Ad-hoc Analytical Framework of Bitcoin Investigations for Law Enforcement

Hiroki KUZUNO††, Member and Giannis TZIAKOURIS††, Nonmember

SUMMARY

Bitcoin is the leading cryptocurrency in the world with a total market cap of nearly USD 33 billion, [1] with 370,000 transactions recorded daily [2]. Pseudo-anonymous, decentralized peer-to-peer electronic cash systems such as Bitcoin have caused a paradigm shift in the way that people conduct financial transactions and purchase goods. Although cryptocurrencies enable users to securely and anonymously exchange money, they can also facilitate illegal criminal activities. Therefore, it is imperative that law enforcement agencies develop appropriate analytical processes that will allow them to identify and investigate criminal activities in the Blockchain (a distributed ledger). In this paper, INTERPOL, through the INTERPOL Global Complex for Innovation, proposes a Bitcoin analytical framework and a software system that will assist law enforcement agencies in the real-time analysis of the Blockchain while providing digital crime analysts with tracing and visualization capabilities. By doing so, it is feasible to render transactions decipherable and comprehensible for law enforcement investigators and prosecutors. The proposed solution is evaluated against three criminal case studies linked to Darknet markets, ransomware and DDoS extortion.

key words: digital forensics, cryptocurrency, investigation

1. Introduction

According to the Financial Action Task Force (FATF) [3], a cryptocurrency is a digital representation of value that can be digitally traded and functions as i) a medium of exchange; and/or ii) a unit of account; and/or iii) a store of value, but does not have legal tender status in any jurisdiction. Over the years cryptocurrencies, especially Bitcoin [4], have exponentially grown both in value and users. A number of reasons have contributed to their establishment as a suitable alternative to fiat currencies. Firstly, they assume operation in trust-less environments where transacting parties are self-interested. Secondly, they eliminate the need for a central authority that handles transactions and issues money, thus returning the control of money back to their users/owners while alleviating steep charges. Finally, the use of asymmetric cryptography and decentralized peer-to-peer (P2P) methodologies enable users to pseudo-anonymously transact without revealing their real identities.

Bitcoin is heavily grounded in the Blockchain technology, a distributed ledger that records verified transactions. The Blockchain maintains and publicly illustrates the transactions performed in the Bitcoin network. The published data only contains information related to the Bitcoin addresses of the transacting parties. This therefore provides a pseudo-anonymity to users, as third parties can only see that a transaction was performed but not the identities of the users behind the addresses. Despite the fact that Bitcoins and other cryptocurrencies (i.e. altcoins) are used daily by legitimate users to exchange money and purchase goods or services in a secure and anonymous manner, their pseudo-anonymous and decentralized nature also make them a powerful tool for criminals. An increasing number of criminals are establishing illicit markets in the Darknet, such as AlphaBay [6], and using cryptocurrencies as their digital financial instruments for receiving payments while concealing their identities and evading prosecution. Bitcoin is unquestionably the preferred digital currency in the Darknet, proven by its extensive use in extortion cases (i.e. ransomware), money laundering, terrorism financing and other illicit activities.

It is imperative that law enforcement agencies are able to identify and thwart online criminal activities that are linked to cryptocurrencies, including Bitcoin. Significant work exists on the analysis of the Blockchain. The existing work has mainly focused on the analysis of Bitcoin transactions, and in particular the identification and analysis of the various relationships between Bitcoin addresses [7]–[10]. The existing research has resulted in a number of commercial websites and software tools that are able to index and analyse Bitcoin transactions and addresses [11]–[15]. Although these solutions provide a significant insight into the Bitcoin ecosystem, including the behaviour and patterns of their users, they are not built with law enforcement in mind. Therefore, they are not able to produce evidence which can be used in court cases as they do not comply with the established investigation policies of most countries and are not certified for recording electronic evidence which is of the utmost importance for law enforcement and policymakers [16]. Another aspect limiting existing solutions and their usage by law enforcement agencies is their high cost. As police organizations often operate with strict monetary and legal restrictions (for example as relates to sharing sensitive information) they need access to certified, low-cost forensic solutions which adhere to their explicit needs.

In this work, we propose a Bitcoin analytical process and software system that can be used by law enforcement agencies for investigating the relationship between
Blockchain transactions and addresses. The proposed analytical framework is driven by previous and ongoing cybercrime investigations, which enable us to better comprehend the functional requirements and tools needed by law enforcement to effectively investigate criminal activities in the Blockchain. The proposed system facilitates a combination of analytical methodologies that provide: i) statistical information concerning the transaction history of Bitcoin addresses, ii) graphical representation of the relations between different transactions and addresses, iii) discovery of transaction paths, and iv) clustering of known Bitcoin addresses into distinctive groups (i.e. wallets). Our work makes the following contributions to the field of Blockchain analytics:

- Proposes an ad-hoc analytical process tailored to the investigative requirements of law enforcement agencies for the analysis of criminal activity in the Blockchain.
- Delivers a cost-effective system that law enforcement agencies in INTERPOL’s member countries can use when investigating suspicious Bitcoin addresses to extract comprehensible and meaningful data that can be presented in court for prosecution.

To assert the effectiveness of the proposed solution, we have evaluated it in three different case studies related to Darknet marketplaces, ransomware and DDoS extortion. More specifically, we have used our system to analyze more than 117 million Bitcoin addresses gathered from 2009 to 2015 and examined the usefulness of the reported data and outcomes in assisting forensic investigators to provide attribution and identify those responsible for crimes relating to cryptocurrencies.

2. **Bitcoin Overview**

Bitcoin is a peer-to-peer electronic cash system founded on the principles of public key cryptography and decentralization. Bitcoins can be used similarly to conventional fiat currencies to purchase items and services electronically. The Bitcoin network assumes operation in an untrusted environment with self-interested users that need to control and protect their digital accounts, called ‘wallets’. The main characteristic of Bitcoin is its potential for pseudo-anonymity. All users in the Bitcoin network are able to exchange Bitcoins while keeping their identities concealed. The only information that is publicly shared, as part of the Blockchain, are the Bitcoin addresses of the transacting users. Each user can generate a different Bitcoin address for each transaction, making it more difficult for these transactions to be traced back to them. In the occasion where a user uses the same address for multiple transactions, it is easy to identify and track its overall activity in the network. Such an example is the use of static Bitcoin addresses by illegal vendors in Darknet markets, ransomware and DDoS extortion cases where criminals and hackers want to receive payment from their clients or victims.

Each user maintains an up-to-date copy of the Blockchain, which contains all the transactions performed in the Bitcoin network. The Blockchain is a sort of shared public ledger in which verified Bitcoin transactions are recorded in containers called blocks. Each block consists of a hash of the previous block (used as a reference/link in the chain), a list of unverified transactions and a random integer value (called a nonce) that is used to produce the proof-of-work [17] for verifying the transactions in the block. Each transaction contains the sender’s signature, the public key of the next owner, the hash value of the previous transaction, and the public key of the next owner which is signed by the private key of the issuer to ensure the payment.

To verify new transactions, miners (users that verify Bitcoin transactions and generate new Bitcoins) cluster all the unverified transactions in a single block and verify them by implementing a proof-of-work. This is achieved by incrementing the nonce value in the block and then hashing the whole block until the generated hash contains a specific number of zero bits in the prefix of the hash (number of zero bits vary based on the number of miners in the network). Once a proof-of-work is found, the block is verified and it is added to the Blockchain.

In order for a user to transfer Bitcoins it is necessary to use a Bitcoin client to be part of the network. The client maintains an e-wallet which records the Bitcoin addresses and transactions of a user. To better illustrate how a user can forward Bitcoins, we use the following example (Fig. 2): Alice wants to send Bitcoins to Bob, where Carol plays the role of a miner. To initiate the payment, Alice generates a new transaction in which she specifies the amount of Bitcoins that she wishes to send to Bob, as well as Bob’s Bitcoin address, and submits the transaction to the network for verification. In order for Bob to access the transferred Bitcoins, he needs to use its private key to sign the transaction. As the transaction is not verified, Carol picks up this transaction and tries to verify it by implementing a proof-of-work for
the whole block of transactions. In the case that Carol is able to solve the challenge and generate an appropriate proof-of-work, Alice’s transaction will be verified and added to the Blockchain.

3. Approach

This section introduces our Bitcoin analytical framework and examines the functionalities of the engineered software system. We provide an in-depth examination of the following functionalities: i) transaction analysis of Bitcoin addresses, ii) activity profiling of Bitcoin addresses (statistical analysis), iii) correlation analysis and mapping between Bitcoin addresses, and iv) clustering of Bitcoin addresses to a single wallet.

The objective is to provide a practical analytical process that will assist law enforcement agencies and digital crime analysts when investigating Bitcoin addresses, by extracting useful and comprehensible information that can be used for prosecuting criminals.

3.1 Bitcoin Analytical Framework

The proposed framework consists of three components: an indexer, an analysis module and a web interface (Fig. 3). The main functionality of the indexer is to record Bitcoin addresses and transactions, at runtime, and label them as suspicious, if applicable. Bitcoin addresses are labeled as suspicious in the cases that have already been reported by law enforcement agencies. To determine if a Bitcoin address is involved in illegal activities, the proposed indexer uses a web-crawler to automatically search the clearnet and Darknet for incriminating e-evidence. Additionally to the web-crawler, our system supports the manual enhancement of Bitcoin addresses by allowing investigators to label them based on relevant information gathered from the international law enforcement community that can lead to attribution (i.e. multi signature) [18]. To index the Blockchain, we use a Bitcoin client such as Bitcoind [19] to join the Bitcoin network and download the latest transactions in a local database. Keeping the database updated is critical, as any missing information could jeopardize the outcome of an investigation. The analysis module is used to identify the transaction history and behavioral information of a targeted Bitcoin address. The results from the analysis module can be classified under four different groups: i) statistical information describing the activity of an address (number of transactions, active months), ii) a graphical representation of the relations between various transactions and Bitcoin addresses, iii) the transaction paths for each Bitcoin address and iv) a cluster containing Bitcoin addresses that belong to the same wallet. Finally, our framework facilitates a web interface that allows users to easily access the underlying mechanisms of the proposed system.

A user can investigate a targeted Bitcoin address by providing different types of information to our system, such as the Bitcoin address, the amount of Bitcoins exchanged or the date that a transaction occurred. By doing so, our system can assist law enforcement agencies in identifying possible associations between targeted Bitcoin addresses. It can also provide estimations concerning the likelihood of a Bitcoin address being involved to criminal activities based on information provided by law enforcement agencies around the world as well as open-intelligence data collection from the clearnet and Darknet. Finally, our tool can cluster Bitcoin addresses into different groups based on the types of criminal activities that were involved in, hence assisting law enforcement agencies to identify the key criminals involved
in each type of crime (money laundering, financing terrorism) and eventually helping them to link these addresses to specific wallets for attribution. The process followed by our analytical framework is illustrated in Fig. 4. Although our framework can support Blockchain investigations and provide valuable information concerning Bitcoin addresses, it is still necessary for INTERPOL member countries to conduct additional investigations to verify the explicit links of a Bitcoin address to criminal activities.

In the following sub-sections, we provide an in-depth analysis of the main functionalities of the proposed analytical framework.

### 3.2 Indexing and Searching

Ensuring that Bitcoin investigations are based on up-to-date data is crucial for law enforcement agencies. Although substantial work exists on indexing the Blockchain [13], [14], current solutions typically share their information with third parties which is not permitted in the law enforcement context due to the classified nature of investigations.

To ensure that INTERPOL’s local investigation database is constantly updated with the latest Bitcoin addresses and transactions, we have developed a real-time monitoring process that identifies and archives new addresses and transactions. Our monitoring process uses the Bitcoind client (a command-line interface for Bitcoin that can pass transactions. Our monitoring process uses the Bitcoind client (a monitoring process that identifies and archives new addresses and transactions, we have developed a real-time database is constantly updated with the latest Bitcoin addresses and Bitcoin values) (lines 7 – 18).

Once Bitcoin transactions are indexed, a user can search for specific or targeted Bitcoin addresses. To search our database $DB$ (Algorithm 2) a user or investigator needs to provide a Bitcoin address $a_1$ to retrieve its transaction records $R$. It is possible for a user to use different input data to acquire information for addresses of interest, such as: i) Bitcoin addresses, ii) block and transaction hash value, iii) Bitcoin transaction value, iv) fiat amount (i.e. US dollar, Fig. 5) and/or v) specific dates. By doing this, we allow law enforcement investigators to use a variety of attributes when exploring the Blockchain. This can prove critical in Bitcoin investigations where law enforcement officers possess different, and often limited, information concerning the criminal activities they are investigating.

The information returned to users comprises the entry records $(a_i, v_i, id, type) ∈ DB$ that contain or match the provided Bitcoin address $a_1$ (lines 2 – 6).

The entry records returned include statistical information describing the number of transactions performed, the balance history, the first/last seen dates, the transaction

---

**Algorithm 1** Indexing of Blockchain

**Input:** $B = \{b_1, \ldots, b_n\}$, a set of blocks in the Blockchain  
**Output:** $DB$, the indexed Database  

- $b = \{T_1, \ldots, T_n\}$, a set of transactions in a block  
- $T_i = (A_{src}, A_{dst}, id)$, a group of source/destination addresses with transaction id  
- $A_{src} = \{(a_1, v_1), \ldots, (a_m, v_m)\}$, a set of source Bitcoin addresses  
- $A_{dst} = \{(a_1, v_1), \ldots, (a_n, v_n)\}$, a set of destination Bitcoin addresses  

1: for all $b_i ∈ B$ do  
2: for all $T_i ∈ b_i$ do  
3: call `IndexingTransaction(T_i)`  
4: end for  
5: end for  

**Algorithm 2** Searching of Indexed Database

**Input:** $a_1$, a bitcoin address  
**Output:** $R = \{r_1, \ldots, r_n\}$, a transaction record set of $a_1$  

- $DB = \{(a_1, v_1, id, type), \ldots, (a_n, v_n, id, type)\}$ is Indexed Database  
- $type = "src"$ or "$dst$"  

1: $R \leftarrow \emptyset$, an empty set  
2: for all $(a_i, v_i, id, type) ∈ DB$ do  
3: if $a_i = a_1$ then  
4: append $(a_i, v_i, id, type)$ to $R$  
5: end if  
6: end for
Fig. 5 A sample result illustrating the information returned when searching for Bitcoin addresses by USD value. The results show 44 transactions between $10 and $20, dated from 01/09/2010 to 09/09/2010.

Fig. 6 A sample result from querying a Bitcoin address.

Fig. 7 A sample result illustrating statistical information for transactions with their timestamps.

Fig. 8 A sample result of a calendar view from querying a Bitcoin address.

amounts, tagged information (added by law enforcement agencies from the analysis of open-source information gathered from clearNet and Deep Web as well as other relevant data describing the types of the Bitcoin addresses via the use of appropriate string prefixes (i.e. multi signature)) (Fig. 6). Furthermore, investigators can also access information related to the weekly or hourly transaction patterns (Fig. 7) of a targeted Bitcoin address, which can be presented in a calendar view (Fig. 8). By providing this information to law enforcement agencies, it is possible for digital forensic analysts to better identify criminal activity in the Blockchain while allowing them to better estimate the number of victims involved in a crime based on the number of transactions, as well as the time zone that a criminal is operating in based on the time of the transactions.

3.3 Visualizing and Path Finding

To provide law enforcement investigators with meaningful and comprehensible data and e-evidence, the proposed system uses a visualizer that graphically illustrates the various relationships between Bitcoin addresses (Fig. 9). More specifically, our system visualizes the outgoing and ingoing transactions from and to Bitcoin addresses which are repre-
Fig. 9 Illustrates the relationships between a Bitcoin address and other addresses in the network.

Algorithm 3 Visualizing Bitcoin Address’s Transactions

Input: \(a_1\), a bitcoin address
Output: \(IN = \{(a_1, v_1), \ldots, (a_n, v_n)\}\), a bitcoin address set of \(a_1\)
\(OUT = \{(a_2, v_2), \ldots, (a_m, v_m)\}\), a output bitcoin address set of \(a_1\)
\(T_i = (A_{src}, A_{dst}, ID)\) is a group of src/dst addresses with transaction id
\(A_{src} = \{(a_1, v_1), \ldots, (a_n, v_n)\}\) is a set of src address with bitcoin
\(A_{dst} = \{(a_2, v_2), \ldots, (a_m, v_m)\}\) is a set of dst address with bitcoin
1: \(T(a_1)\), a function returns set of transaction \([T_1, \ldots, T_n]\)
2: \(IN \leftarrow \{\}\), an empty set
3: \(OUT \leftarrow \{\}\), an empty set
4: 
5: for all \(T_i \in T(a_1)\) do
6: \(A_{src} \leftarrow T_i\)
7: \(A_{dst} \leftarrow T_i\)
8: for all \((a_j, v_j) \in A_{src}\) do
9: \(\{\text{if } a_i \in IN \text{ then}\}
10: \(a_i, v_j \leftarrow IN\)
11: \(\text{update } (a_i, v_j + v_i) \text{ to } IN\)
12: \(\text{else}\)
13: \(\text{append } (a_i, v_i) \text{ to } IN\)
14: \(\text{end if}\)
15: \(\text{end for}\)
16: for all \((a_j, v_j) \in A_{dst}\) do
17: \(\{\text{if } a_i \in OUT \text{ then}\}
18: \(a_i, v_j \leftarrow OUT\)
19: \(\text{update } (a_i, v_j + v_i) \text{ to } OUT\)
20: \(\text{else}\)
21: \(\text{append } (a_i, v_i) \text{ to } OUT\)
22: \(\text{end if}\)
23: \(\text{end for}\)
24: 

addresses \(A_{src}\) and their \((a_i, v_i)\) data, where \(a_i\) denotes a Bitcoin address and \(v_i\) a Bitcoin value. If the examined address \(a_i\) already exists in our records, meaning it has sent Bitcoins to the targeted address before, we update or append the Bitcoin value and transaction reference for that address \(v_j\) to \(v_i\) of \(a_i\). If \(a_i\) is a new, unrecorded address sending Bitcoins to the target address for the first time, a new reference or link is created and the associated data is appended (lines 8 – 15). Similarly, we follow the same procedure to graphically represent the extracted data from the destination addresses \((a_j, v_j) \in A_{dst}\) (lines 16 – 23).

Given a Bitcoin address \(a\), we define address transactions as:

\[G_{in}, G_{out} = \sum_{t_i \in T(a)} \text{VisualizingTransaction}(t_i).\]

where \(G_{in}\) is a graph containing the inbound transactions of \(a\) and \(G_{out}\) a graph containing the outbound transactions of \(a\). Both graphs consist of vertexes and edges, where vertexes represent Bitcoin addresses and edges represent transactions. It is possible to retrieve additional information concerning the transactions made by a Bitcoin address by selecting a vertex in the graph. By doing so, our system will identify all the transactions associated with that vertex, generate a new graph and merge it with the existing graph. Furthermore, each vertex is associated with a hyperlink that contains the statistical information of each Bitcoin address.

Added to the graphical representation of Bitcoin addresses and their transactions, the proposed system can demonstrate the transaction path for transactions of interest. The goal is to create a transaction flow between the initial sender and the end receiver to identify all the intermediate addresses used, in order to identify possible third parties that are involved in a criminal activity. To identify a transaction path, our system (Algorithm 4) takes two Bitcoin address \(a_1, a_2\) as an input. It then iteratively accesses all the transactions \(t_i \in T(a_j)\) which originate from the provided source Bitcoin address and presents all the destination Bitcoin addresses recorded in these transactions and links them together, when applicable, thus resulting to a transaction path \(p\) between \((a_i, a_n)\) (lines 12 – 15).

Given Bitcoin address \(a_i\) and \(a_j\), we define path finding as:

\[\text{PathFinding}(a_i, a_j) = \begin{cases} \text{false, if path is not found} \\ P, \text{ if } a_j = a_i \\ \sum_{t_i \in T(a_i)} \sum_{a_k \in T(a_i)} \text{PathFinding}(a_k, a_j) \end{cases}\]

where \(T\) is a function that links transactions to Bitcoin addresses \(a, t\) is a transaction containing a set of Bitcoin addresses and \(P\) is the path of visited Bitcoin address for a specific transaction. To reduce the computational overhead of our algorithm and improve the readability of its outcomes, we allow digital crime investigators to set a threshold for
Algorithm 4 Path Finding between Bitcoin Addresses

Input: $a_1, a_2$ are bitcoin addresses
Output: $P = \{(a_1, a_2), \ldots, (a_m, a_2)\}$ is path to $a_2$ from $a_1$
1: $T(a_1)$, a function returns set of transaction $[t_1, \ldots, t_n]$ (i.e., $t = [a_1, \ldots, a_n]$, a set of bitcoin address
2: $l$, a loop guard is non negative value
3: $m \leftarrow 0$
4: $P \leftarrow \{}$
5: $\text{PathFinding}(a_1, a_2)$
6: function PathFinding($a_i, a_j$)
7: \begin{align*}
8: & \quad \text{if } m \geq l \text{ then } \\
9: & \quad \quad \text{return } false \\
10: & \quad \text{end if} \\
11: & \quad \text{for all } t_i \in T(a_i) \text{ do } \\
12: & \quad \quad \text{for all } a_n \in t_i \text{ do } \\
13: & \quad \quad \quad \quad p \leftarrow (a_i, a_n) \\
14: & \quad \quad \quad \quad \text{append } p \text{ to } P \\
15: & \quad \quad \text{if } a_n = a_j \text{ then } \\
16: & \quad \quad \quad \quad \text{return } P \\
17: & \quad \quad \text{end if} \\
18: & \quad m \leftarrow m + 1 \\
19: & \quad \text{return call PathFinding(a_n, a_j)} \\
20: & \end{for} \\
21: & \text{return } false \\
22: & \end{for} \\
23: & \text{return } false \\
24: & \end{function}

Fig. 10 Illustrates the transaction path for Bitcoin addresses.

Fig. 11 A sample result illustrating a clustered four Bitcoin addresses belong to a single wallet.

3.4 Clustering

Each Bitcoin client maintains a wallet which is used as a local database for storing Bitcoin addresses. To generate a Bitcoin address, a series of tasks are performed by each client. First, a private key (k) is created which consists of 256 randomly selected bits (64 characters, each 4 bits). Next, a public key (K) is derived from the private key with the assistance of a one-way cryptographic function called elliptic curve multiplication [20]. Finally, to generate a Bitcoin address (bA), it is necessary to create a one-way cryptographic hash function of the public key. To do so, it is necessary to compute a SHA256 hash and then compute a RIPEMD160 hash of the result, generating a 20-byte number address $bA = \text{RIPEMD160(SHA256(K))}$. To make the Bitcoin addresses easily comprehensible, as well as to determine if there are any errors in the address, the hash is encoded with Base58Check which uses a base-58 number system and a checksum. The generated Bitcoin addresses can then be used to forward and/or receive Bitcoins. For each transaction, a user can create a different set of private-public keys and subsequently Bitcoin addresses. In order for a user to transfer Bitcoins, it is necessary to specify the destination Bitcoin address as well as to verify the payment by signing the transaction with its private key once the transaction is carried out.

As the only information publicly available to the Bitcoin network is the Bitcoin addresses involved in a transaction, tracing these addresses back to their owners and clustering them into their respective wallets is a difficult undertaking. To assist law enforcement agencies to link Bitcoin addresses to their wallets and ultimately to their users, our system dynamically identifies the instances and transactions where a targeted address have been involved. All the identified transactions are then clustered into a single wallet (Fig. 11).

Our clustering algorithm (Algorithm 5) takes a set of Bitcoin addresses $A$ as an input and outputs $C$ a set of clusters, each containing Bitcoin addresses belonging to same
Algorithm 5 Clustering Bitcoin Addresses

Input: $A = \{a_1, \ldots, a_n\}$, a set of bitcoin address
Output: $C = \{c_1, \ldots, c_m\}$, a set of cluster

1. $T(a_i)$ function associating transaction id per $a_i$
2. $G \leftarrow \{\}$, an empty set
3. for all $a_i \in A$ do
    4. $g_i \leftarrow \{a_i\}$
    5. for all $a_j \in A$ do
        6. if $j \neq i$ then
            7. call FindCommonTransaction($a_i, a_j, g_i$)
        8. end if
    9. end for
10. append $g_i$ to $G$
11. end for
12. end for
13. for all $g_i \in G$ do
    14. $c_i \leftarrow g_i$
    15. for all $g_j \in G$ do
        16. if $j \neq i$ then
            17. call MergeGroup($c_i, g_j$)
        18. end if
    19. end for
20. append $c_i$ to $C$
21. end for
22. end for
23. function FindCommonTransaction($a_i, a_j, g_i$)
24. for all $t_k \in T(a_i)$ do
25.     for all $t_l \in T(a_j)$ do
26.         if $t_k = t_l$ then
27.             append $a_j$ to $g_i$
28.         goto finish
29.     end if
30. end for
31. end for
32. finish:
33. end function
34. function MergeGroup($c_i, g_j$)
35. for all $a_m \in g_j$ do
36.     if $a_m \in c_i$ then
37.         for all $a_n \in g_j$ do
38.             if $a_m \neq a_n$ then
39.                 append $a_n$ to $c_i$
40.             end if
41.         end if
42.     end if
43. end for
44. break
45. end if
46. end for
47. end function

our clustering algorithm has two main phases. The first phase (i.e. FindingSameTransaction) aims to iteratively check whether Bitcoin addresses have remitted BTC in a single transaction. Each Bitcoin address $a_i \in A$ is appended to a group $g_i$. We then retrieve other Bitcoin address $a_j \in A$ and try to identify whether there is a common transaction between these addresses. If a common transaction is found, group $g_i$ is created and recorded as a subgroup in group $G$ (lines 4 – 12).

The second phase (i.e. MergeGroup) aims to identify whether different groups share common Bitcoin address. If different groups share the same addresses, they are then assigned to a single cluster. In particular, we access each group $g_i \in G$ and assign it to a cluster $c_i$, if any other groups $g_j \in G$ share any of their Bitcoin addresses they are also clustered under $c_i$. Once the iterative process finishes, a cluster $c_i$ is appended to $C$ which contains all the identified wallets/clusters (lines 14 – 22).

Given a Bitcoin address list $A$, we define our clustering process as:

$G = \sum_{a_i \in A} \sum_{a_j \in A, i \neq j} \text{FindingSameTransaction}(a_i, a_j)$

$C = \sum_{g_i \in G} \sum_{g_j \in G, i \neq j} \text{MergeGroup}(g_i, g_j)$

where $G$ is a set of Bitcoin addresses that share an identical transaction and $C$ denotes a Bitcoin cluster that maintains the addresses belonging to a single wallet.

The proposed clustering approach is not only restricted to the identification of Bitcoin addresses that are directly linked to specific transactions, but also to Bitcoin addresses that are implicitly linked via intermediate or other related transactions. To better demonstrate this, consider the example where Bitcoin address $A$ and Bitcoin address $B$ remit BTC to Bitcoin address $X$ in a single transaction ($z$) and Bitcoin address $B$ and Bitcoin address $C$ remit BTC to Bitcoin address $K$ in a single transaction ($r$). Our system is able to infer that since Bitcoin addresses $A$ and $B$ were used as input Bitcoin addresses in transaction $z$, from a single user, they belong to the same wallet. Similarly, as Bitcoin addresses $B$ and $C$ have been used as input addresses for transaction $r$ they also belong to the same wallet. However, what is not easily identifiable is the relation between Bitcoin addresses $A$ and $C$. As both transactions $z$ and $r$ have Bitcoin address $B$ in common, also Bitcoin addresses $A$ and $C$ must belong to the same wallet (Fig. 12).

By providing clustering information concerning targeted Bitcoin addresses, we enable law enforcement agencies and digital crime analysts to better identify and estimate the illegal profits of criminals as well as the type and extent of their operations, thus providing better attribution.
4. Evaluation and Discussion

In this section, we provide our experimental setup and present our findings concerning three Bitcoin investigation case studies related to Darknet markets, ransomware and extortion cases. In particular, this section aims to determine if the analytical features of the proposed solution can provide digital crime analysts with substantial capabilities to effectively investigate criminals in the Blockchain.

4.1 Experimental Setup

To conduct our experiments, we have recorded a large number of Bitcoin addresses and transactions in INTERPOL’s local database - 57,723,195 unique Bitcoin addresses were recorded between 2009 and 2015. Noting that the number of Bitcoin addresses in 2015 (57,723,195 Bitcoin addresses) were 1,700 times as many as in 2009 (32,611 Bitcoin addresses) (Fig. 13). It is exponentially more complex to analyse the transaction history of a Bitcoin address due to the increased number of relations between addresses.

4.2 Bitcoin Addresses Investigation

To determine the effectiveness of our tool, we have revisited three well-known cybercrime case studies (Silk Road market, CryptoLocker ransomware and DD4BC extortion) and compared our findings with the forensic evidence previously reported from law enforcement agencies involved [15], [21], [22].

Due to the varying nature of these cases, different Bitcoin investigation techniques were required. Throughout our experiments, we observed that our tool could reveal critical information to law enforcement investigators which could ultimately lead to prosecutions. The added value of our tool is summarized in Fig. 14, illustrating for each of the examined case studies the information initially available to law enforcement agencies before using our tool, and the additional information or e-evidence acquired by using our solution.

4.2.1 Silk Road Market

The Silk Road market was one of the biggest Darknet marketplaces in the history of the Dark Web, with a large number of vendors and buyers trading more than 10,000 different products, of which 70% were narcotics. Due to the massive attention that the Silk Road market acquired, it became a compelling target for law enforcement and intelligence agencies. As a result, in October 2013 the US Federal Bureau of Investigation (FBI) shut down the website and arrested its administrator. Tracing the Bitcoin addresses of Silk Road vendors and exposing their real identities was a complex undertaking due to the pseudo-anonymous nature of cryptocurrencies in conjunction with the anonymizers (i.e. Tor) used by vendors for accessing the Silk Road and Bitcoin exchange marketplaces. To investigate these addresses, law enforcement investigators need to first identify and gather the Bitcoin addresses associated with the Silk Road market. To do this, digital crime analysts use web crawling tools to conduct an in-depth analysis and indexing of the contents of the illicit markets for identifying possible occurrences of criminal Bitcoin addresses. Once the targeted addresses are acquired, the Blockchain analytics phase can be initiated for the extraction of information that can support prosecutions.

To demonstrate how our tool, in conjunction with open-source investigation, can assist law enforcement officers to identify the criminals behind Bitcoin addresses, we apply our investigation process on a selected Bitcoin address from the Silk Road market: ‘1LDNLr’\(^1\). By using a web-crawler it was possible to identify critical information concerning the Bitcoin address in question, including a post containing an e-mail address linking it to the popular Bitcoin forum bitcointalk.org, which provided us with additional information.

\(^1\)1LDNLreKJ6GawBHPgB5yfVLBERi8g3SbQS
on the operations of the owner of the Bitcoin address. Once we gathered sufficient information verifying the involvement of the selected Bitcoin address with criminal activities, we then conducted a Blockchain analysis, using the tool to trace the transaction history and behaviour of the selected address. It was able to observe that ‘1LDNLr’ has repeatedly made remittances to the address ‘1933ph’, indicating an explicit transaction path between the two addresses, with ‘1933ph’ receiving a total of 111,111 BTC from ‘1LDNLr’ (Fig. 15 shows the transaction history of “1933ph”). The analytical process has indicated a close relationship between the two Bitcoin addresses, ‘1LDNLr’ and ‘1933ph’, demonstrating that both of them belonged to a single owner. Due to existing reported information for Bitcoin address ‘1933ph’, it was feasible to associate this information with the Bitcoin address ‘1LDNLr’ and thus reveal the identity of the owner.

4.2.2 CryptoLocker Ransomware

The CryptoLocker ransomware attack occurred from September 2013 to May 2014. It was propagated via infected e-mail attachments as well as via the Gameover ZeuS botnet. Once activated, CryptoLocker encrypted certain types of files stored on local and mounted network drives using RSA public-key cryptography, with the private key (de-encryption key) stored only on the malware’s control servers. To decrypt the disks, CryptoLocker requested from its victims a payment of 2.0 BTC to a specific Bitcoin address.

Our first investigative step was to estimate the number of CryptoLocker victims, as well as the amount of money received by the criminals. To this end, we used our tool to search for transactions of 2.0 BTC occurring during the timeframe of the attack. Our search query returned a long list of Bitcoin address and transactions, of which many were not relevant to this investigation. However, it was possible to identify that two of the returned addresses were likely involved in the CryptoLocker attack, as the volume of their incoming transactions was irregularly high with a large number of Bitcoin addresses remitting to them 2.0 BTC (Fig. 16 shows a large number of incoming transactions for Bitcoin addresses ‘1KP72f’ and ‘18iEz6’).

Once we identified the suspect addresses, we needed to further trace their activities by examining their transaction histories, with a view to verifying their explicit link to the CryptoLocker attack. Our findings demonstrated an explicit relation between the attack and these addresses due to their timeline of operations, the high volume and frequency of incoming transactions, the payment amounts of 2 BTC, as well as their appearance in various Darknet forums related to the engineering and deployment of malware binaries. The identification of these addresses allowed us to trace their transaction paths and to observe that the ransomware author/owner forwarded the Bitcoins received to another Bitcoin address, and then used this address in a well-known cryptocurrency exchange market to exchange them with a different cryptocurrency, thus increasing their untraceability. This type of information can serve as a critical lead in an investigation, as a forensic investigator can contact a cryptocurrency exchange market and request additional information for the transactions in question which can ultimately lead to the identification of their owners (this practice is subject to the jurisdiction and privacy policies of these markets).

4.2.3 DD4BC Extortion

DD4BC (DDoS “4” Bitcoin) is considered to be one of the biggest extortion incidents, with more than 140 companies affected in 2014. The criminal group was extorting companies by sending e-mails to their victims warning of a distributed denial-of-service (DDoS) attack capable of 400-500 Gbps unless they agreed to pay a certain amount of Bitcoins.
The relation path between ‘1FsVcd’ and ‘3Bgk1o’. Demonstrates a mixing / tumbling process. One of the generated flows is associated with ‘3Bgk1o’.

within a set time limit. To demonstrate the effects of their attacks and test the defenses of their victims, DD4BC simultaneously launched a ‘warning’ DDoS attack of around 10-15 Gbps.

For this case, we initiated our investigation using the two Bitcoin addresses provided by the DD4BC group to their victims. The objective of this investigation is to identify the addresses involved in this extortion case as well as to estimate the amount of Bitcoins received by the extortionists. DD4BC group used the following Bitcoin address: ‘1H2bst’ and ‘17WQov’. By using the visualization feature of our system, we were able to determine that both of the Bitcoin addresses had an explicit link to a third address ‘1FsVcd’. In particular, three transactions occurred between ‘1H2bst’ and ‘1FsVcd’, and 14 transactions from ‘17WQov’ to ‘1FsVcd’. We then used our path finder to identify the Bitcoin flow of the transactions initiated by the initial Bitcoin addresses, which revealed that all the Bitcoins in ‘1FsVcd’ had been forwarded to two new Bitcoin addresses: ‘3Bgk1o’ and ‘3Q5r9g’ (Fig. 17 shows the Bitcoin flow between ‘1FsVcd’ and ‘3Bgk1o’). Additionally, we were able to observe that (Fig. 18) Bitcoin address ‘3Bgk1o’ had sent money to an extensive number of Bitcoin addresses, with some of them recurring transactions at different times which is a behaviour usually exhibited by mixers or tumblers. Tumblers are used by criminals to enhance their anonymity by mixing potentially identifiable Bitcoins with others to mask the trail back to the original source. Finally, we were able to identify that ‘3Q5r9g’ had forwarded the laundered Bitcoins to another address and then exchanged them with another cryptocurrency (Fig. 19).

5. Related Work

Extensive work has been conducted on the analysis of the Bitcoin network [23]–[25]. This work has resulted in a number of commercial and open-source tools. Despite their usefulness, the forensic results produced by the open-source tools are often difficult to interpret, as they are intended for subject matter expert users. Furthermore, they do not facilitate advance functionalities, such as address clustering and Bitcoin web-crawlers, which can be critical for Bitcoin police investigations. Commercial Bitcoin tools can be quite expensive, often out of the reach of the budgets of many police agencies. Lastly, the existing solutions have
not been designed with policing in mind and the explicit requirements concerning the collection and presentation of evidence which can be admissible in court.

One of the most well known open-source tools for analysing the Blockchain is Blockchain.info [2] (similar web-tools: BitcoinChain.com [26] and Blocktrail.com [27]), which provides users with basic information concerning market capitalization, transaction volume, hash rate and Bitcoin transactions. Blockchain.info allow users to search for specific Bitcoin addresses to trace the origin of these coins, as well as how much BTC is stored at any particular wallet address.

Another well-known open-source tool is bitcoin-charts.com [28], which offers a technical analysis of the price trends of cryptocurrencies. The site provides users with a number of customizable charts illustrating price trends.

Bitnodes [29] estimates and visualizes the size of the Bitcoin network on a world map. To achieve this, the tool gathers data from all Bitcoin nodes that run the Bitcoin protocol version 70001.

WalletExplorer.com [12] is an online Bitcoin analytics tool which analyses and clusters Bitcoin addresses in real time based on a heuristic clustering method. The main goal of the solution is to gather and visualize statistical information concerning the reuse of Bitcoin addresses in the Blockchain, revealing problematic services within the Bitcoin ecosystem and recording progress over time.

Spagnuolo M. et al., proposed Bitiodine [15], a modular framework that parses the Blockchain and groups Bitcoin addresses that may belong to the same user or wallet. It also provides end-user with visualization capabilities for illustrating the extracted information from the Bitcoin network. Bitiodine clusters users based on open-source crawled information concerning their identity or actions.

Some notable commercial Bitcoin analytics tools are the products of Chainalysis [11] and Eliptic [30], which provide users with a variety of features for the analysis of Bitcoin addresses and transactions. They facilitate different visualizations for illustrating the various relationships between Bitcoin addresses, along with detailed forensic information concerning each transaction. They also use clustering algorithms for grouping multiple addresses under a single wallet.

Ron D. et al., [8] propose a Bitcoin transaction analytics tool that analyses the activity of Bitcoin addresses. The proposed tool supports the generation of a large graph illustrating the Bitcoin transactions, as well as a clustering algorithm that demonstrates the Bitcoin addresses that belong to a single user or wallet.

Moser M. et al., [31] propose a methodology for tracing and tracking Bitcoin address after anonymized transactions. The authors have traced the transactions of three different mixing services (i.e. Bitcoin Fog, Blockchain.info and BitLaundry), via the use of heuristic clustering. This allows them to uncover relationships between various Bitcoin services and narrow down the list of potential suspicious addresses linked to money laundering or ransomware [7–9, 22, 31, 32].

5.1 Comparison with the State-of-the-Art Solutions

To better illustrate how the proposed solution advances the current state-of-the-art in Bitcoin investigation tools (in the context of policing) we have made a comparison between the analytical features of our solution and three other commercial, open-source solutions [2], [12], [15] that are often used by law enforcement for Bitcoin investigations (Table 1). The reasoning for omitting considering paid-to-use tools in our comparison is due to their high-price which makes them unappealing for the vast majority of the INTERPOL member countries.

The performed comparison demonstrated that the only tool satisfying all the (law enforcement) identified requirements for Bitcoin investigations is the proposed solution. Although, Bitiodine [15] facilitates visualization capabilities for illustrating the relations between Bitcoin addresses and/or clusters, however it is difficult to trace/follow the transaction flows of specific Bitcoin addresses which is crucial for law enforcement. To bypass this limitation our system implements a more interactive methodology that supports an in-depth capability for investigating specific Bitcoin addresses and their relationship with other addresses.

Finally, despite that software tools such as the WalletExplorer.com [12] support a basic level of Bitcoin clustering which can assist in attribution, the use of these tools are still a challenge for the law enforcement community. This arises due to their open-source nature which do not facilitate appropriate security mechanisms (i.e. user logging tools, etc.) to warrant the secure and confidential handling of classified data that are often collected and used by law enforcement agencies. Thus, rendering law enforcement agencies unable to upload their classified Bitcoin data (coming from various sources such as, clearNet, Deep Web crawling, seized
data from suspects, etc.). Forensic investigators necessitate “air-gapped” software tools that are closely monitored to warrant the integrity of the investigation process itself and the handled data which is currently not considered by any of the related work examined.

6. Conclusions

Bitcoin can be perceived as a paradigm shift in the e-payment ecosystem. The pseudo-anonymous nature of this solution has allowed users across the globe to purchase items and exchange money in a safe and anonymous manner. However, this technology has also attracted significant attention from criminals, who exploit it for their illicit purposes. In order for law enforcement agencies to be able to conduct in-depth investigations in the Blockchain, it is essential that ad-hoc forensic tools are developed for Bitcoin analytics. These tools need to adhere to the explicit requirements (laws, functionalities required, tool verifications) and constraints of law enforcement agencies to make them suitable for e-evidence collection and prosecutions.

In this work, we have proposed an ad-hoc Bitcoin analytical process and software system for the analysis of Bitcoin transactions and addresses. Our work is driven by real-life police case studies as well as the explicit requirements and constraints of law enforcement agencies around the world. We have reported on the effectiveness and applicability of our solution by applying it to three real-world case studies on Darknet markets, ransomware attacks and DDoS extortion. We have shown that our tool can assist digital crime analysts and law enforcement agencies to investigate the Blockchain and combat cybercrime with the extraction of critical and coherent information that can be used in the court system for prosecuting criminals. Potential future work from INTERPOL’s Innovation Centre is to initiate the testing phase of the proposed Bitcoin analytics tool with the assistance of the INTERPOL National Central Bureaus (NCBs) in our 192 member countries to identify any system limitations and bugs prior to its official release.

Acknowledgments

We would like to thank Pieter Hartel, Costel Ion and Christian Karam for their valuable comments and feedback.

References

[1] CoinMarketCap, “CryptoCurrency Market Capitalizations,” https://coincap.com/, accessed Sept. 20. 2017.
[2] Blockchain.info, “Bitcoin Block Explorer,” https://blockchain.info/, accessed Feb. 06. 2016.
[3] The Financial Action Task, “Virtual Currencies – Key Definitions and Potential AML/CFT Risks,” https://www.fatf-gafi.org/publications/methodsandtrends/documents/virtual-currency-defintions-aml-cft-risk.html, accessed Dec. 22. 2015.
[4] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” https://bitcoin.org/bitcoin.pdf, accessed Feb. 11. 2016.
[5] Reddit, “DarkNetMarkets,” https://www.reddit.com/r/DarkNetMarkets/, accessed Sept. 12. 2017.
[6] DeSnake, “AlphaBay Market,” http://pwoah7to6au2pul.onion/, accessed Sept. 12. 2017.
[7] S. Meiklejohn, P. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G.M. Voelker, and S. Savage, “A Fistful of Bitcoins: Characterizing Payments Among Men with No Names,” Proc. 2013 Internet Measurement Conference, Barcelona, Spain, pp.127–140, Oct. 2013.
[8] D. Ron and A. Shamir, “Quantitative Analysis of the Full Bitcoin Transaction Graph,” Financial Cryptography and Data Security 2013, Lecture Notes in Computer Science, vol 7859. Springer, Japan, Okinawa, pp.6–24, April 2013.
[9] F. Reid and M. Harrigan, “An Analysis of Anonymity in the Bitcoin System,” Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, USA, Boston, Oct. 2011.
[10] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J.A. Kroll, and E.W. Felten, “SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies,” Proc. 36th IEEE Symposium on Security and Privacy, USA, San Jose, pp.104–121, May 2015.
[11] Chainalysis Inc, “Chainalysis Reactor,” https://chainalysis.com/, accessed Feb. 06. 2016.
[12] A. Janda, “WalletExplorer.com: Smart Bicoin Block Explorer,” https://www.walletexplorer.com/, accessed Feb. 06. 2016.
[13] P. Stevens, “Blockchain,” http://blockchain.com/blog/, accessed Nov. 22. 2016.
[14] I. Brugere, “Bitcoin Transaction Network Dataset,” http://comphio.cs.uic.edu/data/bitcoin/, accessed Nov. 22. 2016.
[15] M. Spagnuolo, M. Federico, and Z. Stefano, “BitiOnion: Extracting Intelligence from the Bitcoin Network,” Financial Cryptography and Data Security 2014, Lecture Notes in Computer Science, vol 8437. Springer, Christ Church, Barbados, pp.457–468, March 2014.
[16] M.W. Carrie, “An Historical Perspective of Digital Evidence: A Forensic Scientist’s View,” International Journal of Digital Evidence, vol.1, issue.1, pp.1–9, Spring, 2002.
[17] BitcoinWiki, “Proof of Work,” https://en.bitcoin.it/wiki/Proof_of_work, accessed Sept. 12. 2017.
[18] BitcoinWiki, “List of Address Prefixes,” https://en.bitcoin.it/wiki/List_of_address_prefixes, accessed April 19. 2018.
[19] The Bitcoin Foundation, “Bitcoin Core,” https://bitcoin.org/, accessed Feb. 06. 2016.
[20] J.W. Bos, J.A. Halderman, N. Heninger, J. Moore, M. Naehrig, and E. Wustrow, “Elliptic Curve Cryptography in Practice,” Financial Cryptography and Data Security 2014, Lecture Notes in Computer Science, vol.8437. Springer, Christ Church, Barbados, pp.157–175, March 2014.
[21] K. Liao, Z. Zhao, A. Doupe, and G.-J. Ahn, “Behind Closed Doors: The Handled Data which is Currently Not Considered by Any of the Related Work Examined,” IEICE TRANS. INF. & SYST., VOL.E101–D, NO.11 NOVEMBER 2018
[28] Bitcoincharts, “Bitcoincharts,” https://bitcoincharts.com/, accessed Feb. 06. 2016.
[29] 21.co, “Bitnodes,” https://bitnodes.21.co/, accessed Feb. 06. 2016.
[30] Elliptic Inc, “Elliptic,” https://www.elliptic.co/, accessed Feb. 06. 2016.
[31] M. Moser, R. Bohme, and D. Breuker, “An Inquiry into Money Laundering Tools in the Bitcoin Ecosystem,” eCrime Researchers Summit 2013, San Francisco, USA, Sept. 2013.
[32] T. Moore and N. Christin, “Beware the Middleman: Empirical Analysis of Bitcoin-Exchange Risk,” Financial Cryptography and Data Security 2013, Lecture Notes in Computer Science, vol.7859, Springer, Japan, Okinawa, pp.25–33, April 2013.

Hiroki Kuzuno received the M.E. degree in Information Science from the Nara Institute of Science and Technology, Nara, Japan, in 2007. Since joining SECOM in 2007, he has been actively engaged in the research field of information security. He is now a Cyber-Security Researcher at SECOM Singapore and INTERPOL Global Complex for Innovation. He is a member of IEICE, IPSJ.

Giannis Tziakouris is a Digital Crime Analyst at the INTERPOL Global Complex for Innovation, working at Innovation Centre directorate. His current research involves various projects related to Darknet and cryptocurrencies. Prior to that, he was a PhD researcher in Cyber-Security and a Senior Teaching Associate at the University Of Birmingham, UK.