Data Analytics for Blockchain Forensics

André Martinez
Instituto Superior Tecnico, Universidade de Lisboa
Lisbon, Portugal
andre.martinez.pereira@gmail.com

***Abstract*—Blockchain has been given a major focus in the most recent years as its innovating technology has revolutionized the way information is stored on a decentralized way on its network and the speed at which as its transactions are executed that would normally take days on traditional banking. A major problem is surfacing in this space, as cybercriminals use this technology for their own benefit stealing millions of crypto assets and getting out free of charge.**

This work will focus on explaining how cybercriminals execute their schemes on the blockchain network, showing step by step the process on how they launder money on this network and providing a methodology on how to gather proof on the blockchain network and a tool to gather the overall information of the analytics of multiple projects and analyse their smart contracts to see if they are exploitable or safe for investors.

Key words: Blockchain, Cybercrminals, Smart Contracts, Mixer, Zk-snark

I. INTRODUCTION

Blockchains are secure and tamper proof digital ledgers [1] implemented in a distributed environment and usually without a central figure of authority. They allow users to record transactions in a shared ledger and in the blockchain network no transaction can be changed once published. In the beginning of the creation of the blockchain concept around 2008 when BTC was created the blockchain concept was combined with other technologies and computing concepts to create the cryptocurrencies we have today which is cryptocurrencies protected through cryptographic mechanisms instead of a central bank or other figures of authority.

The first blockchain [2] that follows this definition was Bitcoin. Inside the Bitcoin blockchain the information represents electronic money and is attached to an address. Bitcoin users can transfer information to another user and the Bitcoin blockchain records this transfer publicly, allowing all participants of the network to verify the validity of the transactions inside of the blockchain. The Bitcoin blockchain is stored and collaboratively managed by a decentralized and distribute participants. This, along with cryptographic mechanisms, makes the blockchain resilient to attempts to alter the information inside the public ledger. Blockchain technology is the foundation of cryptocurrencies, users utilize public and private keys to sign and securely transact within the system while keeping pseudonymity [3].

The first blockchain design was created by Nakamoto and has evolved since then with new types of blockchains with smart contracts [4]. That allows you to create decentralized applications inside blockchains or provide liquidity for decentralize finance. With this, some malicious people use these tools for their own benefit since creating their own tokens so they can steal the users who invested on that asset or exploit a vulnerability in a smart contract on a blockchain to take out the liquidity of a certain exchange exploiting a smart contract function and with that start swapping the tokens that they stole from that contract and then they send the stolen tokens to a private mixer so we can't initially track them.

Developers try to help fix this issue by using convolutional networks [5] to distinguish between good addresses and suspect addresses that could be involved in money laundering to help track suspected money flow. The problem with many of these tools is that they aren't open source [6] or require payment to be able to use their services [7] making it harder to develop solutions faster to address these problems.

To analyse these projects and do smart contract audits on them I developed a software named CoinFetch that uses the CoinGecko and the Go + Security API to gather data analytics from the overall project such as price, volume, social media presence through the CoinGecko API and smart contract audits from projects in the BSC and Ethereum blockchain that tells us if the smart contract is safe or exploitable through the Go + Security API.I also developed the first documented forensic analysis of a cryptocurrency scam and a smart contract exploit in the BSC and Ethereum blockchain.

II. RELATED WORK

The work that has been done before in this area focuses on smart contracts exploits on the Ethereum Network showing us the different vulnerability's that a contract has [8]. It also focuses on extracting Ethereum raw data tool to gather data analytics [9] and the possible use cases that they give and an analysis of the Monero mixer that tries to trace the user identity of the person who used the mixer and explain how they were able to find the vulnerabilities of that mixer that allowed for the traceability of those individuals [10].

The first work is about the types of exploits on a Ethereum smart contracts to have an understanding of what types of

exploits a smart contract can suffer exposing their vulnerabilities. Smart contracts define what is allowed to do with them and who can interact with them, the most important case is when the contract will deposit on a user since in the case it has a vulnerability it can be exploited taking all the money also known as reentrancy attack

When these types of contracts are vulnerable the attacker will steal portions or all the funds that were inside the contract which will result on a loss of assets for the owner of the contract. To identify such vulnerabilities in these contracts we need to analyse the contracts code to reveal critical vulnerabilities that might be used to withdraw the funds of those contracts. The hackers are able to do this due to four critical EVM instructions and we can label these four instructions into two categories: The first two instructions cause a direct transfer of the contract and the other two instructions allow random execution of Ethereum bytecode within the context of those contracts.

The second work focus on blockchain data analytics framework named eXplore Blockchain ETH (XBlock-ETH) [9], that analyses Ethereum data. The tool gathers three types of Ethereum data that are blocks, traces, and receipts. This data will help researchers to investigate and analyse Ethereum data in a useful way. The data was obtained from the Ethereum blockchain. It collects three types of blockchain data that as previously mentioned was block, receipt and trace. We can describe each of these fields as the following:

Block: Block data is being kept in the Ethereum blockchain, each one of these blocks consists of two components the first one is the block header where it gathers the basic information of the block and the second component is block transactions where the body of the block is built.

Trace: Trace data is the run-time data that was generated in Ethereum Virtual Machine. Trace data is referred to the data that is obtained during the execution and can be displayed in three types: Create obtains information about the creator of the contract, its code, and the balance that a smart contract starts when it is deployed. Call happens when cryptocurrencies or information are being transferred into different addresses in the Ethereum network, Reward happens when the miners on Ethereum mine a block and that reward depends of their contribution to mine it.

Receipt: When the transaction is executed, some states on Ethereum have been changed and we need to know what was that changed that happened. To reduce the expenses of the users that use those smart contracts those contracts leave an Event that details what happened in that transaction and that can be read by the users.

XBlock-ETH focus on three areas: Blockchain System Analysis, Smart Contract Analysis and Cryptocurrency Analysis.

Blockchain System Analysis: The XBlock-ETH can process data from blockchains and can use it to focus on these types of blockchain system analysis that are: decentralized analysis since its data offers a good overall understanding of the transactions that happened on Ethereum making us obtain multiple data for decentralized analysis and predicting gas price costs making the users save a lot of cost fees paying the minimum required to do that transaction and not pay extra gas for the transaction.

Smart Contract Analysis: The XBlock-ETH can be used in the studies of smart contracts in these different areas that are: Similarity between contracts since there is a great similarity between the smart contract codes and the calling of those codes making the developers able to know what would be the user experience before hand, detection of vulnerability in contracts since a number of malicious attacks on the Ethereum blockchain have made investors lose huge losses in cryptocurrency assets, detection of fraud since smart contracts are used by a lot of investors they can be used to scam them and example of these would be ICO contracts with vulnerabilities to exploit the contract and go away with their money and being able to detect that fraud would safeguard a lot of investors.

Cryptocurrency Analysis: XBlock-ETH can be used to analyse cryptocurrency in the following three fields: Crypto transaction analysis of cryptocurrency transactions to capture the order flow of those transactions to help us detect possible money laundering schemes, Analysis of cryptocurrency prices in which price analysis consists of three steps that begin by collecting the price information from exchanges, find patterns between the price to be able to help make better decision making and predict possible future prices to obtain profit. The last field is fake user detection that its used to detected possible fake users that are used to improve the activity ranking of a project to make investors believe the project has more activity that it actually has

The third work done in tries to trace the transaction flow in the Monero mixer. Monero is a privacy-oriented cryptocurrency focused on hiding the identity of its users by including worthless coins called mixins to camouflage the actual coins that are spent. The author evaluated two weaknesses in the Monero's mixin strategy. The first weakness is that about 62% of the transaction inputs with mixins are vulnerable to be found by process of deduction and with this actually find the real input that was used to send the real coins and the second weakness is that Monero mixins are able to be distinguish by the age of when those coins were created and with this identify with a high degree of confidence the real input of that transaction.

III. BINANCE SMART CHAIN ECOSYSTEM

Analysing the ecosystem inside of the Binance Smart Chain (BSC) [11] we can see that BSC is a blockchain that runs parallel to Binance Chain (BC). BSC was created to start using smart contracts since BC didnt use them.

The BC blockchain serves the following purposes: to be able to do transactions to send and receive cryptocurrencies, token creation control and the circulating supply in the network through multiple functions such as: creating coins, deleting coins and stack or unstake coins.

Binance Smart Chain (BSC) started working in September 2020 and they created their smart contract inspired mainly by the Ethereum smart contract design. When they were designing BSC they focused on four different aspects that are: Keeping the BSC and BC blockchain separated in case of them went down it wouldn't affect the other, creation of a stake model to guarantee the participation in voting on the BSC proposals and block creation and validation of the BSC, They have compatibility with Ethereum since they copied the general design of Ethereum smart contracts, they asked Ethereum for help for the BSC in which the difference is the BSC consensus mechanism is PoSA in which the users of the BSC can use their BNB to become validators on the BSC network and they developed cross chain mechanism to be able to easily connect between the BSC and BC. To be able to do transactions on the BSC we need to use something as gas to pay for the cost of the transaction that is the currency of that ecosystem and that's when we will use BNB that is the coin that Binance uses to pay for the gas fees.

BNB was created in July 2017. It started as an Initial Coin Offering to gather money from multiple investors and to do that, they would incentivize offering discounts on the trading fees on the Binance exchange using BNB to motivate investors to invest on their cryptocurrency. BNB now serves three purposes that are the following: Pay gas fees in the BSC network, to stake the BNB on the BSC and get rewards from it and to perform cross-chain transactions.

IV. TORNADO.CASH

The mixer Tornado Cash [12] is a decentralized privacy protocol that is built on Ethereum. Users can deposit and withdraw tokens in addresses that are different from the ones that made or took the deposit, with this they increase transaction privacy between their deposit and withdrawal addresses.
The protocol creates a secret hash before admitting the deposit and the hash. When the user wants to withdraw their crypto, they input the secret hash to prove they are the one who deposited the funds. Since the funds go through the liquidity pools of Tornado Cash we can't connect the transaction of the person who deposited and the one who withdrawal providing privacy.

Tornado Cash is owned in by its community since the ownership was transferred in 2020 when the developers gave their control to them making the protocol decentralized completely. Tornado Cash uses zk-SNARK and is able to hide the transaction information making it untraceable and ready to be deposited on crypto wallets like Metamask or Trust Wallet. Using Zero proof the protocol improves its privacy by making the communication of transactions secure without revealing important transaction information.

Tornado Cash has anonymity mining designed to provide Tornado Cash with liquidity to be able to hide their transactions. The user by interacting with the Tornado Cash protocol receives points that are deposited on a protected account that can be used to convert them to TORN that's the currency used on the Tornado Cash protocol.

Tornado Cash was a popular protocol in the crypto space it came with a lot of controversy, the primary one occurred recently August 2022 when U.S. Treasury implemented sanctions for using this protocol. But these sanctions only started forming in March 2022 where Roman Semenov stated to Bloomberg News [13] that sanctions on decentralized protocols are impossible since there isn't a way to control a fully decentralized space.

The Treasury Department on August of 2022 applied sanctions taking down the Tornado Cash website and forbidding US users of using that protocol since the majority of transactions in there are supposed to be for money laundering purposes.

The Treasury Department stated that Tornado Cash was used for the money laundering that should be worth around of $7 billion in cryptocurrencies since the beginning of the protocol which it even includes a cryptocurrency theft by Lazarus group that goes around of 455$ million in multiple cryptocurrencies stolen. All currency that is taken out of Tornado Cash is now associated with the imposed sanctions of the Treasury Department and that implies that the exchanges or businesses that use Tornado Cash services will be considered to have the addresses who use those services tainted and the Treasury of Department should go after those individuals.

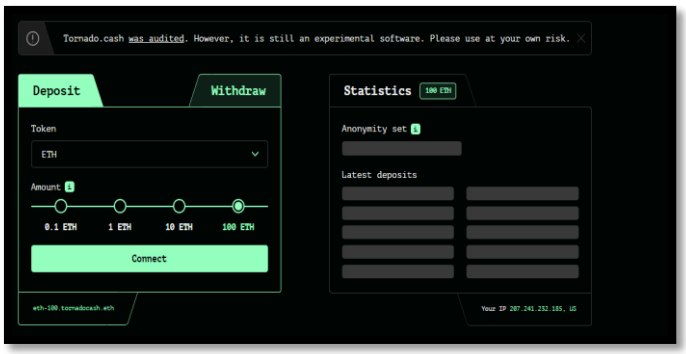

Figure 1: Tornado Cash Mixing Service

The person who wanted to use the mixing services, in the case of this work, cybercriminals after stealing set amount of tokens of different protocols of the liquidity pool and swapped them for Ethereum he would then send the Ethereum after connecting his wallet to this website depositing the Ethereum according to set quantity defined by the amount above and it would send the Ethereum to a Mixing pool and the cybercriminal would get a hash to recover the stolen Ethereum on an account defined by him and couldn't be traced since the tornado cash application is used as a 3rd party that sends the money to a new account not giving us a direct link of the stolen money.

We can try to guess if an abnormal amount of Ethereum gets out of Tornado Cash soon after the theft that is stolen money with high probability but only a rookie cybercriminal would do that since in general, they would leave the money in the mixing process for some while to not be traceable. The solution that I was going to propose is what the U.S Treasury did that's making all accounts that interacted with this service suspicious and couldn't take out the money in an exchange that is linked to his KYC and would be easily be caught since he wouldn't be able to explain with a sound explanation where did this money came from.

V. BLOCKCHAIN ANALYTICS FOR FORENSIC ANALYSIS

To do forensic research in the blockchain I developed CoinFetch that is a tool that gathers information from two applications. The first one is CoinGecko that we use to gather information about crypto in general, this allows us to gather: price data, volume data, social media websites, contract address, Liquidity scores, the all-time high price and all-time low. The second application is Go + Security that we use to do the smart contract audits giving us different metrics to access the smart contract security such as: it has a function to take back ownership of that contract, if it can change the balance of one of the tokens holders, if its a honeypot contract were this contract calls another function from a different contract, whether he can blacklist or whitelist a user, has the self-destruct function and have external calls to other contracts.

The two main functions of this tool are to do basic analysis to gather overall information about these projects or through advanced analysis do smart contract audits to verify their safety.

The Basic analysis functionality of my tool consists in calling the CoinGecko API and gathers all the information about price, volume and social media websites, contract address, Liquidity scores, The price % change up to two hundred days ago also with the all-time high price and all-time low.

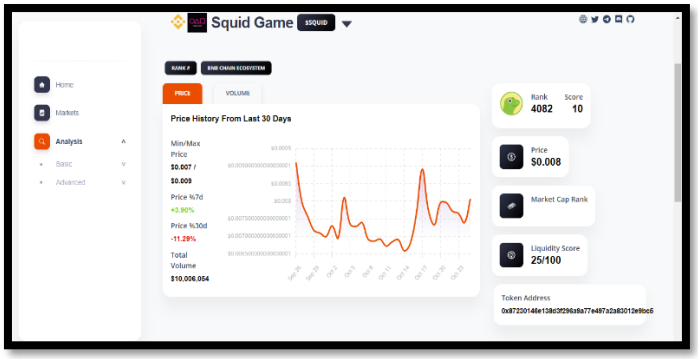

Figure 2: Basic analysis to gather information about Squid Game Token

The advanced analysis tool consists in doing smart contract audits to smart contracts addresses that we gather through the CoinGecko API and using the Go + Security API from the information that we gathered we can do smart contracts audits that are compatible with the BSC and ETH blockchain. This function has two parts the Overall shows all the information, Contract audits, contract ABI and contract address we are analysing. The second is Security Information that only shows the information of the smart contract audit.

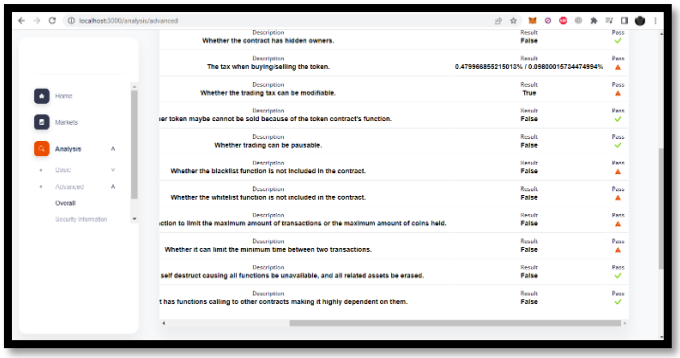

Figure 3: Smart Contract Audit using Go + Security

SQUID GAME TOKEN

The Squid Game Token was a scam that occurred between October and November of 2021 and was the first case that got major attention going to main stream media since investors weren't allowed to sell their tokens and that allowed the price of the token to soar from one cent to 3400 $ and collapsed to 0$ making investors loose around 3.38 million $ [14]. Using the CoinFetch tool we manage to gather the overall information of this project using the Basic Analysis from the tool that you can see on Figure 2 and then got the URL to their website.

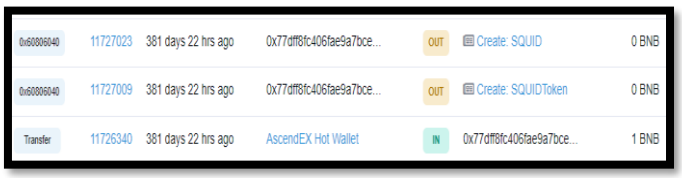

Figure 4: Cybercriminal receives money and creates contracts

The user cybercriminal started by receiving 1 BNB from the AscendEX Hot Wallet so he would be able to create the contracts since that's the price to create a smart contract in the Binance smart chain. The money came from a hot wallet from an exchange that didn't have KYC (we can not identify the person because that exchange doesn't collect that information) besides since it came from a hot wallet that's a wallet that exchanges use to send the transactions from users to not be directly linked from that transaction for privacy issues, we can't link it to cybercriminals.

The hacker then proceeds to swap the money to them send it to a mixer that in this case was Tornado.Cash leaving the cybercrminals untraceable starting from this point.

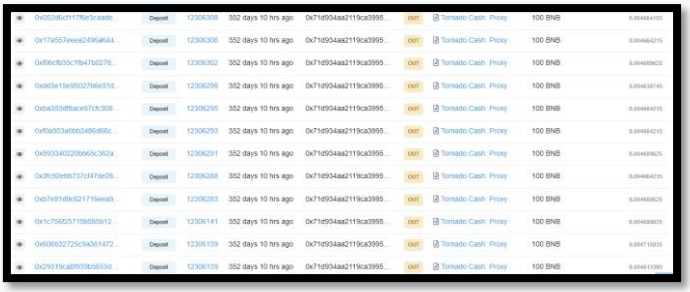

Figure 5: Squid Game Token cleaned on Tornado.cash

LIQUID EXCHANGE

On August 19th of 2021, [15] this exchange was hacked for a total of 90 million $. The exchange announced the incident through twitter indicating that their hot wallets were compromised. The incident happened because the hot wallet wasn't protected correctly and the hackers got access to it via 3 possible ways, a phishing attack, malware or an inside job making the hackers have access of the wallet and allowing to transfer assets. After that the hacker started to swap the tokens using multiple ways, he sends the money to CEX to take it out which doesn't make much sense unless he managed to fool the KYC system of the Huobi and Bitlaxy since they should have the ID of the hacker behind this. They also used Tornado.cash, in this case, they made a mistake in which they send directly the money to 1 of the wallets that would be used to receive the money from Tornado.cash and we can see what he did after it.

The total amount stolen from the two hot wallets amounts to 90$ million at the time of transaction and the cybercriminal starts swapping the other tokens it stole to ETH in multiple different wallets to me easier to clean the money on the mixer and in this case other platforms.

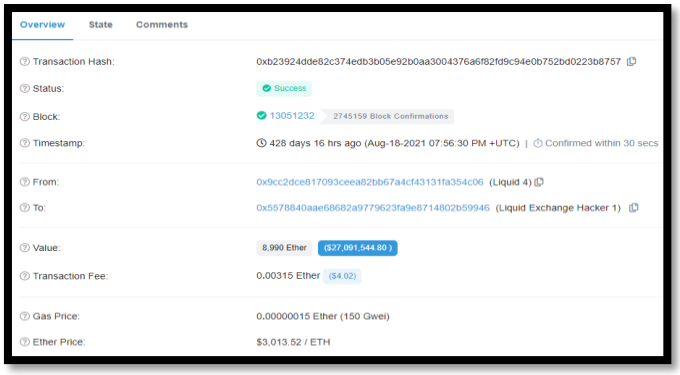

Figure 6: 27$ million stolen on a single transaction

The hacker after swapping the money then proceeds to send the money to Tornado.cash but he made a mistake that made us identify the wallets he used after the mixer.

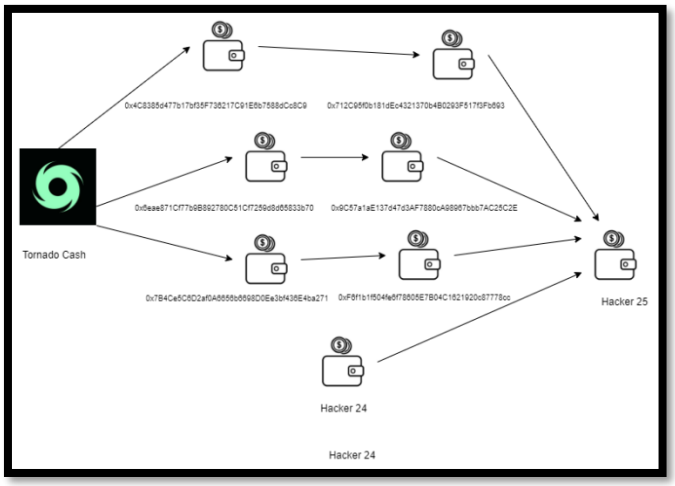

Figure 7: Following the Mistake of the hacker

Following the trail of the hacker we can see that he then proceeds to use RenBTC to switch to another blockchain to in the future try to bring the money to the real word. We manage to track the wallets that received the money and they havent moved till today. The hacker recently proceeded to move some address on this scheme to launder a portion of money that he left behind and then proceeded to use the Tornado Cash application to clear his trails, even after stealing such amount of money the hacker hasnt been caught and its still operating inside the blockchain.

Comparison Between the Squid Game Case and the Liquid Exchange Case

Table 1:
Comparison between Squid Game and Liquid Exchange

Comparison	Scam Token	Wallet Exploit	Stolen funds	Swapping	Tornado.Cash
SQUID	X		14 million $	X	X
Liquid Exchange		X	90 million $	X	X

By analysing these two cases and in comparison, we can see that even though the methods on how this scam started in different ways the process in which they try to launder the money is the same. Swapping the tokens stolen from investors by exploiting the smart contracts to make the cybercriminals able to steal the funds from the investors and then proceed to mix the money to erase the trail they left behind using Tornado.Cash and in the future cybercriminals will probably use some mixing tool even more advanced than this one making it even harder to find them. In the Squid Game Token case they got away with 14 million dollars and on the Liquid Exchange case they stole around 90 million $.

We can observe that exchanges are the highest target for exploits for the huge potential of profits in they manage to exploit their smart contract and obtain their funds and get away with them. With this comparison we were capable to create a good methodology on how to gather evidence on an efficient matter on decentralized finance exploits and scams, the blockchain since its immutable makes this evidence unforgeable and provides us with high detail on how every transaction was done inside it through the power of data analytics. With this we created a forensic model that can be used on the future to gather this evidence.

VI. Evaluation

We created CoinFetch in node.js using React to have a clean and simple user interface. The software connects with two different API end points that are CoinGecko and Go + security. We use the CoinGecko API to obtain data about cryptocurrency projects that are gathered in the CoinGecko repository providing us with clean data about the overall information of the project. The second API is Go + Security that is an API that allows us to do smart contract audits by using the information that CoinGecko provides us about the contract address and verify if that address its safe for use or it can be exploitable.

The framework that we elaborated to analyse cryptocurrency scams consists in three phases: Find the scam token developer address or the first transaction after an exploit of a hot wallet to find the initial flow the transactions The second phase consists in documenting the swapping transactions with their respective wallets to able to trace the money flow inside the blockchain. Finally, the last phase is finding the cryptocurrency mixer wallets that should be at the end of the transaction flow when the hackers deposit the money inside the mixer making us lose their trail.

The results obtained were satisfactory since we managed elaborated a tool that was able to gather information from a project in the BSC or ETH network to provide us with the overall scope of what that project consists and provide a solution to be able to audit smart contracts on those networks to see if the contract has any vulnerabilities that can be exploited to steal the funds from crypto users guarantying with this the safety of the individual that uses the CoinFetch tool if he chooses to use those smart contracts.

The second solution I provided was creating a simple framework that helps people with no prior crypto forensic research understand in a simple manner how those cybercriminals operate in the blockchain network and how to track their crypto transaction flow with no additional tool.

VII. Conclusions

With this work were able to analyze how decentralized finance works and provide a methodology to analyze these scams and exploits in a pretty efficient matter on how everything works since cybercriminals start using their modus operandi to swap the tokens after exploiting the smart contract for their benefit either by creating a honeypot contract or through malware/phishing to obtain the credentials to an exchange hot wallet to steal their funds.

To then learned how a mixer works and how the cybercriminals send the investors funds to it to clear evidence that links them to that stolen money using Tornado.cash.

The only solution that I can provide to avoid cybercriminals launder money is make a watchlist of all the wallets that used a mixer and if they aren't able to justify where did those funds came from its highly likely they were stolen even though the use of a mixer is not illegal this is the only way I can imagine to keep safe the assets of stolen investors. So, the methodology it's still the same but the medium in which is done may differ but with this work we were able to learn how to use the power of data analytics to analyse in a forensic matter the blockchain and help guide people understand how to do a forensic investigation in this matter.

ACKNOWLEDGMENT

I would like to thank my thesis supervisor Miguel Nuno Dias Alves Pupo Correia for supporting me throughout the development of my thesis and to whom I dedicate the success of this project and also my family and friends who gave me the necessary support to finish this work.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. [Online]. Available: https://www.debr.io/article/21260.pdf.

[2] M. D. Pierro, "What Is the Blockchain?," *in Computing in Science & Engineering*, vol. 19, no. 2017, pp. 92-95, September/October 2017.

[3] Fergal Reid, Martin Harrigan, "An Analysis of Anonymity in the Bitcoin System," 22 June 2011. [Online]. Available: https://arxiv.org/pdf/1107.4524.pdf.

[4] D. G. Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger," 2017. [Online]. Available: https://gavwood.com/paper.pdf.

[5] Mark Weber, Daniel Karl I. Weidele, Giacomo Domeniconi, Claudio Bellei, Charles E. Leiserson, Jie Chen, Tom Robinson, "Anti-Money Laundering in Bitcoin: Experimenting with Graph," 31 June 2019. [Online]. Available: https://arxiv.org/pdf/1908.02591.pdf.

[6] Dinesh Srivasthav P, Lakshmi Padmaja Maddali,Vigneswaran R, "Study of Blockchain Forensics and Analytics tools," *2021 3rd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS),* 2021.

[7] Bitqquery, "Bitquery," 18 August 2020. [Online]. Available: https://bitquery.io/blog/best-blockchain-analysis-tools-and-software.

[8] Johannes Krupp,Christian Rossow, "teether: Gnawing at Ethereum to Automatically," *27th USENIX Security Symposium,* 2018.

[9] Peilin Zheng , Zibin Zheng , Jiajing Wu, Hong Ning Dai, "XBlock-ETH: Extracting and Exploring," *IEEE Open Journal of the Computer Society 1 (2020),* pp. 95-106.

[10] Malte Möser, Kyle Soska, Ethan Heilman, Kevin Lee, Henry Heffan, Shashvat Srivastava, "An Empirical Analysis of Traceability in the," 13 April 2017. [Online]. Available: https://arxiv.org/abs/1704.04299.

[11] R. J. Dolor, "Binance Smart Chain (BSC) Explained | A Beginner's Guide," FinBold, 30 May 2022. [Online]. Available: https://finbold.com/guide/binance-smart-chain/#Introduction.

[12] Bybit, "Tornado Cash: How It's Stirring Up a Storm in the Crypto World," Bybit, 4 September 2022. [Online]. Available: https://learn.bybit.com/defi/what-is-tornado-cash/.

[13] M. Shen, "Crypto Mixer Tornado Cash Says Sanctions Can't Apply To Smart Contracts," 10 March 2022. [Online]. Available: https://www.bloomberg.com/news/articles/2022-03-10/crypto-obfuscator-tornado-says-sanctions-cant-affect-smart-contracts.

[14] BBC News, "Squid Game crypto token collapses in apparent scam," BBC News, 2 November 2021. [Online]. Available: https://www.bbc.com/news/business-59129466.

[15] Donovan, "Tracking the Stolen Assets from the Liquid Exchange Hacking: Laundering Process, Exchanges Involved, Post Tornado Cash?," Sentinel Protocol Team, 31 August 2021. [Online]. Available: https://medium.com/sentinel-protocol/tracking-the-stolen-assets-from-the-liquid-exchange-hacking-acd94e01c762.

[16] R. J. Dolor, "https://finbold.com/guide/binance-smart-chain/#Introduction," 30 May 2022. [Online]. Available: https://finbold.com/guide/binance-smart-chain/#Introduction.

[17] Bybit, "https://learn.bybit.com/defi/what-is-tornado-cash/," 2022. [Online]. Available: https://learn.bybit.com/defi/what-is-tornado-cash/.

[18] Haozhe Zhou, Amin Milani Fard , Adetokunbo Makanju, "The State of Ethereum Smart Contracts Security: Vulnerabilities, Countermeasures, and Tool Support," 27 May 2022. [Online]. Available: https://www.mdpi.com/2624-800X/2/2/19.