Grenze International Journal of Engineering and Technology, June Issue

Convergence of Machine Learning with Blockchain Technology

Yojana[1] and Yogesh Chaba[2]
[1-2]Guru Jambheshwar University of Science and Technology, Hisar, India
Email: yojana.011@gmail.com, yogeshchaba@yahoo.com

***Abstract*—Recently, ML and BCT convergence has become the most disruptive and trending technology. ML uses a large database, which emphasizes the synchronization and security of data. To handle the former issues, use the decentralization structure of BCT for the ML data, which is also immutable and uses distributed ledger, which makes the framework more secure and data reliable. ML offers intelligent machines which work similarly to hu- mans. In this paper, both technologies summarized , how they work, and their types. Also, discuss some research objectives of how ML with BCT integrated and how this technology use for analyzing traffic.**

***Index Terms*— Machine Learning (ML), Block-chain Technology(BCT), Traffic Analysis.**

I. INTRODUCTION

Let us first get familiar with the traditional approach used, where the programmer evaluates the model output and creates a new rule to make models more accurate. Traditional programming is time-consuming, and so machine learning came into existence, which is a subset of artificial intelligence. Almost a decade ago Solomonoff 1957 first time wrote about machine learning. Machine learning (ML) rises in the early 2000s and is highly motivated by human learning behavior, where input and output are given and the machine figures out how to solve the problem. Machine learning is the term used for cognition of simple algorithms and as a statistical method used by a system to complete a task without explicitly programming. ML's main strength is in the formation of different computer algorithms that give access to data from different machines[1]. In this way, machines take their own decisions, and based on data, they make predictions. There are a distinct number of areas where machine learning is used as a solution to their problems. ML examines the features of a newly presented object (problem) and assigns it to one of the pre-defined classes (algorithms). While there are several applications of ML-like classifying loans as Low, Medium, and High Risk, assigning products into categories and sub categories, speech recognition, image recognition, and traffic prediction [2]. These are data-driven applications that use a vast amount of data. To synchronize and secure the data, a third party is required, which is not efficient to use. For that purpose, technology used for synchronization and security, i.e. Block-chain Technology (BCT). In 2008, a group of un- known people which is also known as Nakamoto, first time published a study about bitcoin[3]. Until mid-2010, Nakamo- to continued to conspire with another software developer for the further work of bitcoin, and by using the bitcoin, they implemented a decentralized block-chain framework for the first time. Blockchain technology is a database that stores records, which are called blocks. In a network, every block connect with another block. This format of the framework is preferred to be a ‘digital ledger [4]. The

Grenze ID: 01.GIJET.8.2.12_1

encrypted digital code of the owner, which validates the transaction and secures data from alteration, certifies every transaction in the ledger. Hence, the information in the digital ledger is highly protected. Immutability, decentralization, and transparency these factors are the backbone of blockchain technology [5]. These three features opened its door to the utmost variety of applications. Blockchain Technology categorized such applications into no of groups, i.e., healthcare, services related to financial, business and industry, and other new applications. Our concern is related to the security of data, this technology provides security at three levels. First, decentralization excludes the possibility of data alteration [6]. Second, cryptographic technology is used for the security of data and to confirm that data can't be changed without approval [7]. In the last, the consensus mechanism provides security for the whole framework [8].This analysis observed that the joint consideration of Machine Learning and Block-chain Technology brings a significant impact on the case study of traffic analysis. The systematic review of both the technology is the primary driver for conducting this research. In particular, try to attempt to address this by answering the following three questions:

a) How does ML BCT technology develop over time?
b) How do ML and BCT work, discussed their framework, their types, and their characteristics?
c) How does the convergence of ML BCT framework affect and find out the benefit of both the technology when integrated?

Our work shares added to a thorough understanding of the ML BCT features framework and provides a scenario of the current state of both technologies. In this paper try to highlight the interest growing based on the content analysis approach and classify three key research aims: (i) Identify and classify the ML BCT technology and their based In Section II presented an overview in which the background of Machine Learning and Block Chain Technology is discussed, their architecture is explored how they work, their features, and types. After introducing this technology in Section III, present the descriptive analysis of the integration of this technology for the use case: Traffic Analysis. The last, the conclusion, is presented in Section 4.

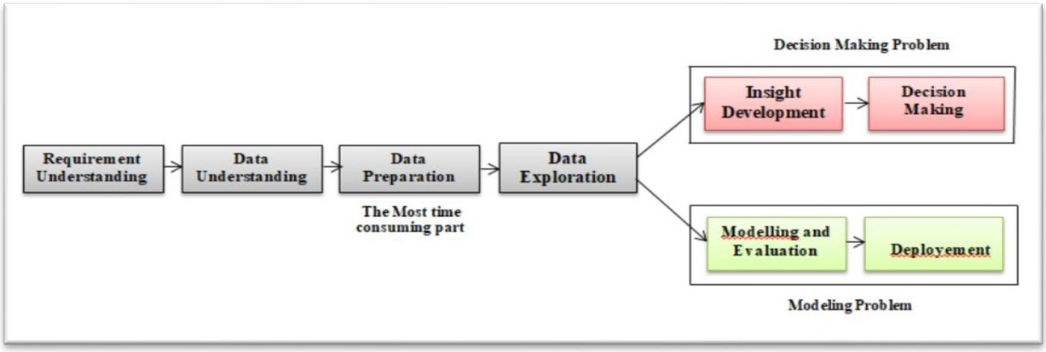

Fig 1

II. Overview

A. Machine Learning

In the principle, ML consider as an application that learns from the experience. Without explicitly programming, ML can teach a computer [9]. ML enclosed its work from a distinct set of areas, including information theory, artificial intelligence, control theory, computational complexity, probability, and statistics [10]. Nowadays, researchers discuss the framework of classification where the major goal is to focus on the use of the particular algorithm. To understand more deeply the workflow of the framework, discuss a detailed structure of ML. The workflow of the whole framework shows in Fig 1. In the first step, User have to understand the data according to their needs and then analyze the sources from which raw data is collect, better the quality of data better it will be for the modeling. After finalizing the data source, gathered all the data according to the use of users. The next step is to prepare the data in which divide the data into two parts which are required or not, then the required data is divided into two phases, i.e., the training set and the testing set. With suitable algorithms and techniques on the training set, build a model where the training set is used as an input, and for the testing phase, take the remaining data. After dividing the data, explored the data in more detail, remove the data which is not in use, fill the null value if required.

Types of ML approaches:

- *Supervised learning*: It's a type of classification method. To generate a new prediction of outcomes,

train the labeled data. In this type of learning, learning continues as far as the algorithm reaches an acceptable point. Based on training data, the algorithm continuously concludes outputs, which are continually correct [11] [12].

- *Unsupervised Learning*: Unsupervised Learning is also known as the clustering method, where labeling of the dataset is not required; only data as input is required. The goal of this learning is to learn about data by the de- sign of infrastructure or how data is distributed [13].

ML algorithms grow over traditional algorithms since ML algorithms axiomatically learn from the data. Data that is given as an input to algorithms of ML is divided into two phases, one is a training set and the other is a testing set which already discussed. If the size of the training data set increases, the truthfulness of ML algorithms increases [14]. There are two major approaches to ML, which are Supervised Learning and Unsupervised Learning, as already mentioned previously. According to some researchers, they present that the truthfulness (accuracy) of Supervised Learning is much superior to Unsupervised Learning, as shown in Table1. The outcome showed that supervised technique is less powerful than unsupervised techniques. However, both techniques work better because they present little distinction in the truthfulness of attack, which is visible and invisible [15].

TABLE I.

Types of ML	*ML Algorithms*	*Accuracy*
Supervised	*Naïve Baise, SVM, Random Forest*	*91% - 95%*
Unsupervised	*K-Mean*	*91% - 98%*

Traditional approaches are applied to earlier works and their main goal is to find out useful information from huge data packets, which can be useful for applications like analysis of security, and profiling of users [19][16]. Our goal in using ML is to the analysis of the traffic that extract from social media and identify attacks [17]. In addition, technologies of ML require no human intervention for change in the traffic patterns; ML does it by detecting the controller of the network [12] and accomplishing no. of goals as present in Table 2. Although over from many years of research presents that Machine Learning is very helpful for traffic analysis, it revolves around the significance of typical training, through which devices that find out deviation should be continuously trained because of the evaluation of the types and characteristics of attacks. Certain criteria are to be studied to regulate how methods are effective for detecting the attack type [17]. In Table 2 Machine Learning techniques are summarized which is used for traffic analysis.

B. Block chain Technology

The block chain is a secured protocol enabling peer-to-peer exchange in the network [3]. Block chain is an unchangeable set of records that are cryptographically associated with each other. Peers have agreed to a distributed ledger of transactions, stored as immutable blocks that are connected to one upon on a decentralized network secured by cryptography [25]. The distributed ledger is not copied, it allows for shared data or information with every node of the network. Why did Block chain come into existence? The traditional approach Client-Server network uses they centralize its architecture, which means the server has all the information in one place which can easily be updated according to changes of data. As already know, most applications are using huge information, and no business runs using this information. Block chain stores data in a distributed ledger that can be accessed, and updated, ensuring all the records are in the correct order in the network which makes this technology ideal for transferring information that is unaltered and transparent. If all the information is summarize the as whole thing, a framed BCT using these elements, i.e., immutable, decentralized, distributed ledger which makes BCT powerful.

Key Elements of Block chain:

- *Distributed Ledger Technology:* All the members of the network can access the distributed. Ledger. Its record of transactions is immutable. They shared the ledger among every member of the network, all information is recorded only once, and end duplicate data from the ledger. There are three types of permission given to the ledger, according to which the whole process acts those are Private, Consortium, and Public permission.
- *Immutable Records:* Once the information is stored in the ledger, it can't alter or shared by every member of the block even if it contains an error when there is any position in the ledger then with the permission of all members ledger is updated accordingly.

TABLE II.

Authors	*Technique used*	*Objectives*
Paxson [20]	Machine learning technique	ML methods have been applied for detection of spam more effectively than intrusion detection because the detection of inconsistency is best for detecting different forms of attacks.
Zheng [23]	Machine learning SVM is used.	Classify the energy which is used as data flow.
Suthaharan [21]	Machine learning use techniques that are of supervised learning.	By learning the properties of the network, intrusion traffic is classified.
Laskov [15].	Machine learning (SVM), (kNN) γ-algorithm, k-means	For detection of malicious activities, compared both types of learning i.e. supervised and unsupervised.
Mukkamala [22]	SVM technique used in Machine Learning	Identify patterns or features that define user behaviors to build classifiers for recognizing anomalies
Vinoth [24]	Naïve Bayes and Decision tree algorithms of ML are used.	Traffic that is used in the network is classified.
M.Sofka [19]	Using the supervised learning technique of Machine Learning.	Identify security threats that are already known and previously unseen.

- *Decentralization:* This is one of the most crucial elements of the BCT. In a decentralized framework, it doesn't need a central node. For storing data, the recording of data ledger concept is in use, the ledger which is shared with every block or node in the framework does all the work [26].

There are three types of permission given by the authority for the block-chain framework, those are: Public Block-chain, Consortium Block-chain and Private Block chain. The selection of the type of block chain is based on the requirement and place where the block chain is applied. If anyone can verify any transaction between blocks of BCT and data is accessed by everyone who wants to read it, then prefer the Public Block chain. When main lead want to give some permission for some blocks to read and for some blocks permission is not given, there is no access for it or set the priorities according to which user can access in that case use the Consortium Block chain. In a Private Block chain, no one is allowed to access the data, and anyone present in the network can not verify any transaction. There are problems when BCT is not in use is difficult to monitor asset ownership and transfers in a trusted business network [27]. BCT gave a solution by giving different permission to different parts which already discussed.

Structure of Block chain

In a particular order number of blocks with transactions is connected in some order then it's considered a Blockchain. Fundamental Concepts of a data structure are used for the structure of Block chain: I) Pointer ii) Linked Lists

Pointers: The pointer carries the data of another node or variable present in the network or points out to the address of the node.

Linked List: Every node or block saves its respective information on the network. With the help of a pointer, they link it to the other node.

In the first block, it doesn't involve the pointer. They designed the structure of the BCT framework like a linked list data structure so it reaches the last block or node of the network which has only a pointer. In the block chain framework, they connected blocks of records in a particular sequence, which is shown in Fig. 2.

Block chain Architecture Components and its working:

Components of Block chain Architecture:

- *Node*: users within the architecture of Block chain.
- *Transaction*: when data, record, or any information is transferred from one block to another.
- *Block*: It's like a data structure that comprises every record of transaction executed between nodes.

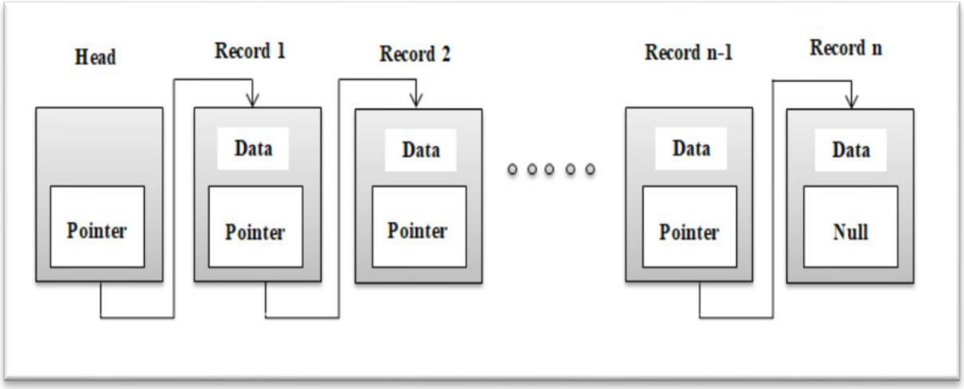

Fig 2

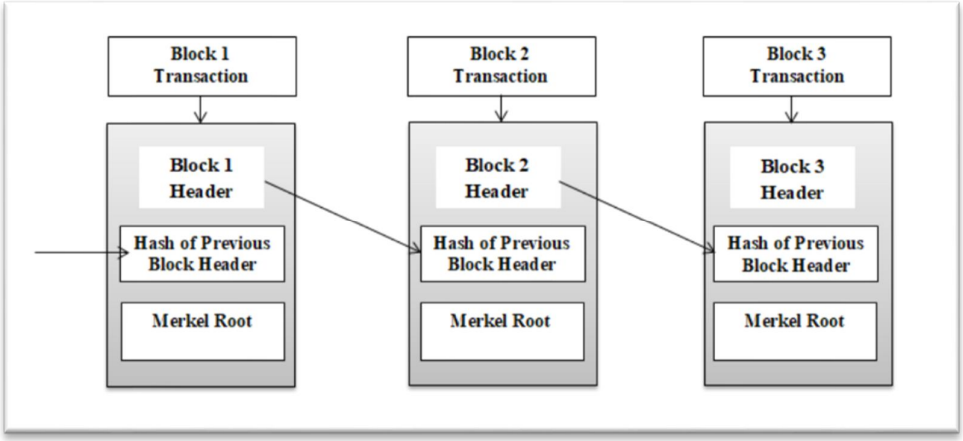

Fig 3

- *Chain*: Arrangement of blocks in a particular order.
- *Miners*: Miners is the node that verifies the new node whenever a new node wants to add to the framework.
- *Consensus Protocol*: Consensus protocol is a set of rules which operations of block chain should follow.

In the block chain architecture, whenever a new transaction is executed, it creates a new block. The digitally signed proof assures of its authority for every new transaction. Before the addition of this block into the network, it verifies by every node of the architecture. It presented the internal structure of the block in Fig 3, in which all block has its header, hash value of the previous block, and Merkel Root. Merkel Root is a type of data structure, like a Binary Hash Tree. This data structure has several data blocks that contain hashes that conclude all the transactions in a block. It secures data more efficiently.

How Block chain Works:

Block is a package of transactions, which comprise the previous block hash. All the blocks in the network are linked us- ing their hash value, and data can't be changed in one block without breaking the chain [5]. Every block shares the same ledger. Now here gave a detailed overview of how Block chain works. The first step is to request a transaction, and they create a block that represents the transaction. The next step is broadcasting the blocks in the network. The network ver i- fies transaction validation. After that node receives a reward for the proof of work, the block is added to the existing block chain, and then the transaction is completed. An overview of block chain, its framework objectives, survey of Block chain is presented in Table 3.

III. Convergence of Machine Learning with Blockchain

The joint consideration of Machine learning with Block chain technology helps to improve the problem which is faced by the applications of machine learning. By using the BCT's different properties, ML security and synchronization problem- solve to some extent. Machine Learning is also used to increase the time taken to

reach consent by creating routes for better data sharing synchronization. Further, it creates an event to architect an improved model by using the favorable benefits of BCT. ML algorithms use information or data as input for their processing and to reach outcomes. ML algorithms perform better when data is possessed from the data warehouse or a repository that is secure, credible, and valid. Block chain used the distributed ledger which makes BCT powerful.

TABLE III.

Authors	*Objective*	*Findings*
Meng et al. [28]	In Intrusion Detection use of the block-chain present	A different application of Block-chain is discussed and their future scope.
Alexopoulos et al. [29]	Integration of the Intrusion system with the block chain increases the trust and accountability in the system.	Design considerations thoroughly Explained
W.Kanoun et al. [32]	Irregularities of the detection system using Block chain is discussed.	use of bandwidth is extremely low which is about 0.246%
S.Chawathe et al. [33]	For monitoring self-organizing maps in Real-time using block chain.	Present a linear relationship for running time for the number of attributes, different instances, and iterations.
Sarda et al. [34]	To prevent work-history related frauds using block chain	Validate the history of work and excess use of block chain
Sgantzos et al. [30]	Using a block-chain genetic algorithm Implemented and their result presented.	Discussed architecture and future Applications and their scope.

Hence, the information in the digital ledger is highly protected. The convergence of ML and block chain can create an The encrypted digital code of the owner, which validates the transaction and secures data from alteration, certifies every transaction in the ledger immutable, secure, decentralized, and immutable system for eminently perceived data.This concept outcome in significant advancements for data security and information in many areas, which include banking, financial and legal data. Some properties of BCT for ML emerge a significant effect:

- The intelligence of Decentralization Framework: Any decision of the process involves multiple users, and every user performs their own task which is trained under the training which is common for every user. In the same way, it combined different ML users in groups, which gave the network security and help in solving the scheduling issues.
- Security of Data Enhanced: The ML algorithm works on a vast amount of data and information for the security of data. BCT provides security to the data. Block-chain store database in the blocks, as has been already studied. Every block is signed digitally which means every block has its private keys which makes ML more secure and gave valid outcomes.
- High Efficiency: BCT work on decentralized architecture which uses the distributed ledger to save all the in- formation of the particular block which gave immense efficiency to the network. With ML as already know, it uses a vast amount of data for processing when ML is integrated with BCT, then the concept of decentralization applies to the ML framework, which helps to synchronize the data and gave efficient outcomes.

TABLE IV.

Machine Learning	Block chain Technology	Convergence of ML with BCT Benefits
Resilient (Volatile)	Unchangeable	Discernment of Decentralized Structure
Dynamic	Deterministic	Resolve Credible Issues
Centralized	Decentralized	Security of data Increase
The main concern in data, Information, and decision	Main Concern on Security	Efficiency high

The integration of machine learning with block chain enhances block chain's underlying architecture and boosts Machine Learning potential. Block chain & ML perfectly complement each other and very much are the two pillars on which future innovations are to be built. The main features of block chain may bring a lot of advancement for ML techniques. Because of its decentralized element, block chain can conceivably kill the inconveniences of bringing it together. In Table 4, present the benefits which get after integration of technologies.

IV. CONCLUSION

In this paper, a detailed overview of Machine Learning and Block chain Technology is given. Also discussed how Block chains Technology and its main features: Immutable, Decentralization framework, and using distributed ledger are integrated with Machine Learning, enhance and solve issues of Machine Learning. Discussed A case study of traffic analysis is discussed for which these technologies are used. The framework of Machine Learning and its types also discussed and summarized techniques of Machine Learning used for traffic analysis. Overview of both the technology shows that they gave the best outcomes when they are integrated.

REFERENCES

[1] T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, Inc., 1 ed., 1997.
[2] P. Louridas and C. Ebert, "Machine learning," IEEE Software, vol. 33, pp. 110–115, Sep. 2016.
[3] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*, 21260.
[4] G. Wood, ``Ethereum: A secure decentralized generalised transaction ledger," *Ethereum Project Yellow Paper*, vol. 151, pp. 1_32, Apr.2014
[5] G. B. Mermer, E. Zeydan, and S. S. Arslan, "An overview of blockchain technologies: Principles, opportunities and challenges," in 201826th Signal Processing and Communications Applications Conference (SIU), pp. 1–4, May 2018.
[6] Schmidt, C. G., & Wagner, S. M. (2019). Blockchain and supply chain relations: A transaction cost theory per- spective. *Journal of Purchasing and Supply Management*, *25*(4), 100552.
[7] Lu, Y. (2018). Blockchain and the related issues: A review of current research topics. *Journal of Management Analytics*, *5*(4), 231-255.
[8] Singh, M., Singh, A., & Kim, S. (2018, February). Blockchain: A game-changer for securing IoT data. In *2018 IEEE4th World Forum on Internet of Things (WF-IoT)* (pp. 51-55). IEEE.
[9] Zhou, W., et al. *Internet traffic classification using feed-forward neural network*. in *2011 International Confer- ence on Computational Problem-Solving (ICCP)*. 2011. IEEE.
[10] Buczak, A.L. and E. Guven, *A survey of data mining and machine learning methods for cyber security intru- sion detection*. IEEE Communications Surveys & Tutorials, 2015. **18**(2): p. 1153-1176.
[11] Shafiq, M., et al. *Network traffic classification techniques and comparative analysis using machine learning algorithms*. in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*. 2016. IEEE.
[12] Zamani, M., et al. *A DDoS-aware IDS model based on danger theory and mobile agents*. in *2009 International Conference on Computational Intelligence and Security*. 2009. IEEE.
[13] Dainotti, A., A. Pescape, and K.C. Claffy, *Issues and future directions in traffic classification*. IEEE network, 2012. **26**(1): p. 35-40.
[14] Scarfone, K. and P. Mell, *Guide to intrusion detection and prevention systems (idps)*. 2012, National Institute of Standards and Technology.
[15] Laskov, P., et al. *Learning intrusion detection: supervised or unsupervised?* in *International Conference on Image Analysis and Processing*. 2005. Springer
[16] Parsaei, M.R., et al., *Network traffic classification using machine learning techniques over software defined networks*. International Journal of Advanced Computer Science and Applications, 2017. **8**(7): p. 220-225.
[17] Scarfone, K. and P. Mell, *Guide to intrusion detection and prevention systems (idps)*. 2012, National Institute of Standards and Technology *fication*. Procedia Computer Science, 2015. **60**: p. 784-791.
[18] Namdev, N., S. Agrawal, and S. Silkari, *Recent advancement in machine learning based internet traffic classi-*
[19] Bartos, K., M. Sofka, and V. Franc. *Optimized invariant representation of network traffic for detecting unseen malware variants*. in *25th {USENIX} Security Symposium ({USENIX} Security 16)*. 2016.
[20] Sommer, R. and V. Paxson. *Outside the closed world: On using machine learning for network intrusion detec- tion*. in *2010 IEEE symposium on security and privacy*. 2010. IEEE.
[21] Suthaharan, S., *Big data classification: Problems and challenges in network intrusion prediction with machine learning*. ACM SIGMETRICS Performance Evaluation Review, 2014. **41**(4): p. 70-73.
[22] Mukkamala, S., G. Janoski, and A. Sung. *Intrusion detection: support vector machines and neural networks*. in*proceedings of the IEEE International Joint Conference on Neural Networks (ANNIE), St. Louis, MO*. 2002.
[23] Zheng, N., et al. *You are how you touch: User verification on smartphones via tapping behaviors*. in *2014 IEEE 22nd International Conference on Network Protocols*. 2014. IEEE.
[24] Jamuna, A. and V. Ewards, *Survey of Traffic Classification using Machine Learning*. International journal of advanced

research in computer science, 2013. **4**(4).
[25] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, con- sensus, and future trends," in 2017 IEEE International Congress on Big Data (BigData Congress),pp. 557–564,June 2017
[26] 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, pp. 1–1, 2019.
[27] M. Vukoli´c, "The quest for scalable blockchain fabric: Proof-of-work vs. bft replication," in Open Problems in Network Security (J. Camenisch and D. Kesdo˘gan, eds.), (Cham), pp. 112–125, Springer International Publish- ing, 2016.
[28] 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. 10179–10188, 2018.
[29] N. Alexopoulos, E. Vasilomanolakis, N. R. Ivánkó, and M. Mühlhäuser, "Towards blockchain-based collabora- tive intrusion detection systems," in Critical Information Infrastructures Security (G. D'Agostino and A. Scala, eds.), (Cham), pp. 107–118, Springer International Publishing, 2018.
[30] K. Sgantzos and I. Grigg, "Artificial intelligence implementations on the blockchain. use cases and future ap- plications," Future Internet, vol. 11, no. 8, p. 170, 2019.
[31] P. Sarda, M. J. M. Chowdhury, A. Colman, M. A. Kabir, and J. Han, "Blockchain for fraud prevention: A work-history fraud prevention system," in 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (Trust- Com/BigDataSE), pp. 1858–1863, Aug 2018.
[32] M. Signorini, M. Pontecorvi, W. Kanoun, and R. D. Pietro, "BAD: blockchain anomaly detection," CoRR, vol. abs/1807.03833, 2018.
[33] S. Chawathe, "Monitoring blockchains with self-organizing maps," in 2018 17th IEEE International Confer- ence On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), pp. 1870–1875, Aug.