



Traceability of Cryptocurrency Transactions using Blockchain Analytics

Kanwal Gagneja¹, Andre Goode¹, David Rentos¹ and Karim Rezk¹

¹Florida Polytechnic University, Florida, USA

Received 01 Aug. 2019, Revised 06 Feb. 2020, Accepted 23 Feb. 2020, Published 01 Mar. 2020

Abstract: Cryptocurrencies are increasingly used as a medium to conduct illicit activity. The fundamental concept empowering Bitcoin and other cryptocurrencies is the blockchain, which is a distributed ledger technology that transactions are stored on. The entire ledger, which is widely distributed among the peer-to-peer network, is susceptible to the temporal analysis of the transaction to be performed. This type of blockchain analytics threatens the expected privacy of transactions and has resulted in the growing number of privacy-oriented solutions that reduce the traceability and increase the anonymity of crypto transactions. Despite the evolving technological advancements of privacy-oriented features for cryptocurrencies, traceability of transactions is still possible. In this paper, the core features of blockchain are reviewed as well as their resistance to traceability. Existing countermeasures that attempt to obfuscate user activity are also considered. Also, a prototype software solution is proposed that could be used in the incidence response of criminal activity involving cryptocurrencies.

Keywords: Blockchain, Incident Response, Temporal Analysis, Privacy, Traceability

1. INTRODUCTION

Bitcoin (BTC) launched in early 2009 through the release of a highly-cited academic paper by the anonymous founder(s) Satoshi Nakamoto. The paper aimed to solve the problem of secure money transfer by introducing a peer-to-peer (P2P) technology known as blockchain [1]. Since then, the Bitcoin network has become recognized as a secure digital form of transacting. This acceptance has led to a growing adoption by vendors and consumers but has subsequently also contributed to a growing number of criminals using the cryptocurrency to perform illegal transactions [2].

Bitcoins are sent to public keys, which act as addresses to cryptographic wallets that store the currency. These addresses are cryptographic hashes that allow users to anonymously send currency without having to link their identity to the transaction directly. Because of this, Bitcoin operates in a semi-anonymous fashion where transactions may seem anonymous when viewed in isolation. Historically, authorities have utilized Bitcoin's public ledger to assist in investigations identifying transactions over the Bitcoin ledger to specific suspects. For example, Husam et al. were able to successfully link user's Bitcoin wallets to transactions that were linked to onion services, which are sites on the anonymity network known as The Onion Ring (TOR) [3]. In addition to the drug trade marketplace on Silk Road, Bitcoin has been a

common factor in many reports detailing the use of cryptocurrencies in numerous cybercrimes, such as a payout for ransomware and financing know terror organizations [4], [5]. As a result of the traceability provided by the blockchain, covert cryptocurrency users have discovered various methods to aid in better anonymity, which will be highlighted later in this paper. Despite these obfuscation techniques, the core architecture of the blockchain still permits analysts the ability to trace transactions.

Using incident response and data analysis techniques on the blockchain-enabled cryptocurrencies, our work will review the techniques that remove the pseudo-anonymity from transactions. The next two sections review the concepts of blockchain technology, which make it inherently traceable. Section IV discusses the technique and software tool implementation that can assist authorities and investigators in future cases involving bitcoin. The implementation also highlights bitcoin alternatives that improve privacy among blockchain transactions. Section V explains the countermeasures of blockchain analysis. The last section concludes the paper.

2. OVERVIEW OF BLOCKCHAIN

The distributed ledger technology that serves as a foundation for Bitcoin consists of a list or chain of transactions that are stored in a group called a block. The



blockchain continues to grow as more transactions occur. This ledger service runs through a distributed protocol where connected nodes form a consensus for the protocol and serve as validators [6].

A core concept of blockchain technology is the distributed consensus mechanism that is used to implement ledger consistency and resiliency. A permission-less (public) blockchain architecture allows nodes to freely join and participate in the network, whereas a permission-based (private) blockchain typically restricts the nodes that can participate in the computation and validation of transactions [7].

The blockchain is a cryptographic technique that allows secure authentication of a transaction as a part of a global ledger. Whenever a digital transaction is ready to be transferred over the blockchain, the transaction is recorded in a global ledger. It means that at any point, any person in the blockchain network can instantly determine the validity of a new transaction. As an example, let's say that we have an entity A, which contains data on the blockchain network as depicted in fig. 1 below. If A wants to send B the data, the transaction is broadcasted to the network and added to the ledger, which anyone within the network can review at any time. Before each transaction is added into the ledger, it is verified by the rest of the network. Only transactions that have been validated can exist on the ledger, as shown in fig 2.

To prevent attackers from falsifying transactions on the network and provide secure transmission, any time a new transaction is proposed, special nodes on the network known as “miners” verify the transaction by computing whether or not the transaction is valid. In fig 3, entity A tries to send the data to entity D. However, the miners in the network will determine that this transaction cannot occur because A previously transferred the data to entity B. To prevent a single point of failure, the ledger is distributed to everyone on the network. A financial reward is offered to miners once transactions have been validated, which ensures the integrity of the ledger. A blockchain can allow anyone on the network to determine who sent any piece of data, and when they sent it. It also solves the problem of double-spending as it prevents a recipient from accepting a transaction from a sender who references the output of a previous transaction.

There are several features of the blockchain that give it advantages over traditional ledger architectures.

A. Peer-to-peer

The distributed nature of the blockchain allows each entity access to the entire history of transactions. Therefore, it is not controlled by a third-party entity, removing the risk of single-point-of-failure that exist in centralized ledgers.

B. Immutable

Bitcoin's proof-of-work validation service makes it impossible to feasibly attack the integrity of the chain of transaction blocks. It makes blockchain-based architectures highly auditable as each block added to the chain can be independently verified for authenticity. To launch a successful attack of falsifying the ledger, it would require the majority of nodes in the entire network to be malicious.

C. Resilient

By requiring nodes to validate entries through a proof-of-work system, the blockchain is protected against any attacks that target the consensus reached by participating entities. If an attacker wished to overcome this, it would require them to redo the proof-of-work computation for that block and all other blocks after that in addition to surpassing the continuing work of benign nodes [1].

D. Low cost

Nodes are freely able to participate in the network and are the entities that participate in the consensus determination. It removes the dependence of financial institutions and, using mining incentives, encourages the participation of more nodes, which results in a transactional cost advantage [8].

E. Programmable Logic

Transactions on the ledger can be triggered by programmable events that allow for the blockchain to be easily accessed and interacted with [9].

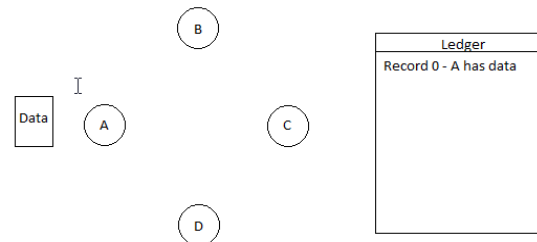


Figure 1. Node A contains the data.

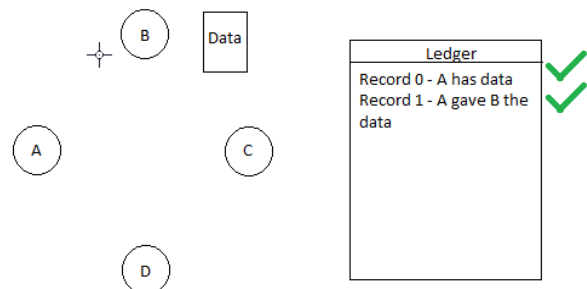


Figure 2. All of the transactions on the network are recorded.

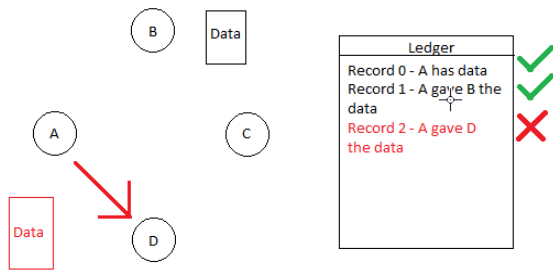


Figure 3. Node A tries to send Node D the data.

3. TRACEABILITY OF CRYPTOCURRENCIES

Some Bitcoin users assume a level of anonymity due to the usage of addresses that either send or receive amounts of the cryptocurrency. Bitcoin uses the Elliptic Curve Digital Signature Algorithm (ECDSA) to generate the asymmetric keys used in transferring ownership of Bitcoins. After attaching a digital signature of the hash of the previous transaction using the source's private key, which is linked to a user's wallet, and including information of the public key of the recipient [10].

Incident responders, along with the research community, have utilized two main techniques to overcome the pseudo-anonymity that Bitcoin creates through its cryptography mechanism. The first involves network analysis of the events on the peer-to-peer network while the second focus is on the transaction records on the blockchain [11].

The pseudo-anonymity of Bitcoin is vulnerable to exposure through the network level. Biryukov et al. have detailed how clients are identified by their established outgoing connections, which serve as an identifier to trace to specific users in an investigation [12]. While this method is more invasive, network analysis is possible in a passive manner. Simply through observation of transaction relay traffic, mappings were able to associate Bitcoin addresses to IP addresses [13] successfully. There is a lot of information online that can be combined with network analysis to assert conclusions of association. For example, large exchanges of Bitcoin can provide investigators with identifiable information such as IP addresses, which can be linked to certain transactions [14].

Analysis of the transactions embedded on the blockchain can also assist in identifying unique users. More specifically, by using a graph of all Bitcoin transactions and a statistical approach, [15] can identify that several large (>50,000 BTC) transactions originated from a single source. Similarly, by combining a statistical approach to a subset of simulated transactions, Androulaki, Elli, et al. concludes that the profile of users can be revealed using a behavior-based clustering approach [16]. A clustering technique was also used by Meiklejohn, Sarah, et al., which involves grouping transactions by known users and user types [17].

Furthermore, when used in combination with publicly available information scrapped from Bitcoin coin web forums and social media networks such as Twitter, transactions to specific entities can be linked to originating from known users [18] [19]. Using this public data capture technique in combination with transactional graph analysis, information helpful in incident response is extracted. The software tool introduced in the next section explores using simple transactional analysis as an incident response tool in Bitcoin investigations.

4. SOFTWARE TECHNIQUE TO BLOCKCHAIN ANALYSIS

A report released by the blockchain forensics firm Chain-analysis notes the growing trend of using Bitcoin and other cryptocurrencies in crypto crime, such as through black markets on Tor but also in scams, ransomware, and hacking [20]. Academic, government and digital forensics organizations have shifted a lot of focus to the traceability of cryptocurrency transactions on a blockchain. This paper presents a software tool created to assist in the analysis of the Bitcoin blockchain to track outgoing transactions. There were several design considerations when creating a tool that assists in the exploration of the blockchain, which at the time of publication, was over 164GB, according to blockchain.info. The large value of transactions that occur make it difficult to trace the movement of money.

A. Software tool overview

The proposed software accepts a Bitcoin address as an input. Then, JSON files are retrieved from the internet using a web-scraper. A series of calls are made to blockchain.info to retrieve any transactions that include the target address. Blockchain.info provides a service that allows users to easily explore Bitcoin transactions throughout the entire blockchain without the need to maintain the most up-to-date ledger. Then, to create a master JSON file with all the outgoing transactions originating from the target address, any result that lists the address as a destination is excluded. Using the complete list of outgoing transactions, and implemented data visualized depicts the source address with references to outgoing transactions sorted by the month and year that they took place. Fig. 4 summarizes the logic of the presented tool.

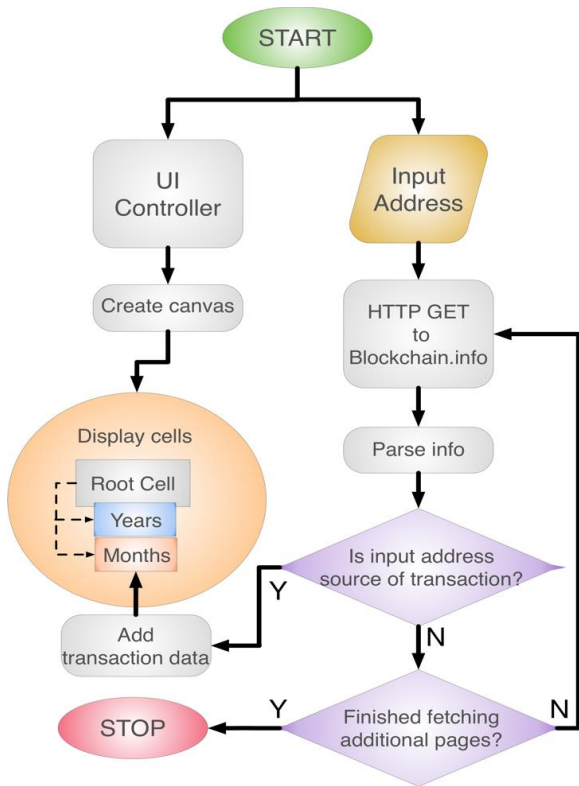


Figure 4. Flow diagram of the proposed software.

B. Limitations to the program

To simplify the output for users of the software, only outgoing transactions are displayed for each target address. It was a design choice due to the limited scope of the research in identifying outgoing transactions of a specific address rather than the incoming amounts or a combination of both thereof. It restricts the utility of the application as in some investigations, and it may be needed to consider incoming transactions to realize the potential of blockchain analytics fully. Additionally, this tool is a manual investigation technique which allows the user to find paths between addresses.

C. Example of usage

To showcase the software’s graphical user interface and its ability to retrieve outgoing transactions based on an address, we examine the wallet address for the Tails project, which solicits donations to a published Bitcoin address 1BvBMSEYstWetqTFn5Au4m4GFg7xJaNVN2. Once a target address is entered, the program begins fetching transaction information linked to that input. The tree depicted in fig 5 shows the visualization created by the program given the specific address. As highlighted in fig 6, the root of the tree is linked to various other nodes that represent each transaction year since the creation of the Bitcoin ledger. Linked to that cell are 12 smaller cells that represent each month in that transaction year. The

sizes of the cells are dynamic, which allows the user to see where more transactions took place easily. For example, in fig 6, it is assumed that the first month of 2017 had more transactions affiliated with the specified address when compared to the visibly smaller second month. Cells represented in orange contain transactions, whereas the gray nodes represent that no outgoing transactions took place. Continuing, a user can then click on the month cell to bring up a graph that depicts all the transactions within that period. Fig.e 7 shows the list of 20 transactions that took place within January 2017. Specific transactional info is revealed when the user selects a leaf. It causes the software to display an information panel which gives information about that specific transaction. Figure 8 depicts the cell info that is displayed. Its total bitcoin value details the transaction, the size of the transaction on the ledger, the number of input and output addresses, along with its hash value.

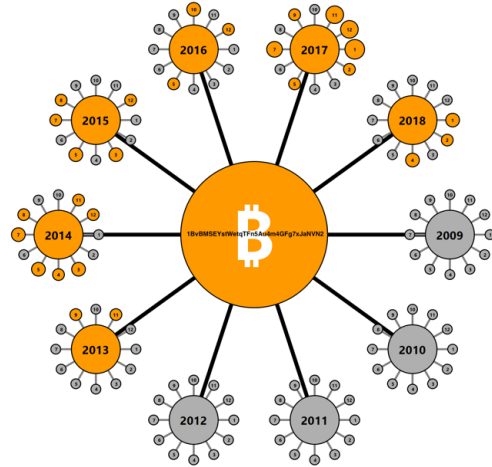


Figure 5. The graph is rooting from the source address.

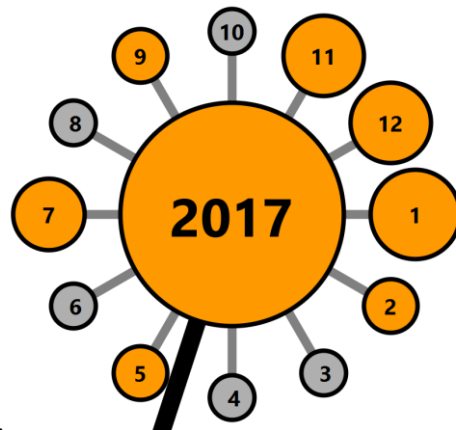


Figure 6. This cell represents the transactions from 2017 sorted by month.

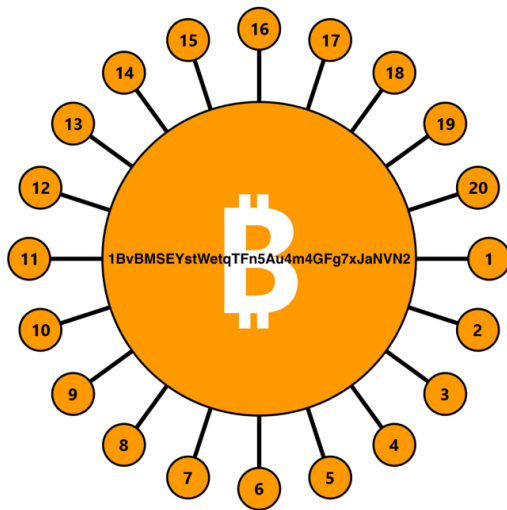


Figure 7. A grouping of 20 outgoing transactions.

Cell Info	
Date:	DECEMBER 20, 2017
Total Value:	1.99868059BTC
Size:	226 bytes
Num Inputs:	1
Num Outputs:	2
Hash	219c40884c49148dde0dd4d666 e3d628f9d12d053d6d7453b6e5 b198f90b1282

Figure 8. The cell info table which displays after a user selects a transaction.

5. COUNTERMEASURES TO BLOCKCHAIN ANALYSIS

Since all Bitcoin transactions are stored on the distributed ledger and each update to the ledger is done through a digitally signed block, each transaction is essentially linked to a pseudonymous address. To better protect transactional privacy, there are a couple of methods that work to obscure any information relating to the sender or receiver of an amount of Bitcoin. While these techniques can be employed in an effort to perform illicit activities (e.g., laundering, ransomware, and terrorism funding), these techniques aim to increase the financial privacy of law-abiding Bitcoin users. While reviewing these techniques, users must also be aware that government entities such as law enforcement and regulators may still be able to extract intelligence using other techniques and with the power to force exchanges to release records of user accounts through subpoenas.

Reuse of bitcoin addresses allows for obvious correlations to exist in the blockchain. With simple blockchain analysis, one can show exactly which transactions were affiliated with a specific address, such as in a forensic investigation [21]. Because of this, one of the most basic obfuscation techniques is to generate different bitcoin addresses for each transaction. As mentioned before, addresses are hashes that can be created freely and easily by users. Therefore, if a user creates a new address for each transaction is increases, the amount of work needed to link transactions to other addresses. In addition to creating new addresses, the structure of a transaction can be randomized to help hide the total amount sent to the recipient, and the change that is sent back to the sender.

Another obfuscation technique is known as mixing, also referred to as a tumbler. This technique relies on the cooperative effort of a group of senders to mix their coins [22]. It typically is achieved by sending the amounts to a service, which then repays most of their original payment minus a small fee that is charged by the service [23]. A disadvantage to using a mixer is the exposure to a single-point-of-failure if a centralized mixing service is used.

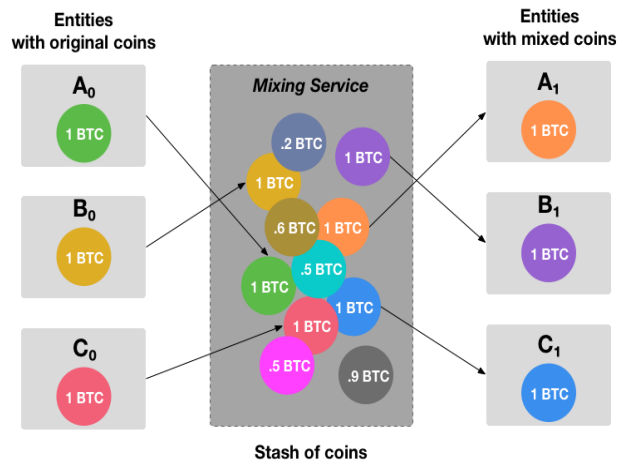


Figure 9. A hypothetical example of bitcoin mixing where the participants end up with equal values but different coins.

The mixing patterns of some services can also be relatively obvious, allowing the detection of such a service with simple analysis [244]. Some community suggestions attempt to solve this issue by using a peer-to-peer infrastructure where not only are the participants unaware of the destination addresses, but this information remains unknown to the mixing service [25], [26], [27]. A diagram depicting a hypothetical situation where three sources of Bitcoin use a mixing service to obfuscate the source of the original bitcoin amount is shown in fig 9.

Alternative coins, or altcoins, have emerged onto the cryptocurrency market as challengers to Bitcoin's



marketplace dominance. A growing number of altcoins place a higher emphasis on transactional privacy, which is a demonstrated weakness concerning anonymity with Bitcoin. These newer currencies seek to improve upon Bitcoin's lack of true privacy by improving existing techniques and introducing other privacy features to reduce traceability [28]. One of the most popular cryptocurrencies, Monero, has taken presence as the replacement to Bitcoin on the dark web and is ranked as one of the top cryptocurrencies as measured by market capitalization.

Some attempts to implement a cryptographic anonymity scheme which further decentralized the anonymization process such as with Zerocoin, which is a protocol allowing users to mix their Bitcoin using decentralized mixing [29]. An improvement to this protocol, called Zerocash, instead implemented a "zero-knowledge non-interactive arguments of knowledge" (zk-SNARK). This validation method improves the speed and cost of transactions, which can remain completely anonymous since balances cannot be deduced from the blockchain [30]. Since that research into shielded transactions with zk-SNARK was released, an increasing number of altcoins have implemented it within their frameworks. One of which, Bitcoin Private, is a coin forked from the Bitcoin framework, which integrates the usage of zk-SNARK to allow for transactional metadata to be hidden to likability [31].

6. DISCUSSION AND RELATED WORK

Improving the effectiveness of investigations through more advanced analytical techniques will contribute to a better understanding of privacy within Bitcoin and can be applied to similarly-based cryptocurrencies. With a focus on data mining and big data analytics, better approaches and investigative practices will emerge from the research. Several papers describe methods that work to reduce the anonymity of Bitcoin transactions using machine learning approaches. Harlev, M.A., et al. describe a technique that leverages Gradient Boosting to achieve high accuracy of predicting the category of previously unidentified transactional clusters [32]. These works have further provided an impetus to the industry to develop automated tools to perform an incident response on the Bitcoin's public blockchain.

7. CONCLUSION

The blockchain technology has shown resiliency to attack of the public ledger mechanism, but this allows for an openness that threatens the privacy of transacting users. Incident response and data analysis techniques on the Bitcoin blockchain work to remove the pseudo-anonymity from Bitcoin transactions. This paper overviewed the core concepts which make up for this pseudo-anonymity as well as techniques used in the

traceability of transactions. Software solutions can provide investigators with easy-to-use tools and blockchain data visualizers to speed up transaction analysis. There are many attempts to fix the lack of privacy following these discoveries, including using external services, improving blockchain techniques, and suggestions of entirely new frameworks oriented around transaction privacy. However, Bitcoin's adoption level and market dominance still allow for many crypto transactions to be vulnerable.

REFERENCES

- [1] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." (2008).
- [2] Ermakova, Tatiana, et al. "Bitcoin: Drivers and Impediments." (2017).
- [3] K. Kaur, X. Xiaojiang Du and K. Nygard, "Enhanced routing in Heterogeneous Sensor Networks", IEEE Computation World'09, pp. 569-574, Athens, Greece, Nov. 15-20, 2009.
- [4] Weimann, Gabriel. "Going dark: Terrorism on the dark Web." *Studies in Conflict & Terrorism* 39.3 (2016): 195-206.
- [5] Gagneja K.K. and Nygard K., "Heuristic Clustering with Secured Routing in Heterogeneous Sensor Networks", IEEE SECON, New Orleans, USA, pages 51-58, June 24-26, 2013.
- [6] Swanson, Tim. "Consensus-as-a-service: a brief report on the emergence of permissioned, distributed ledger systems." The report, available online, Apr (2015).
- [7] Gagneja K.K., "Pairwise Post Deployment Key Management Scheme for Heterogeneous Sensor Networks", 13th IEEE WoWMoM 2012, San Francisco, California, USA, pages 1-2, June 25-28, 2012.
- [8] Meiklejohn, Sarah, et al. "A fistful of bitcoins: characterizing payments among men with no names." *Proceedings of the 2013 conference on Internet measurement conference*. ACM, 2013.
- [9] Fleder, Michael, Michael S. Kester, and Sudeep Pillai. "Bitcoin transaction graph analysis." *arXiv preprint arXiv:1502.01657* (2015).
- [10] *The Changing Nature of Cryptocrime*. Chainalysis, 2018
- [11] Möser, Malte, and Rainer Böhme. "Anonymous alone? Measuring bitcoin's second-generation anonymization techniques." *Security and Privacy Workshops (EuroS&PW), 2017 IEEE European Symposium on*. IEEE, 2017.
- [12] Gagneja K.K., "Pairwise Key Distribution Scheme for Two-Tier Sensor Networks", IEEE ICNC, Honolulu, Hawaii, USA, pp 1081-1086, Feb. 3-6, 2014.
- [13] Gagneja K., Nygard K., "Energy Efficient Approach with Integrated Key Management Scheme for Wireless Sensor Networks", ACM MOBIHOC, Bangalore, India, pp 13-18, July 29, 2013.
- [14] Chan, Jeff, Tanya Liu, and Eddie Xue. "Analyzing the Bitcoin Transaction Graph: A Look at Mixers and Traceability."
- [15] Gagneja K.K., Nygard K., "A QoS based Heuristics for Clustering in Two-Tier Sensor Networks", IEEE FedCSIS 2012, Wroclaw, Poland, pages 779-784, Sept. 9-12, 2012.
- [16] K. K. Gagneja, K. E. Nygard and N. Singh, "Tabu-Voronoi Clustering Heuristics with Key Management Scheme for



- Heterogeneous Sensor Networks", IEEE ICUFN 2012, Phuket, Thailand, pages 46-51, July 4-6, 2012.
- [17] Miers, Ian, et al. "ZeroCoin: Anonymous distributed e-cash from bitcoin." Security and Privacy (SP), 2013 IEEE Symposium on. IEEE, 2013.
- [18] Gagneja K.K., Nygard K., "Key Management Scheme for Routing in Clustered Heterogeneous Sensor Networks", IEEE NTMS 2012, Security Track, Istanbul, Turkey, pp. 1-5, 7-10 May, 2012.
- [19] Sasson, Eli Ben, et al. "Zerocash: Decentralized anonymous payments from bitcoin." Security and Privacy (SP), 2014 IEEE Symposium on. IEEE, 2014.
- [20] A. S. Gagneja and K. K. Gagneja, "Incident Response through Behavioral Science: An Industrial Approach," 2015 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, 2015, pp. 36-41.
- [21] Tirado E., Turpin B., Beltz C., Roshon P., Judge R., Gagneja K., "A New Distributed Brute-Force Password Cracking Technique", Future Network Systems and Security, FNSS Communications in Computer and Information Science, vol. 878, pp 117-127, 2018
- [22] Caleb Riggs, Tanner Douglas, and Kanwal Gagneja, "Image Mapping through Metadata," Third International Conference on Security of Smart Cities, Industrial Control System, and Communications (SSIC), Shanghai, China, 2018, pp. 1-8.
- [23] Keely Hill, Gagneja K.K., "Concept network design for a young Mars science station and Trans-planetary communication", IEEE MobiSecServ 2018, Miami, FL, USA, Feb. 24-25, 2018.
- [24] Javier Campos, Slater Colteryahn, Gagneja Kanwal, "IPv6 transmission over BLE Using Raspberry PI 3", International Conference on Computing, Networking and Communications, Wireless Networks (ICNC'18 WN), March 2018, pp. 200-204.
- [25] Gagneja K., Jaimes L.G., "Computational Security and the Economics of Password Hacking", Future Network Systems and Security. FNSS 2017. Communications in Computer and Information Science, vol. 759, pp. 30-40, Springer, 2017.
- [26] Ziegeldorf, Jan Henrik, et al. "Coinparty: Secure multi-party mixing of bitcoins." Proceedings of the 5th ACM Conference on Data and Application Security and Privacy. ACM, 2015.
- [27] Gagneja K.K. Ranganathan P., Boughosn S., Loree P., and Nygard K., "Limiting Transmit Power of Antennas in Heterogeneous Sensor Networks", IEEE EIT2012, IUPUI Indianapolis, IN, USA, pages 1-4, May 6-8, 2012.
- [28] C. Riggs, J. Patel, and K. Gagneja, "IoT Device Discovery for Incidence Response," 2019 Fifth Conference on Mobile and Secure Services (MobiSecServ), Miami Beach, FL, USA, 2019, pp. 1-8.
- [29] S. Godwin, B. Glendenning and K. Gagneja, "Future Security of Smart Speaker and IoT Smart Home Devices," 2019 Fifth Conference on Mobile and Secure Services (MobiSecServ), Miami Beach, FL, USA, 2019, pp. 1-6.
- [30] Keely Hill, Kanwalinderjit Kaur Gagneja, Navninderjit Singh, "LoRa PHY Range Tests and Software Decoding - Physical Layer Security", 6th IEEE International Conference on Signal Processing and Integrated Networks (SPIN 2019), 7 - 8 March 2019.
- [31] Alexandro Ruiz, Carlos Machdo, Kanwal Gagneja, Navninderjit Singh, "Messaging App uses IRC Servers and any Available Channel", 6th IEEE International Conference on Signal Processing and Integrated Networks (SPIN 2019), 7 - 8 March 2019.
- [32] Iansiti, Marco, and Karim R. Lakhani. "The truth about blockchain." Harvard Business Review 95.1 (2017): 118-127.



Dr. Kanwal Gagneja is an Assistant Professor in the Department of Computer Science at Florida Polytechnic University, Florida, USA. She graduated with her PhD in Computer Science from North Dakota State University, Fargo, USA. Her research is focused on large scale Distributed Systems (P2P, Grid, Cloud), Cybersecurity, Digital Forensics, Sensor Networks, Networks Security and Semantic Web Technologies.



Andre Goode is working as a Data Engineer at Disney Streaming Services, New York, USA. He graduated from Florida Polytechnic University in 2018.



Karim Rezk is working as an IT support specialist, California, USA. He graduated from Florida Polytechnic University in 2018.



David Phillip Rentos is working as Network Administrator Trainee at Melbourne, Florida, USA. He graduated from Florida Polytechnic University in 2018.