A PROPOSAL FOR AN OPEN-SOURCE BITCOIN FORENSICS TOOL

Pedro Henrique Resende Ribeiro, Pedro Leale and Ivan da Silva Sendin

Faculdade de Computação, Federal University of Uberlândia, Uberlândia, Brazil

ABSTRACT

Over recent decades, the global financial paradigm has experienced significant transformations, notably the emergence and adoption of cryptocurrencies. The escalating prominence of assets like Bitcoin has inadvertently catalysed a surge in illicit activities associated with the currency. Consequently, the forensic examination of transactions within blockchains becomes imperative for the detection and surveillance of malevolent undertakings. This research delineates a preliminary pipeline for a Bitcoin forensic analysis tool. Moving forward, the ambition is to conceptualize and empirically validate this tool utilizing data procured from blockchain and ancillary sources. The methodology will harness Open-Source Intelligence (OSINT), clustering of Bitcoin addresses, and an exhaustive financial analysis. Upon finalizing the pipeline, the implementation of an open-source instrument is envisioned, poised to confer substantial advantages to the broader cryptocurrency milieu.

KEYWORDS

Cryptocurrency, Bitcoin, Forensic Analysis, OSINT, Clustering.

1. INTRODUCTION

In the contemporary era, technological advancements have catalysed the proliferation and ubiquitous integration of myriad technologies, notably cryptocurrencies. Such innovations facilitate internet-centric payment systems, obviating the necessity for traditional financial intermediaries in transactional processes [1]. Bitcoin, conceived in 2008, stands as a quintessential exemplar of cryptocurrencies, witnessing an exponential surge in its acclaim over the past years [2]. Nonetheless, the adoption of this technology is not devoid of pitfalls, encompassing potential transactional fraudulence and its exploitation for illicit endeavours, encompassing money laundering [3], Ponzi Schemes [4], and illicit trade like drugs and sex commerce [5].

The confidentiality inherent in cryptocurrency transactions has garnered considerable attention in academic literature. While privacy remains pivotal for the mainstream assimilation of cryptocurrencies, it concurrently invites scrutiny, particularly in the context of malevolent activities leveraging cryptocurrencies for remuneration [6]. The employment of pseudonymous protocols within the Bitcoin ecosystem facilitates malefactors in obfuscating transactional provenance, geolocation, and financial dominion [3] [7]. Consequently, there emerges an imperative for the formulation of analytical tools accentuating the potential misappropriation of Bitcoin.

The Bitcoin Blockchain epitomizes transactional lucidity. Its architecture permits accessibility and manipulation, especially when synergized with potent clustering heuristics like the multi-input heuristic [8]. This heuristic, when initialized with a specific address set, endeavours to

Dhinaharan Nagamalai (Eds): NLAII, VISCOM, CCSITA, AIDD, BCYIoT -2023
pp. 173-178, 2023. IJCI – 2023 DOI:10.5121/ijci.2023.1206013

identify additional addresses potentially under analogous ownership, thereby enriching the data reservoir for subsequent forensic dissection.

Open-Source Intelligence (OSINT) is predicated on the systematic aggregation of data to distil strategic insights on a designated subject [9]. Within the ambit of decentralized finance, OSINT has manifested its indispensability as an instrumental apparatus for the meticulous analysis of transactional integrity within blockchain network. The proficiency in tracing transactions and pinpointing users within blockchain infrastructures can yield invaluable intelligence for fraud deterrence, criminal investigations, and the surveillance of dubious undertakings [10].

The objective of this work is to elucidate the confluence of clustering methodologies and OSINT in discerning illicit blockchain activities and their correlation to tangible users. The strategy encompasses grouping data, such as Bitcoin transactional inputs, attributable to a specific entity, be it collective or individual. Subsequent phases involve the extraction of intelligence from digital platforms, including social media, discussion forums, and pertinent websites, with the intent of associating discovered addresses with the personal data of the transactional agents.

2. BACKGROUND

This segment endeavours to elucidate extant instruments for blockchain analysis and furnish the requisite technical minutiae of the blockchain imperative for analytical undertakings.

2.1. Economic and Behavioural Aspects

Subsequent to its genesis, Bitcoin has garnered the allegiance of multitudes, facilitating transactions of substantial magnitude and thereby engendering a pronounced economic reverberation. Its inherent autonomy has engendered polarized discourses, inciting deliberations regarding its prospective trajectories and the environmental impacts it has already caused [11].

The inherent freedom and higher level of anonymity compared to traditional currency, has rendered it a magnet for illicit endeavours, such as the online sale of narcotics following early proof-of-concept transactions [12]. It has also been utilized for ransomware payments [13].

Several works available in the academic literature underscore the utility of economic attributes as forensic analytical tools, encompassing transactional history [14], the Gini coefficient, address lifetime, transactional frequency, and transacted values [4], inter alia.

The transactional historic facilitates the tracing of fund provenance and the discernment of antecedent transactions tethered to a specific address, thereby unveiling interconnections amongst entities, and bolstering financial malefaction probes [14]. The Gini coefficient quantifies affluence concentration within addresses or transactions, potentially flagging nefarious activities such as capital hoarding by malevolent actors [15].

Address lifetime possesses salience; short-term addresses might insinuate surreptitious endeavours, such as transient address utilization to obfuscate fund origins. Conversely, enduring addresses proffer insights into transactional histories and user credibility [14]. Elevated transactional frequencies or substantial cryptocurrency volumes might be harbingers of money laundering, market manipulation, or other illicit activities, spotlighting atypical patterns warranting in-depth analysis [4].

2.2. Open-Source Intelligence

OSINT, originally conceived to satiate military requisites for public domain strategic intelligence, facilitates the extraction of pivotal information [9]. Since its inception, myriad proposals have been proffered, advocating its application across diverse domains.

Cryptocurrency transactional network analysis emerges as a paramount application of OSINT in contemporary times. [16] accentuate that such analytical endeavours potentiate the identification of dubious transactions and the interrelations amongst implicated users. This modus operandi has been harnessed to buttress criminal investigations tethered to cryptocurrency misuse in illicit activities.

Fraud detection in cryptocurrency transactions represents another OSINT application delineated in academic treatises. [17] underscore the efficacy of machine learning in pinpointing suspicious cryptocurrency transactions, enabling the recognition of patterns indicative of illicit undertakings. OSINT's synergy with data analysis can be instrumental in unearthing cryptocurrency transactional fraud and pre-empting financial malefactions.

The emergence of Bitcoin and its cryptocurrency counterparts has precipitated a gamut of apprehensions encompassing security, discretion, and regulatory dimensions. Within this milieu, OSINT has manifested as an invaluable analytical tool. Its prowess in transactional tracing and user identification within blockchain ecosystems proffers invaluable intelligence for fraud mitigation, criminal probes, and the surveillance of suspect activities [10].

Furthermore, OSINT's applicability extends to issues tethered to cryptocurrency confidentiality. While Bitcoin operates within a relatively transparent network, alternative cryptocurrencies endeavour to proffer augmented user anonymity. In this scenario, OSINT methodologies can critically assess these stratagems and pinpoint potential vulnerabilities [18].

2.3. Clustering

In simple terms, Bitcoin transactions involve input and output addresses. Inputs are linked to outputs from previous transactions, creating a chain [19]. During a transaction, inputs are used up, and outputs are labelled as Unspent Transaction Outputs (UTXOs). Each UTXO represents a certain amount of Bitcoin linked to an address. A user's balance is the total of all UTXOs connected to their addresses [19].

Given this, we can use clustering methods, like the multi-input heuristic, to group sets of addresses in Bitcoin transactions that might belong to the same user [8]. This method assumes that if transactions share inputs, then all those inputs belong to the same user.

Another method, the one-time change heuristic, looks at Bitcoin transactions that need a change address. This address is usually used once and belongs to the same user as the input addresses [20]. Besides these methods, other techniques are discussed in the literature, like analysing mixing services and mining payments [21].

3. PIPELINE PROPOSAL

Given the technological paradigms delineated, specifically blockchain forensic clustering methodologies and OSINT, the envisaged pipeline is predicated upon the integration of a pre-defined ensemble of potentially dubious addresses into a clustering algorithm, exemplified by the

multi-input heuristic. In the ensuing phase, the addresses discerned from this procedure will be harnessed in OSINT methodologies to ascertain the tangible entities concomitant with them. Figure 1 provides a schematic representation of the pipeline as postulated in this research.

Moreover, the proposal advocates for the adoption of a data structure optimized for Bitcoin heuristics. This data structure integrates a Bloom filter and a search tree to efficiently store and execute intersection and union operations required by the clusterization algorithm. The data structure will be developed in Rust language, due to its innovative memory management system and fast execution speed.

The OSINT facet will be implemented in Python, imbibing web scraping methodologies and harnessing API-driven data retrieval to collate pertinent information. Additionally, the integration of artificial intelligence methodologies is contemplated to bolster the analytical rigor of the information.

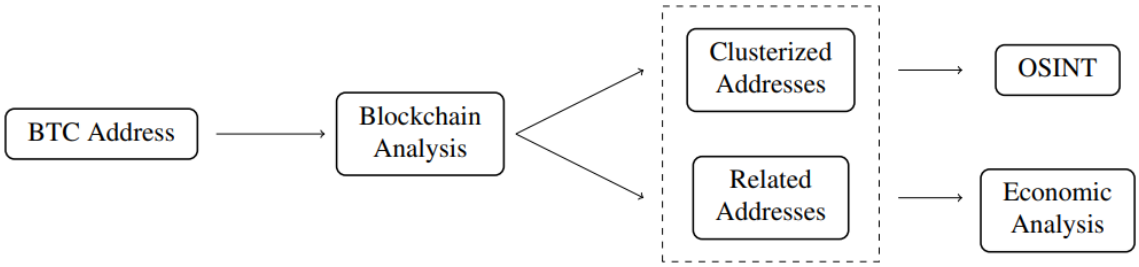

Figure 1. Pipeline proposed to implement a Bitcoin forensics. Once an addressof interest exists, addresses owned by the same entities can be inferredby clusterization, and related addresses are obtained from the Blockchain.Further analyses can be done using OSINT and economic features.

4. CONCLUSIONS

The development of forensic instruments tailored for Bitcoin holds significant promise in bolstering the security of transactions executed within the Blockchain. Within the purview of this research, the amalgamation of clustering techniques, in tandem with Open-Source Intelligence (OSINT), empowers the elucidation of transactional trends and the identification of transactions that bear hallmarks of suspicion. Consequently, this paves the way for pre-emptive measures to discern and thwart malevolent undertakings, thereby fortifying the security apparatus of the financial milieu.

Moreover, a plethora of enterprises, including but not limited to Chainalysis, Ciphertrace, and Elliptic, have entrenched themselves in the realm of Bitcoin forensics, curating tools that are instrumental in pinpointing and probing illicit transactions. This underscores the palpable demand for cryptocurrency forensic instruments in the contemporary market, accentuating the pertinence of this research endeavour.

In conclusion, the judicious deployment of these forensic tools, complemented by proper regulations and awareness initiatives, is paramount to safeguarding the integrity of the decentralized financial architecture and mitigating criminal transgressions. The relentless evolution of Bitcoin forensic methodologies and the refinement of employed techniques are indispensable in navigating the ever-evolving intricacies of malevolent activities, thereby ensuring sustained trust in this burgeoning financial paradigm.

REFERENCES

[1] Mukhopadhyay, U., Skjellum, A., Hambolu, O., Oakley, J., Yu, L., and Brooks, R. (2016).A briefsurvey of cryptocurrency systems. In 2016 14th annual conference on privacy, security, and trust (PST), pages 745–752. IEEE.

[2] Vranken, H. (2017). Sustainability of bitcoin and blockchains. Current opinion in environmental sustainability, 28:1–9.

[3] Monamo, P., Marivate, V., and Twala, B. (2016a). Unsupervised learning for robust bitcoin fraud detection. In 2016 Information Security for South Africa (ISSA), pages129–134. IEEE.

[4] Bartoletti, M., Pes, B., and Serusi, S. (2018). Data mining for detecting bitcoin ponzischemes.

[5] Foley, S., Karlsen, J. R., and Putnins, T. J. (2019). Sex, drugs, and bitcoin: How much illegal activityis financed through cryptocurrencies? Review of Financial Studies,32:1798–1853.

[6] Feng, Q., He, D., Zeadally, S., Khan, M. K., and Kumar, N. (2019). A survey on privacy protection in blockchain system. Journal of Network and Computer Applications,126:45–58.

[7] Monamo, P. M., Marivate, V., and Twala, B. (2016b). A multifaceted approach to bitcoinfraud detection : Global and local outliers. In 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 188–194. IEEE.

[8] Harrigan, M. and Fretter, C. (2017). The unreasonable effectiveness of address clustering. Proceedings - 13th IEEE International Conference on Ubiquitous Intelligence and Computing, 13thIEEE International Conference on Advanced and Trusted Computing, 16th IEEE International Conference on Scalable Computing and Communications, IEEE International, pages 368–373.

[9] Evangelista, J. R. G., Sassi, R. J., Romero, M., and Napolitano, D. (2021). System at ic literature review to investigate the application of open-source intelligence (osint) with artificial intelligence. Journal of Applied Security Research, 16(3):345–369.

[10] Wu, Y., Luo, A., and Xu, D. (2019). Forensic analysis of bitcoin transactions. In 2019IEEEInternational Conference on Intelligence and Security Informatics (ISI), pages167–169. IEEE.

[11] Badea, L. and Mungiu-Pupazan, M. C. (2021). The economic and environmental impact of bitcoin. IEEE Access, 9:48091–48104.

[12] Bohme, R., Christin, N., Edelman, B., and Moore, T. (2015). Bitcoin: Economics, technology, and governance. Journal of Economic Perspectives, 29(2):213–38.

[13] Paquet-Clouston, M., Haslhofer, B., and Dupont, B. (2019). Ransom ware payments inthe Bitcoin ecosystem. Journal of Cyber security, 5(1). tyz003.

[14] Lin, Y.-J., Wu, P.-W., Hsu, C.-H., Tu, I.-P., and wei Liao, S. (2019). An evaluation of bitcoin address classification based on transaction history summarization.

[15] Lin, Q., Li, C., Zhao, X., and Chen, X. (2021). Measuring decentralization in bitcoin and Ethereum using multiple metrics and granularities. In 2021 IEEE 37th International Conference on Data Engineering Workshops (ICDEW), pages 80–87. IEEE.

[16] Tharani, J. S., Charles, E. Y. A., Hou, Z., Palaniswami, M., and Muthukkumarasamy, V. (2021).Graph based visualisation techniques for analysis of blockchain transactions. In 2021 IEEE 46thConference on Local Computer Networks (LCN), pages 427–430.IEEE.

[17] Podgorelec, B., Turkanovic, M., and Karakatic, S. (2019). A machine learning-based method for automated blockchain transaction signing including personalized anomaly detection. Sensors,20(1):147.

[18] Conti, M., Kumar, E. S., Lal, C., and Ruj, S. (2018). A survey on security and privacy issues of bitcoin. IEEE Communications Surveys & Tutorials, 20(4):3416–3452.

[19] Antonopoulos, A. M. (2017). Mastering Bitcoin: Programming the Open Block chain. O'Reilly Media, Inc., 2nd edition.

[20] Zhang, Y., Wang, J., and Luo, J. (2020). Heuristic-based address clustering in bitcoin. IEEE Access,8:210582–210591.

[21] He, X., He, K., Lin, S., Yang, J., and Mao, H. (2022). Bitcoin address clustering method based on multiple heuristic conditions. IET Blockchain, 2(2):44–56.

AUTHORS

Pedro Henrique Resende Ribero Graduated in Mechanical Engineering (2019) from the Federal University of Uberlândia (UFU). Currently, he is a Computer Science undergraduate student at UFU and in the Software Engineering program at the International University Center (UNINTER). Currently enrolled in a postgraduate program in Data Science and Big Data Analytics at the University of Uberaba (UNIUBE)

Pedro Leale Computer Science undergraduate student at Universidade Federal de Uberlândia (UFU). Conducted research in the field of bitcoin forensics, focusing on the development of a hybrid data structure.

Ivan da Silva Sendin Graduated in Computer Science from the State University of Londrina (1996) and completed Master's (1999) and Ph.D. (2012) degrees in Computer Science from the State University of Campinas. Currently, he is a professor at the School of Computing at the Federal University of Uberlândia.