Awan and A. Cortesi, "Blockchain transaction analysis using dominant sets," in *IFIP International Conference on Computer Information Systems and Industrial Management*. Springer, 2017, pp. 229–239.

[67] "How Much of All Money Is in Bitcoin?" Https://www.investopedia.com/tech/how-much-worlds-money-bitcoin [Available Online: 06-08-2021].

[68] W. Liang, L. Xiao, K. Zhang, M. Tang, D. He, and K.-C. Li, "Data fusion approach for collaborative anomaly intrusion detection in blockchain-based systems," *IEEE Internet of Things Journal, in Print*, pp. 1–1, 2021.

[69] R. Yang, X. Chang, J. Mišić, and V. B. Mišić, "Assessing blockchain selfish mining in an imperfect network: Honest and selfish miner views," *Computers & Security*, vol. 97, p. 101956, 2020.

[70] Y. A. Hsain, N. Laaz, and S. Mbarki, "Ethereum's smart contracts construction and development using model driven engineering technologies: a review," *Procedia Computer Science*, vol. 184, pp. 785–790, 2021.

[71] T. Pham and S. Lee, "Anomaly detection in bitcoin network using unsupervised learning methods," *arXiv preprint arXiv:1611.03941*, 2016.

[72] F. Jobse, "Detecting suspicious behavior in the bitcoin network," Ph.D. dissertation, Tilburg University, 2017.

[73] D. D. F. Maesa, A. Marino, and L. Ricci, "Detecting artificial behaviours in the bitcoin users graph," *Online Social Networks and Media*, vol. 3, pp. 63–74, 2017.

[74] T.-H. Chang and D. Svetinovic, "Improving bitcoin ownership identification using transaction patterns analysis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018.

[75] M. Signorini, M. Pontecorvi, W. Kanoun, and R. Di Pietro, "Advise: Anomaly detection tool for blockchain systems," in *IEEE World Congress on Services (SERVICES)*, 2018, pp. 65–66.

[76] ——, "Bad: A blockchain anomaly detection solution," *IEEE Access*, vol. 8, pp. 173 481–173 490, 2020.

[77] Z. Il-Agure, B. Attallah, and Y.-K. Chang, "The semantics of anomalies in iot integrated blockchain network," in *Sixth HCT Information Technology Trends (ITT)*. IEEE, 2019, pp. 144–146.

[78] A. Bogner, "Seeing is understanding: anomaly detection in blockchains with visualized features," in *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the ACM International Symposium on Wearable Computers*, 2017, pp. 5–8.

[79] S. Dey, "Securing majority-attack in blockchain using machine learning and algorithmic game theory: A proof of work," in *10th computer science and electronic engineering (CEEC)*. IEEE, 2018, pp. 7–10.

[80] B. Podgorelec, M. Turkanovic, and S. Karakativc, "A machine learning-based method for automated blockchain transaction signing including personalized anomaly detection," *Sensors*, vol. 20, no. 1, p. 147, 2020.

[81] T. Pham and S. Lee, "Anomaly detection in the bitcoin system-a network perspective," *arXiv preprint arXiv:1611.03942*, 2016.

[82] S. Morishima and H. Matsutani, "Acceleration of anomaly detection in blockchain using in-gpu cache," in *IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)*, 2018, pp. 244–251.

[83] C. Chen, X. Chen, J. Yu, W. Wu, and D. Wu, "Impact of temporary fork on the evolution of mining pools in blockchain networks: An

evolutionary game analysis," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 1, pp. 400–418, 2021.

[84] H. Kang, X. Chang, R. Yang, J. Mišić, and V. B. Mišić, "Understanding selfish mining in imperfect bitcoin and ethereum networks with extended forks," *IEEE Transactions on Network and Service Management*, pp. 1–1, 2021.

[85] S. G. Motlagh, J. Mišić, and V. B. Mišić, "The impact of selfish mining on bitcoin network performance," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 1, pp. 724–735, 2021.

[86] J. Chen, X. Xia, D. Lo, J. Grundy, X. Luo, and T. Chen, "Defining smart contract defects on ethereum," *IEEE Transactions on Software Engineering*, pp. 1–1, 2020.

[87] S. Bistarelli, G. Mazzante, M. Micheletti, L. Mostarda, D. Sestili, and F. Tiezzi, "Ethereum smart contracts: Analysis and statistics of their source code and opcodes," *Internet of Things*, vol. 11, p. 100198, 2020.

[88] J. Hirshman, Y. Huang, and S. Macke, "Unsupervised approaches to detecting anomalous behavior in the bitcoin transaction network," *3rd ed. Technical report, Stanford University*, 2013.

[89] J. Lorenz, M. I. Silva, D. Aparicio, J. T. Ascensao, and P. Bizarro, "Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity," *arXiv preprint arXiv:2005.14635*, 2020.

[90] P. M. Monamo, V. Marivate, and B. Twala, "A multifaceted approach to bitcoin fraud detection: Global and local outliers," in *15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2016, pp. 188–194.

[91] S. Ranshous, C. A. Joslyn, S. Kreyling, K. Nowak, N. F. Samatova, C. L. West, and S. Winters, "Exchange pattern mining in the bitcoin transaction directed hypergraph," in *International Conference on Financial Cryptography and Data Security*. Springer, 2017, pp. 248–263.

[92] D. Zambre and A. Shah, "Analysis of bitcoin network dataset for fraud," *Stanford Report*, vol. 27, 2013.

[93] S. Morishima, "Scalable anomaly detection method for blockchain transactions using gpu," in *20th IEEE International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, 2019, pp. 160–165.

[94] H. Baek, J. Oh, C. Y. Kim, and K. Lee, "A model for detecting cryptocurrency transactions with discernible purpose," in *Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2019, pp. 713–717.

[95] S. Farrugia, J. Ellul, and G. Azzopardi, "Detection of illicit accounts over the ethereum blockchain," *Expert Systems with Applications*, vol. 150, p. 113318, 2020.

[96] S. Sayadi, S. B. Rejeb, and Z. Choukair, "Anomaly detection model over blockchain electronic transactions," in *15th IEEE International Wireless Communications & Mobile Computing Conference (IWCMC)*, 2019, pp. 895–900.

[97] S. Y. Yang and J. Kim, "Bitcoin market return and volatility forecasting using transaction network flow properties," in *IEEE Symposium Series on Computational Intelligence*, 2015, pp. 1778–1785.

[98] W. Chen, J. Wu, Z. Zheng, C. Chen, and Y. Zhou, "Market manipulation of bitcoin: evidence from mining the mt. gox transaction network," in *IEEE INFOCOM -IEEE Conference on Computer Communications*, 2019, pp. 964–972.

[99] M. Xie, H. Li, and Y. Zhao, "Blockchain financial investment based on deep learning network algorithm," *Journal of Computational and Applied Mathematics*, vol. 372, p. 112723, 2020.

[100] J. Liu and A. Serletis, "Volatility in the cryptocurrency market," *Open Economies Review*, vol. 30, no. 4, pp. 779–811, 2019.

[101] A. F. Bariviera and I. Merediz-Solà, "Where do we stand in cryptocurrencies economic research? a survey based on hybrid analysis," *Journal of Economic Surveys*, vol. 35, no. 2, pp. 377–407, 2021.

[102] C. Natoli and V. Gramoli, "The blockchain anomaly," in *IEEE 15th International Symposium on Network Computing and Applications (NCA)*, 2016, pp. 310–317.

[103] T. Ide, "Collaborative anomaly detection on blockchain from noisy sensor data," in *IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2018, pp. 120–127.

[104] B. Oh, T. J. Jun, W. Yoon, Y. Lee, S. Kim, and D. Kim, "Enhancing trust of supply chain using blockchain platform with robust data model and verification mechanisms," in *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 3504–3511.

[105] T. H.-D. Huang, "Hunting the ethereum smart contract: Color-inspired inspection of potential attacks," *arXiv preprint arXiv:1807.01868*, 2018.

[106] E. Jung, M. Le Tilly, A. Gehani, and Y. Ge, "Data mining-based ethereum fraud detection," in *IEEE International Conference on Blockchain (Blockchain)*, 2019, pp. 266–273.

[107] J. Chen, X. Xia, D. Lo, J. Grundy, X. Luo, and T. Chen, "Defectchecker: Automated smart contract defect detection by analyzing evm bytecode," *IEEE Transactions on Software Engineering, in Print*, pp. 1–1, 2021.

[108] W. J.-W. Tann, X. J. Han, S. S. Gupta, and Y.-S. Ong, "Towards safer smart contracts: A sequence learning approach to detecting security threats," *arXiv preprint arXiv:1811.06632*, 2018.

[109] X. Wang, J. He, Z. Xie, G. Zhao, and S.-C. Cheung, "Contractguard: Defend ethereum smart contracts with embedded intrusion detection," *IEEE Transactions on Services Computing*, vol. 13, no. 2, pp. 314–328, 2019.

[110] A. Ailijiang, A. Charapko, M. Demirbas, and T. Kosar, "Wpaxos: Wide area network flexible consensus," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 1, pp. 211–223, 2020.

[111] A. Furfaro, L. Argento, D. Saccá, F. Angiulli, and F. Fassetti, "An infrastructure for service accountability based on digital identity and blockchain 3.0," in *IEEE INFOCOM Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019, pp. 632–637.

[112] J. Xu and B. Livshits, "The anatomy of a cryptocurrency pump-and-dump scheme," in *28th {USENIX} Security Symposium ({USENIX} Security 19)*, 2019, pp. 1609–1625.

[113] T. Frankel, *The Ponzi scheme puzzle: A history and analysis of con artists and victims*. Oxford University Press, 2012.

[114] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for cyber physical systems: A survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 1, pp. 746–789, 2020.

[115] X. Sun, F. R. Yu, P. Zhang, Z. Sun, W. Xie, and X. Peng, "A survey on zero-knowledge proof in blockchain," *IEEE Network*, vol. 35, no. 4, pp. 198–205, 2021.

[116] Y. Liu, N. Kumar, Z. Xiong, W. Y. B. Lim, J. Kang, and D. Niyato, "Communication-efficient federated learning for anomaly detection in industrial internet of things," in *GLOBECOM - IEEE Global Communications Conference*, 2020, pp. 1–6.

[117] L. Cui, Y. Qu, G. Xie, D. Zeng, R. Li, S. Shen, and S. Yu, "Security and privacy-enhanced federated learning for anomaly detection in iot infrastructures," *IEEE Transactions on Industrial Informatics, in Print*, pp. 1–1, 2021.

[118] J. Lansky, S. Ali, M. Mohammadi, M. K. Majeed, S. H. T. Karim, S. Rashidi, M. Hosseinzadeh, and A. M. Rahmani, "Deep learning-based intrusion detection systems: A systematic review," *IEEE Access*, vol. 9, pp. 101 574–101 599, 2021.

[119] F. Di Mattia, P. Galeone, M. De Simoni, and E. Ghelfi, "A survey on gans for anomaly detection," *arXiv preprint arXiv:1906.11632*, 2019.

[120] G. Pang, A. van den Hengel, C. Shen, and L. Cao, "Toward deep supervised anomaly detection: Reinforcement learning from partially labeled anomaly data," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 1298–1308.

[121] D. Li, W. E. Wong, and J. Guo, "A survey on blockchain for enterprise using hyperledger fabric and composer," in *6th International Conference on Dependable Systems and Their Applications (DSA)*, 2020, pp. 71–80.

[122] Y.-L. Zhang, L. Li, J. Zhou, X. Li, and Z.-H. Zhou, "Anomaly detection with partially observed anomalies," in *Companion Proceedings of the The Web Conference*, 2018, pp. 639–646.

[123] R. Kiryo, G. Niu, M. C. d. Plessis, and M. Sugiyama, "Positive-unlabeled learning with non-negative risk estimator," *arXiv preprint arXiv:1703.00593*, 2017.

[124] G. Xu, Y. Liu, and P. W. Khan, "Improvement of the dpos consensus mechanism in blockchain based on vague sets," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4252–4259, 2020.