# Analytical Approach for Future Blockchain Forensic Investigation of Bitcoin Transaction Network

**Sumaiya. Samreen[1], B.Sai Sruthi[2], A.Manisha Kumari[2], B.Vaishnavi[2], CH. Chandana Priya[2]**

[1] Assistant Professor, Department of Information Technology, Mallareddy Engineering College for Women, (UGC-Autonomous), Hyderabad, India, sumaiyasamreen_it@mrecw.in.

[2] Student, Department of Information Technology, Mallareddy Engineering College for Women, (UGC-Autonomous), Hyderabad, India.

**Abstract**

Since Satoshi Nakamoto introduced Bitcoin, its popularity as an alternative method of payment has grown tremendously over the past few years. By the end of 2021, the market value of Bitcoin had exceeded $200 billion. One of the unique features of Bitcoin is its pseudonymous nature, as it is not directly linked to user identities like traditional usernames. This characteristic has led to misconceptions about Bitcoin being completely anonymous and has raised concerns about its potential misuse for untraceable transactions during illicit activities. Tracking Bitcoins associated with known addresses is usually manageable. However, it becomes challenging to trace Bitcoins when criminals utilize vague and ambiguous addresses to obfuscate their activities. Furthermore, the global usage of cryptocurrencies, including Bitcoin, continues to increase steadily, making it crucial to monitor Bitcoin transactions more carefully. Unfortunately, conventional methods have proven to be insufficient in effectively analyzing Bitcoin transactions. Therefore, this research focuses on the development of a Bitcoin transaction network (BTN) using pattern matching rules (PMR). Initially, the dataset undergoes preprocessing to identify missing symbols and unknown characters from the forensic blockchain dataset. Then, a Petri-Net model is applied to the pre-processed dataset, helping to identify properties such as timestamps, transaction IDs, work tera hash, and work error details. The Petri-Net model plays a significant role in parsing and constructing the BTN model. Subsequently, PMR conditions are formulated to extract transaction addresses along with their timestamp details. This allows PMR to detect illegal payment addresses by comparing them with known data, thereby identifying potential spam addresses. Additionally, a cache based PMR (CPMR) is applied to detect fraudulent transactions. CPMR stores all previously detected illegal payment addresses, allowing it to ignore those addresses during new transactions. This results in a reduction of fraud transaction detection time and speeds up the overall processing. The approach shows promise in enhancing the efficiency and accuracy of Bitcoin transaction analysis, addressing the challenges posed by the growing use of cryptocurrencies and the need for more robust forensic investigation methods.

## 1. Introduction

Since Satoshi Nakamoto first introduced bitcoin, its popularity as an alternate method of payment has grown significantly over the last several years [1]. At the end of 2021, it was estimated that the market value of Bitcoin had surpassed $200 billion. Bitcoins are often not linked to user identities like usernames. Due to its pseudonymous character [2], Bitcoin is mistakenly thought of as an anonymous mode of payment on the Internet and as a means of enabling untraceable transactions during illicit dealings. Tracking Bitcoins linked to a known address is often not a problem [3]. However, it has been difficult to trace Bitcoins since criminals often use ambiguous and hazy addresses. Figure 1 shows the various bitcoin frauds occurred in different countries [4] like Vietnam, united states, United Kingdom, Ukraine, turkey, south Africa, Russia, south Africa, and China. The bitcoin frauds are

majority based on darknet markets, ransomware, scams, and stolen funds [5]. In order to deal with this, various works aims to separate bitcoin fraud addresses. Generally speaking, some transactions may show commonalities and recurring trends. For instance, bitcoin transactions [6] were used to accumulate Bitcoins often link an output address to a number of input addresses. When monitoring ambiguous and improbable transactions, examining the connections between such input and output addresses may provide insightful information. However, such analysis involves additional challenges [7] like defining the characteristics of bitcoin transactions, successfully identifying the characteristics that can be used to identify suspects. With the help of our pattern matching technology, we have discovered static and dynamic Bitcoin transaction attributes that identify Bitcoin transaction patterns for analysis and locating questionable addresses. The evolution of the Bitcoin gene, which is integrated in Petri-Net transitions [8], is another significant addition. The movement of Bitcoins may be quickly and reliably tracked and analyzed using bitcoin transaction. Additionally, based on the combinations of match rules, this study suggests a set of match criteria to discover transactions and get suspicious addresses [9].

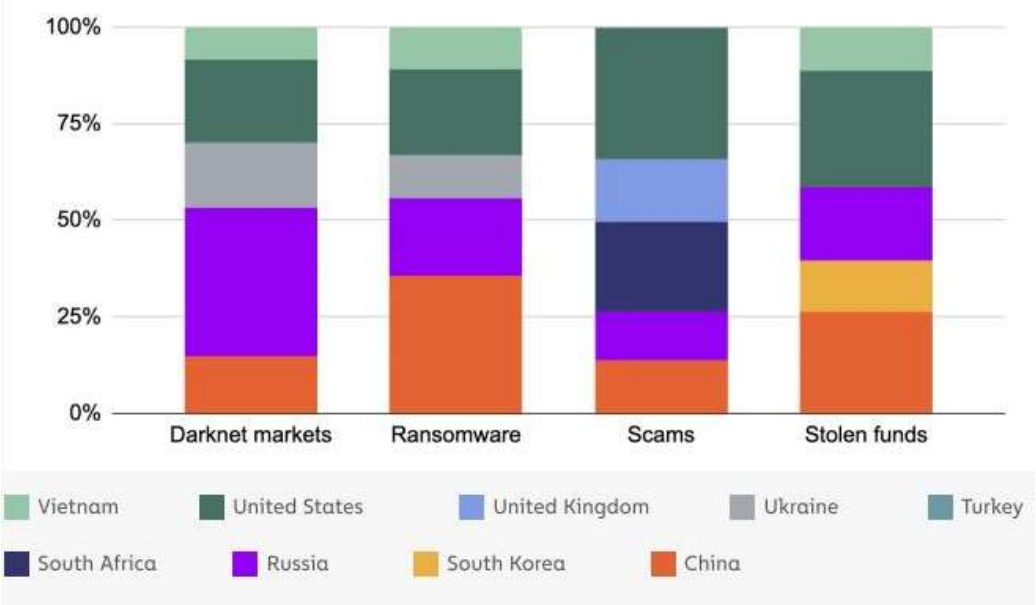


Figure 1. Bitcoin frauds in various countries.

## 2. Literature survey

The owner of bitcoin addresses an is unknown and is under suspicion. However, the owner of Bitcoin addresses a may be inferred if the owner of bitcoin address, which is inside the same cluster as a, is known. Addresses are divided into clusters [9], when they are utilized as transaction inputs. Addresses a and b, for instance, are grouped together into a single cluster if they are used as inputs for transaction t1. Numerous research [10] make advantage of this input address clustering technique. A user graph was built using the bitcoin address clustering (BAC) method, and significant users were identified using PageRank. To recover the "change" from whatever transaction the user has issued. In addition to clustering addresses using the input and change address clustering methods [11], they also assessed the efficacy of the change address clustering approach. BitIodine is a modular framework developed in [12], that parses the blockchain, groups address that are probably owned by the same person or group of users, and visualizes complicated data taken from the Bitcoin network. The two clustering techniques are also used by BitIodine to group addresses. These clusters of addresses utilized as vertexes in several previous flow analysis techniques [13] and the transaction interactions

between vertexes as direction edges. Bitcoins often go from vertexes to vertexes via edges. In their basic Bitcoin flow analysis technique, in [14] authors initially grouped Bitcoin addresses before connecting the clusters using connections made by transactions. Finally, both statistical and graphical tools have been used to study the graph. The hybrid graph [15] was used to examine the graph structure's characteristics that could impact anonymity. But none of these studies have placed enough attention on studying Bitcoin transaction trends.

Transactions involving Bitcoin have been examined using modified Petri-Net [16]. Bitcoin addresses are represented by Petri-Net locations and transitions. This Petri-Net model employed to group addresses and discovered common patterns of behavior, such as the use of a specific address just once. In [17] authors examined disposable addresses to addresses using machine learning based approaches. A power-law distribution characterizes the lengths of these chains. The Bitcoin addresses were employed in these two models as Petri-Net locations or inputs for Bitcoin transactions. On the BTN, inputs for transactions are often coins rather than addresses. As a result, such models are unable to assess and quantify transaction aspects effectively. The expanded Petri-Net for the study of Bitcoin transactions that were presented in contrast to the current approaches that seek to identify behavior factors underlying Bitcoin transactions. According to authors [18], it is possible to recognize and validate Bitcoin users by examining the characteristics of their transactions over time. A data visualization tool called BitCone view was created in [19], to demonstrate the efficiency. They conducted a thorough measuring study of information on the Silk Road that was gathered by web crawling. The data characters have been shown using these visualization techniques. This may use visual perception to identify the transaction patterns in a block. However, this approach is unable to identify subtle trends in large-scale transactional data. BlockChainVis [20] is a program uses visual analytics methods to filter out unwanted information and graphically examine the transactions. Users of BlockChainVis may create simple filters to exclude unwanted information

## 3. Proposed system

Criminals, on the other hand, plan to conceal their Bitcoin addresses in the real world. Due to the paucity of known samples, it is challenging to locate their locations in order to study the transaction attributes, which restricts their practical application. Figure 2 shows the proposed BTN framework. This effort focused on the creation of BTN utilizing PMR. Initially, dataset pre-processing is performed to discover missing symbols and unfamiliar characters in the forensic blockchain dataset. Here, the information is saved into the database using an open-source program called Bitcoin Database Generator. The Petri-Net model is then used to the pre-processed dataset, identifying the time stamp, transaction id, work tera hash, and work error attributes. The Petri-Net model is primarily used to parse and construct the BTN model. Then, PMR conditions are created to retrieve the collected transaction addresses with time stamp data. As a result, PMR identifies illicit payment addresses by comparing known data to illegal addresses. Furthermore, CPMR is used to identify fraudulent transactions, which stores all previously recognized unlawful payment addresses. As a result, for each new transaction, CPMR will disregard any previously recorded (detected) unlawful payment addresses. This effect reduces the time required to identify fraud in transactions and speeds up processing. When compared to existing approaches, the simulations demonstrate that the suggested method TPT, FTDT, and enhanced FTDA.

### 3.1 Pre-processing

The raw forensic blockchain dataset contains noises, missing values, which caused to complicated training of CPMR model. Further, it will reduce the classification, prediction performance. So, the

data preprocessing operation is performed to overcome these problems. The preprocessing operation will replace unknown symbols, missing vales with the known nearest values. The efforts of filtering options that may be used to discover certain transactions or addresses. This approach defined a transaction pattern and instead took into account transaction characteristics independently. Methods of visualization often rely on visual perception to get outcomes. However, certain outcomes often go unnoticed because of the capacity limitations of the human brain. In our solution, the pattern matching algorithm instead of visual perception is used to match transaction patterns. As a result, it is more effective and seldom overlooks details. In order to filter out certain false positive samples using our suggested strategy, marginal distributions of several transaction characteristics are shown using visualization methods rather than being directly analyzed.

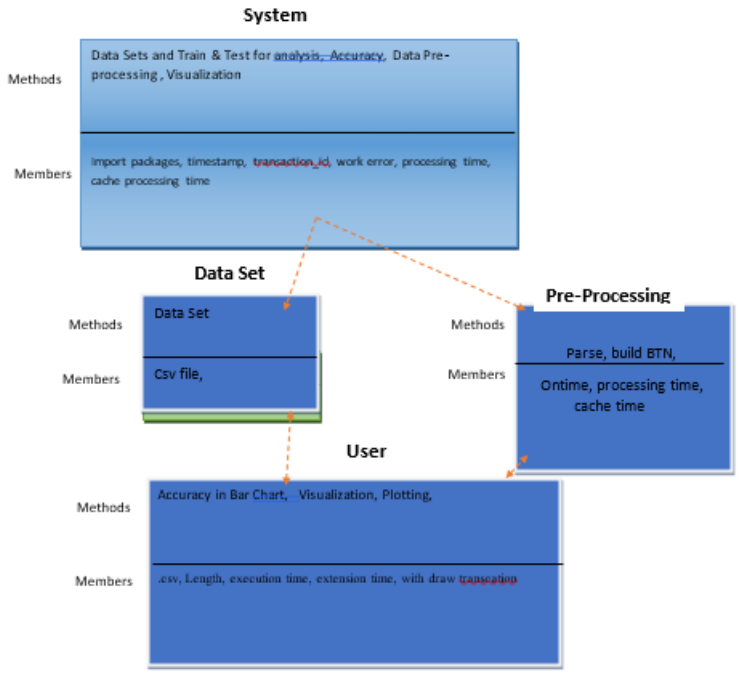


Figure 2. Proposed BTN-CPME block diagram.

### 3.2 Petri-Net Model

A formal mathematical model called a Petri-Net, which is used to explore concurrent and asynchronous processes in distributed systems. An alternative name for it is a place/transition (PT) net. It was initially developed in 1962 by Carl Adam Petri. It is a bipartite directed graph with two different kinds of nodes, locations, and transitions. Directed arcs link the locations with the transitions, indicating which locations serve as inputs before transitions take place and outputs once they do. Arcs can only link locations to transitions or locations to transitions. Tokens are stored in places. Transitions cannot keep any tokens, but places may store an endless amount of them. The distribution of tokens among locations determines the state or marking of a Petri-Net. The Petri-Net are more effective at catching concurrent actions as an assault progressed. Petri-Net have since been used to simulate physical and digital assaults on a variety of systems and networks. An innovative approach for studying the BTN was suggested in this research. This system formalizes Bitcoin transactions as an extended Safe Petri-Net known as BTN. The static and dynamic properties of a Bitcoin transaction are described by its structure and semantic features. The Bitcoin flow analysis may exploit the gene characteristic of bitcoins. Different transaction patterns may be developed based on the qualities that have been stated. It is possible to identify the addresses that fit the patterns. Based on a review of actual case studies, the suggested technique has been shown to be a useful tool for

forensic investigation of future Bitcoin transactions. Pattern expressions are manually designed for our investigations. The next stage will be to create a compiler to automatically turn patterns into code. We shall then keep looking on ways to preserve BTN's interim states. In our tests, the Bitcoin Blockchain is parsed using the open-source application bitcoin database generator. The performance of the analysis is impacted since it does not retrieve information about block and transaction ordering.

## 4. Results and discussion

Further, Table 1 shows that the proposed BTN-CPMR protocol resulted in higher security standards compared to BAC [10], BitIodine [13], and BlockChainVis [20]. Because, the proposed BTN-CPMR approach reduced the TPT (ms), FTDT (ms), and increased the FTDA (%). This research presented an innovative methodology for the investigation of the Bitcoin transaction network. In this architecture, Bitcoin transactions are formalised as an expanded version of the Safe Petri net, which is referred to as BTN. Static and dynamic aspects of a Bitcoin transaction may be understood by its structure and the semantic qualities it has. The DNA feature of Bitcoins may be used for the investigation of the movement of Bitcoins. There are many other transaction patterns that may be defined based on the qualities that have been stated. It is possible to determine which addresses correspond to certain patterns. Based on a review of real-world case studies, the approach that was developed has been shown to be an effective instrument for use in future forensic investigations of Bitcoin transactions.

In our studies, the pattern expressions are created by hand programming. In the subsequent stage, a compiler will be constructed to automatically compile patterns into programmes. This will be the first step. The data included in the Bitcoin Blockchain is enormous; thus, it is possible to reduce the amount of time spent on doing a recent case study by saving intermediate states at various periods in time. Following that, we will proceed to look at further ways that intermediate states of BTN might be kept safe. In our research, the Bitcoin Blockchain is dissected with the help of an open-source programme called Bitcoin Database Generator. However, it does not retrieve information about blocks and transaction orders, which might negatively impact the analysis performance. Another one of our study goals for the future is to investigate this tool's downside, so stay tuned for that! Forensic examination of bitcoin will be aided in the future by our implementation of a fully integrated bitcoin analysis platform.

Table 1: Performance comparison of existing and proposed models.

| Method | FTDA (%) | TPT (ms) | FTDT (ms) |
|---|---|---|---|
| BAC [10] | 91.056 | 43.614 | 42.516 |
| BitIodine [13] | 92.969 | 21.661 | 35.905 |
| BlockChainVis [20] | 93.636 | 17.308 | 17.456 |
| Proposed BTN-CPMR | 98.927 | 9.352 | 8.440 |

## 5. Conclusion

The primary emphasis of this effort was placed on the construction of the BTN-CPMP. In the beginning, the dataset is preprocessed so that any missing symbols or unfamiliar characters in the forensic blockchain dataset may be located and accounted for. The preprocessed dataset is then

subjected to a Petri-Net model application, which detects attributes such as the time stamp, transaction id, work tera hash, and work error. The Petri-Net model was primarily used in order to construct and parse the BTN model. The PMR criteria needed to extract the transaction addresses together with the time stamp data are then generated. Therefore, PMR is able to identify illicit payment addresses by comparing the known data with illegal addresses (spam addresses). In addition, a CPMR is used in order to identify fraudulent transactions. This PMR keeps a record of all unlawful payment addresses that have been identified in the past. Therefore, for every new transaction, CPMR will disregard all of those previously recorded (detected) unlawful payment addresses. This will protect the integrity of the network. This phenomenon produces a decrease in the amount of time needed to identify fraudulent transactions, resulting in a speedup of the processing. According to the results of the simulations, the proposed method led to a reduction in the amount of time required for the processing of transactions i.e., TPT, the amount of time required to detect fraudulent transactions i.e., FTDT, and an improvement in the FTDA when compared to the conventional methods. Further, this work can be extended with deep learning models for improved performance.

**References**

[1] [1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
[2] [2] D. Bryans, "Bitcoin and money laundering: mining for an effective solution," Indiana Law Journal, vol. 89, pp. 1-33, 2014.
[3] [3] M. J. Barratt, "SILK ROAD: EBAY FOR DRUGS: The journal publishes both invited and unsolicited letters," Addiction, vol. 107, pp. 683-683, 2012.
[4] [4] M. Dittus, J. Wright, and M. Graham, "Platform Criminalism: The'lastmile'geography of the darknet market supply chain," in proceedings of the 2018 World Wide Web Conference on World Wide Web, 2018, pp. 277- 286.
[5] [5] G. White. UK company linked to laundered Bitcoin billions, BBC, (2018). Available: https://www.bbc.com/news/technology-43291026
[6] [6] N. J. Ajello, "Fitting a Square Peg in a Round Hole: Bitcoin, Money Laundering, and the Fifth Amendment Privilege Against SelfIncrimination," Brooklyn Law Review, vol. 80, p. 4, 2015.
[7] [7] P. Reynolds and A. S.M. Irwin, "Tracking digital footprints: anonymity within the bitcoin system," Journal of Money Laundering Control, vol. 20, pp. 172-189, 2017.
[8] [8] F. Reid and M. Harrigan, "An Analysis of Anonymity in the Bitcoin System," in proceedings of 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, 2011, pp. 1318-1326.
[9] [9] S. Göbel, A Polynomial Translation of Mobile Ambients Into Safe Petri Nets: Understanding a Calculus of Hierarchical Protection Domains: Springer, 2016.
[10] [10] D. Ron and A. Shamir, "Quantitative Analysis of the Full Bitcoin Transaction Graph," in proceedings of International Conference on Financial Cryptography and Data Security Berlin, Heidelberg, 2013, pp. 6-24.
[11] [11] M. Fleder, M. S. Kester, and S. Pillai, "Bitcoin transaction graph analysis," arXiv preprint arXiv:1502.01657, 2015.
[12] [12] E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, "Evaluating User Privacy in Bitcoin," in proceedings of International Conference on Financial Cryptography and Data Security, Berlin, Heidelberg, 2013, pp. 34-51.
[13] [13] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, et al., "A fistful of bitcoins: characterizing payments among men with no names,"

presented at the Proceedings of the 2013 conference on Internet measurement conference, Barcelona, Spain, 2013.

[14] [14] M. Spagnuolo, F. Maggi, and S. Zanero, "BitIodine: Extracting Intelligence from the Bitcoin Network," in proceedings of International Conference on Financial Cryptography and Data Security, Berlin, Heidelberg, 2014, pp. 457-468.

[15] [15] M. Harrigan and C. Fretter, "The unreasonable effectiveness of address clustering," in proceedings of 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016, pp. 368-373.

[16] [16] C. Zhao and Y. Guan, "A GRAPH-BASED INVESTIGATION OF BITCOIN TRANSACTIONS," in proceedings of Advances in Digital Forensics XI, Cham, 2015, pp. 79-95.

[17] [17] D. D. F. Maesa, A. Marino, and L. Ricci, "Uncovering the Bitcoin Blockchain: An Analysis of the Full Users Graph," in proceedings of 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2016, pp. 537-546.

[18] [18] D. Di Francesco Maesa, A. Marino, and L. Ricci, "Data-driven analysis of Bitcoin properties: exploiting the users graph," International Journal of Data Science and Analytics, September 25 2017.

[19] [19] M. Ober, S. Katzenbeisser, and K. Hamacher, "Structure and anonymityof the bitcoin transaction graph," Future internet, vol. 5, pp. 237-250, 2013.

[20] [20] A. Pinna, R. Tonelli, M. Orrú, and M. Marchesi, "A Petri Nets Model for Blockchain Analysis," arXiv preprint arXiv:1709.07790, 2017.