TxMon : Tool for Monitoring Illegal Bitcoin Transactions

Sejin Jeong
GSIS, KAIST
skyshiri@kaist.ac.kr

Nohyun Kwak
GSIS, KAIST
nhkwak@kaist.ac.kr

ABSTRACT

Recently, as bitcoin price surged, interest in bitcoin investment has been increasing, and in proportion to this, the number of crimes using bitcoin is also increasing steadily. Bitcoin gained lots of media attention for being an anonymity and often used for illegal purposes based on anonymity. But the chain of bitcoin transaction is transparent and traceable. Therefore, it becomes necessary to identify the owner of bitcoin related to suspected of crime. We suggest the tool for monitoring illegal bitcoin transactions. Our perspective is when bitcoin is used in trackable or open place, we can monitor illegal bitcoin transaction. TxMon can extract all bitcoin addresses related to an address that is suspected of illegal deals. TxMon also monitors all transactions in the latest block to see if there are any illegal transactions. The strength of TxMon is that it allows real time monitoring. If transaction occurs between blacklist and whitelist regarded as illegal transaction, then the tool generates an alarm to signal that a transaction needs to be tracked. We evaluated 4 whitelists identified on Youtube and 3 blacklists identified by related works. We crawled bitcoin addresses in *blockchain.info* with increasing depth. TxMon crawled whitelist into 3 depths with an average of 1,312 addresses found in 1,430 seconds and 276 addresses for 138 seconds in blacklist. We tested TxMon whether it properly monitors illegal bitcoin transaction in mining pool and latest block. TxMon alarmed illegal transaction in real time and there was no overload occurred.

Keywords

Bitcoin; Blockchain; Monitoring; Illegal transaction; Deanonymizing

1. INTRODUCTION

Bitcoin is a distributed, cryptographic digital currency that was introduced by Nakamoto in 2008[13]. In order to send and receive bitcoins, a user has to create a key pair, which consist of a public key, that serves as an account identifier, and a private key, that is used to sign transactions. Each transaction has a list of inputs and outputs. The inputs refer to previous transactions, that contain a certain amount of bitcoins, in order to enable all members of the network to verify, that these coins have not already been spent. The transaction usually has two outputs, one output destination is the address (public key) of the recipient, the other output belongs to the sender of the bitcoins. Bitcoin uses a proof-of-work system to verify transactions and to prevent double-spending. Conicts in the system are resolved by majority decisions, with the weight of the vote based on computational power. On average, every ten minutes a new block is created, which bundles a number of valid transactions and refers to the previous block, thereby extending the blockchain. To check, whether the inputs of a transaction have already been spent, all clients keep an index of unspent transactions and reject those with invalid inputs from being integrated into a block [6].

Bitcoin gained lots of media attention for being an anonymity. Anonymity in the bitcoin is based on bitcoin address cannot be mapped to the real identity and new transactions are spread radially, thus the user's IP address will not be exposed. However, the weakness of bitcoin anonymity is reflected in the following areas. The real-name authentication mechanism helps bitcoin service providers to find the addresses that ever deposited and withdrew. Bitcoin address exposed on the internet can be related to its owner. And chain of transaction is transparent and traceable[10].

Bitcoin is often used for illegal purposes based on anonymity. For example, there was a case in which a hacker forcibly encrypted victim's files with ransomware and threatened to pay a bitcoin to decrypt them. This shows that the transaction with bitcoin is used as an illegal means of payment because it has anonymity.

Assuming that Police knew bitcoin address in open place like a bitcoin exchange market denoted as whitelist. They can easily get whitelist by web crawling or acutal visit. And bitcoin addresses suspected of illegal transactions denoted as blacklist can be identified by police through investigation or web crawling. To get a blacklist is harder than whitelist by web crawling, but it's not impossible. Blacklist is anonymous, it's difficult to identify the bitcoin owner, and furthermore it's more difficult to arrest the criminals using bitcoin. However, if blacklist have connection with whitelist, it will be traceable.

In this paper, we propose a tool for monitoring illegal bitcoin transaction. First, crawl bitcoin address for monitoring

WOODSTOCK '97 El Paso, Texas USA

ACM ISBN 0-12345-67-8/90/01…$15.00

DOI: 10.475/123_4

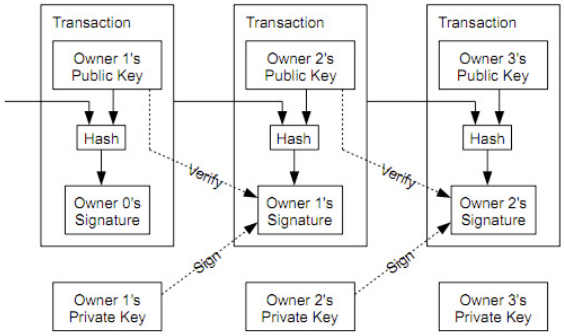

Figure 1: Blockchain of bitcoin transactions

blacklist and whitelist. Second, monitor all transactions in the latest block to see if there are any illegal transactions. Last, monitor unconfirmed transaction from mining pool in real time. Unconfirmed transactions are reflected in the latest block after 10 minutes on average, so if we can monitor illegal connection between blacklist and whitelist in unconfirmed transaction, it can be possible to catch the criminal in advance.

2. BACKGROUND

2.1 Dark Web

Dark web is World Wide Web content that exists on darknets, overlay networks which use the Internet but require specific software, configurations or authorization to access. Dark web forms a small part of the deep web, which is not indexed by search engines. So dark web can be used for hiding criminal activity. Individuals can access dark web using special software such as Tor(The Onion Router). Tor relies on the volunteer computer network to route the user's web traffic through a set of other user computers, so that traffic cannot be traced back to the original user. Once on dark web, users navigate through the same directory as the "Hidden Wiki", which organizes sites by category, just like Wikipedia. Individuals can also search the dark web using a search engine. Search engines are used to search for contraband such as broad or unspecified drugs, guns or counterfeit bills. In dark Web, individuals can communicate through methods such as secure e-mail, web chat, or personal messaging hosted on Tor.

2.2 Blockchain

Blockchain is a continuously growing list of records, called blocks, which is linked and secured using cryptography. Each block contains hash as a link to a previous block, a timestamp and transaction data. By design, blockchain is inherently resistant to modification of the data. Blockchain can serve as open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way. For use as a distributed ledger, blockchain is typically managed by a peer-to-peer network collectively adhering to a protocol for validating new blocks. Once recorded, the data in any given block cannot be altered without the alteration of all subsequent blocks, which needs a collusion of the network majority. Blockchain is secure by design and example of a distributed computing system

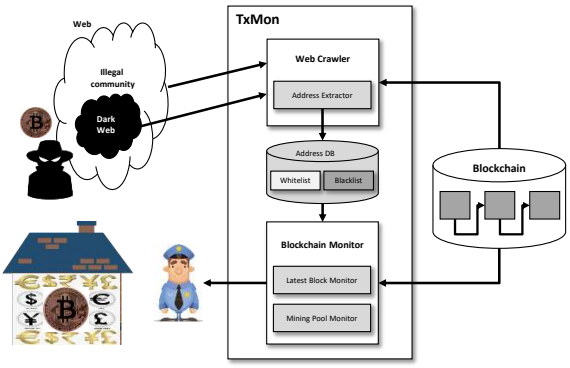

Figure 2: TxMon System Overview

with high Byzantine fault tolerance. Decentralized consensus has therefore been achieved with a blockchain. The first distributed blockchain was conceptualised in 2008 and implemented the following year as a core component of the digital currency bitcoin, where it serves as the public ledger for all transactions[1].

3. DESIGN

Figure 2 shows the overall architecture for monitoring illegal transactions: The client of our solution is the police. The police uses TxMon[1] for monitoring illegal transactions. TxMon monitors the illegal transaction based on the address DB and bitcoin's blockchain. Address DB has the bitcoin addresses crawled from the web. In the following subsections, we will explain in more detail.

3.1 Illegal Transaction

To define an illegal transaction, we must define what the illegal activities are. We consider activities that are generally classified as cybercrime as illegal activities. These are sales of stolen goods(or personal information), weapons, drugs, counterfeit bills, adult contents, and request for hacking or illegal software or illegal gambling. Especially these illegal things happen in the dark web or illegal community.

Darknet websites, called dark web, are accessible only through networks such as Tor ('The Onion Router'). Tor-accessible sites can be identified by the domain '.onion'. Identities and locations of darknet users cannot be tracked due to the layered encryption system. Since people believe that Bitcoin provides anonymity, a lot of transactions have been made through bitcoin in the dark web or illegal community.

3.2 Web Crawler

The web crawler collects the bitcoin address information from the web. Then, the transaction history of the corresponding address is searched in the blockchain. The transactions include address as the receiver or the sender of bitcoin. And we store the address information on Address DB

[1]We call our tool TxMon as monitoring illegal bitcoin transactions

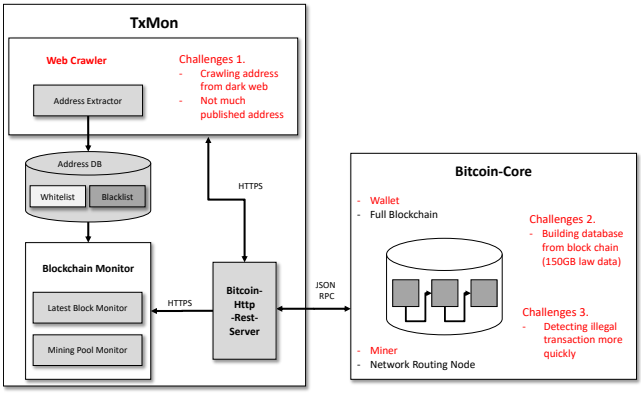

Figure 3: Technical Challenge

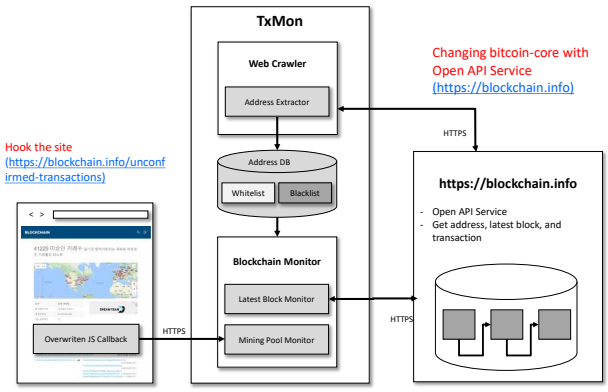

Figure 4: Implementation

as whitelist or blacklist. Addresses found on the dark web or illegal community are regarded as the illegal addresses associated with the illegal transactions and we store them to the blacklist. Any addresses from the trusted web site are regarded as the trusted address and we store them to the whitelist. Trusted web sites can be a licensed crypto currency exchanges or well-known shopping malls where user can trade with bitcoins.

3.3 Blockchain Monitor

The blockchain monitor keeps track of all the stored bitcoin address in the address DB and monitor all transaction from the blockchain by checking whether the new transaction is related with the connection between whitelist and blacklist. Whenever a connection is found between whitelist and blacklist, TxMon reports that transaction to the police. If a criminal attempts to exchange bitcoin for cash through an off-line exchange, the police will be able to catch the criminal by knowing through TxMon's report.

4. IMPLEMENTATION

We tried to implement the design described in the previous chapter with bitcoin-core [5]. Bitcoin-core is the reference client of bitcoin. But there were several technical challenges to implement our solution. Figure 3 shows the difficulties in implementing the TxMon System using the bitcoin core.

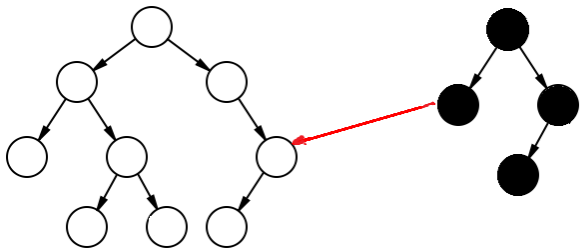

Figure 5: Transaction between blacklist and whitelist Addresses

First is web crawling. We tried to crawl the bitcoin address on the dark web. But, on the dark web, the basic guideline for user is not to reveal the bitcoin address to the public. This is because as soon as someone exposes the bitcoin address, it can become a tracking and deanonymizing target by looking for all transactions associated with that address. Also, trusted service providers do not publish bitcoin addresses well. Most of the moments when we get the bitcoin address from the trusted service provider is the time to make a payment through bitcoin.

Second is wallet in Bitcoin-core. Bitcoin-core consists of wallet, full blockchain, miner and network routing node. Web crawler and blockchain monitor of TxMon have to know blockchain information fully, but the bitcoin-core wallet indexes only the addresses and transactions related to the user or pre-registered info in a separate DB, and does not index the entire address and transactions. Adding an extra address to the bitcoin-core wallet requires a lot of time because the bitcoin-core needs to rescan the blocks that are approximately 150GB in size.

Third is detecting illegal transaction more quickly. Reporting an illegal transaction quickly can be effective in preventing crime. We will explain how to overcome these challenges in the next subsections.

The reason for mentioning that it is difficult to implement the design originally conceived with the bitcoin core, as shown in figure 3, is to share know-how with subsequent researchers to avoid the same mistakes. So our final structure is shown in figure 4.

4.1 Web Crawler

First challenge is Web Crawling. In dark web, bitcoin address is confidential. For example, Darknet website checks the user's identity by compelling the user to tag a specific keyword on facebook. Even on normal web, the published bitcoin address is not more than we expected.

So we used the manually found bitcoin address as the input value, and expanded the relevant addresses from the blockchain. Figure 5 shows the relationship between two address groups, whitelist and blacklist, in a graphical representation. White nodes are addresses on the whitelist and black nodes are addresses on the blacklist. Based on the assumption that the input address is the root node of the tree, we add the addresses that received the bitcoin sent from the input address to the tree as the children node. Also, based on the children node, the addresses that received the bitcoin sent from the addresses of children node are added as grandchildren nodes. This increases the depth[2] of the tree

[2]the first blacklist identified as the root node 0depth, where

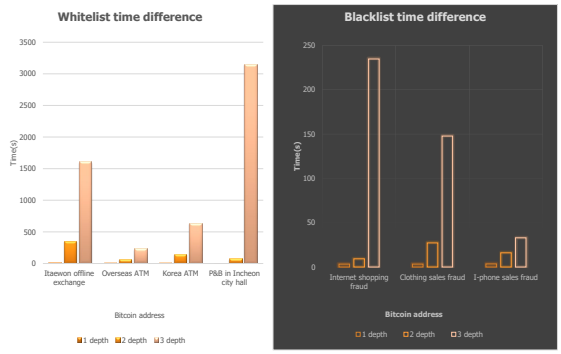

Figure 6: Time difference between whitelist and blacklist

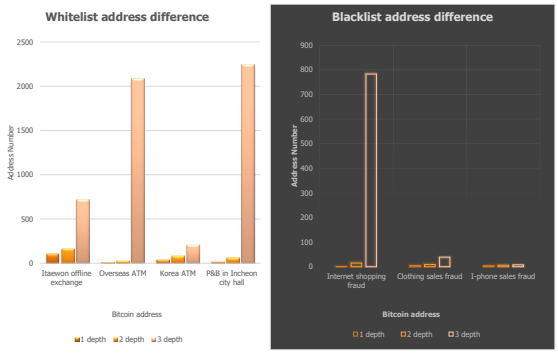

Figure 7: Address number difference between whitelist and blacklist

while we crawl the address in the blockchain. Its intuitive meaning is to track bitcoin transactions on the blockchain. If there is a transaction between whitelist and blacklist like the red arrow in the figure 5, it is an opportunity to investigate illegal addresses from a certain identity as a starting point.

4.2 Building DB from blockchain

Second challenge is Building DB from blockchain. As we described above, the bitcoin-core does not index the entire address and transactions but indexes only a small number of the addresses and transactions in DBMS. So it is necessary to build database(DB) from block chain. In other words, additional meta-data(DB) must be created for blockchain. It needs to manage large DB with a lot of operational know-how.

So we used Open API Service from *blockchain.info* [2]. Although it saved a lot of development time and effort, we had to implement a way to circumvent because of the limited number of requests to send a query.

4.3 detecting an illegal transaction

Third challenge is detecting an illegal transaction more quickly. When a transaction is requested, the transaction enters into the mining pool before it is reflected in the blockchain. So If we monitor the mining pool, we can get a warning about illegal transactions faster.

But Open API is limited to send many requests whenever the new transaction is generated. Furthermore, the query result for the unconfirmed transactions in *blockchain.info* reflects only the 10 most recent occurrences. Therefore, it is very complicated to send a request continuously and to check if there are any missing unconfirmed transactions according to the interval of sending the request.

So we hooked a site which shows unconfirmed transactions in real time. We disabled CSP (Contents Security Policy) with Chrome Extension and overwrote the JavaScript callback function. To be more specific, we analyzed JavaScript of the site showing the unconfirmed transactions, and found the callback function. The callback function renders the unconfirmed transactions on the screen. By replacing the callback function with our callback function, some code were added for sending the information about the unconfirmed transactions to the blockchain analyzer of TxMon to monitor the illegal transaction. We can not send a request to the blockchain analyzer because of CSP. So we disabled CSP with Chrome Extension and we succeeded in sending the request. You can see a demo of this part on YouTube[3].

1depth is the bitcoin address traded with 0depth, and 2depth is the bitcoin address traded with 1depth.

5. EVALUATION

We evaluated 4 whitelists and 3 blacklists. Whitelists are Itaewon offline exchange shop, overseas ATM, ATM in Korea, and Paris-Baguette in Incheon city hall. Whitelist address is easily accessible through web surfing or actual visit. But blacklist is hard to get. Because criminals know the weakness of bitcoin anonymity as "Address exposed on the internet can be related to its owner". In the dark web, bitcoin address is disclosed only after passing several levels of authentication. So we get the blacklist address by police and related works. We crawled this bitcoin address. So we get the blacklist address by police and related works. We crawled this bitcoin address.

5.1 Web crawling bitcoin address

As showed in figure 6, time to calculate 1 depth for whitelist and blacklist is very small. But as depth increased 2 and 3 depth, time increased exponentially. And when we tried to calculate depth 4, number of addresses is above 5,000 and *blockchain.info* denied service with HTTP 500 error in message internal server error. We guessed that the server has blocked in advance because web crawler made too many requests to server.

Figure 7 shows address number difference of blacklist and whitelist as depth increased. This figure is similar as time difference figure. As depth increased 1 to 2 and 3 depth, address number increased too. The main feature of the whitelist is that as depth increased, number of address increased exponentially, but blacklist is can be small even if the depth is increased like the address used as clothing sales fraud and I-phone sales fraud in figure 7. Because whitelist is public and many people can use it, so whitelist increased exponentially as depth increased. But blacklist is used cautiously for illegal transaction, there will not be many transactions at that address and the number of transaction is not likely to increase exponentially as depth increased. If number of address increased exponentially as the depth in-

Place		Address	Total received(BTC)
whitelist	Itawon shop	1Q2ogJniBJLwgnwpRgi5bWRXkKBF3yxrh	81.25
	Overseas ATM	14HV8pzvC7NdWZujmFBFq4ZHZL7fcYgPjH	1778.6
	ATM in Korea	1PABDWQzaGkfKoz8w5XucbStPQiHGneWQY	12.2
	Paris Baguette	1L827x6F5MzpdV61X5nYPWMKqtHBL5dRXL	0.85
blacklist	Internet shopping fraud	139XLp8U6sHSwYqgyuUnDtfTaMGuhFPZHz	2.939
	Clothing sales fraud	1KH5iMuJvrjCEud4PnQfaSig4LSDWEGYY5	9.012
	I-phone sales fraud	17Q4VPBsB4F4pEgJeqogQQDCQHQnfWAMtY	31.146

Table 1: Whitelist and Blacklist

Place		1 depth		2 depth		3 depth	
		time(s)	number	time(s)	number	time(s)	number
whitelist	Itawon shop	11.2	106	352.6	170	1,602.5	717
	Overseas ATM	8.5	17	62.3	32	337.1	2,081
	ATM in Korea	7.7	43	145.6	81	626.2	214
	Paris Baguette	5.0	20	69.9	67	3,152.6	2,239
blacklist	Internet shopping fraud	3.2	1	9.2	15	234.7	783
	Clothing sales fraud	3.2	5	27.1	9	147.8	39
	I-phone sales fraud	3.4	4	16.0	6	32.9	8

Table 2: calculation time and address number as depth increased

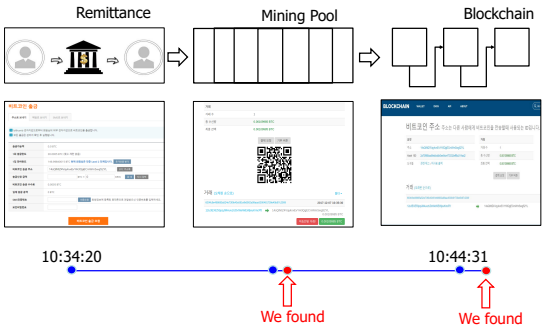

Figure 8: Monitor mining pool and latest block

creased, there will be address that makes the number of address increasing exponentially, and this address is likely to be a whitelist that we do not have. So it needs to exclude those addresses from the blacklist, and we can save computation time to crawl bitcoin address from *blockchain.info*.

5.2 Monitoring mining pool & latest block

We tested TxMon whether it properly monitor illegal bitcoin transaction in mining pool and latest block. First we deposited the bitcoin in *bithumb* cryptographic exchange regarded it as blacklist and then, sended it to the other address regarded as illegal transaction occurred by transaction between blacklist and whitelist. TxMon alarmed 20 seconds after an illegal transaction happened and there was no overload occurred in real time monitoring.

6. DISCUSSION

In web crawling, it's difficult to crawl information from the dark web or illegal community. Criminals are generally concerned about being arrested, so they are reluctant to make accurate expressions and often use surrogate language or conceal them from the police or the general public. Therefore, it's very different from the assumption of existing crawling technology that data is publicly published for the purpose of sharing and it is easy to pattern how to find it. Therefore, more advanced crawling technique should be used. The recently published crawling technique is described in related work.

For the limited requests to *blockchain.info*, continuous web crawling is required to extract addresses at deeper depths. In the evaluation of this paper, TxMon's blacklist and whitelist are crawled into 3 depths. But in the case of blacklist, it's necessary to further increase the depth and trace it to the end. However, if TxMon increases depth, we will not be able to guarantee the consistency of the result of the crawling because the error occurs during the crawling due to the restriction of the number of Open API requests. A simple solution to this problem is to continue crawling. For example, TxMon processes only 1000 requests at first, keeps track of how much crawling has been done, keeps crawling on a daily basis, and maintains consistency.

Using blockchain technology does not guarantee safety. Generally, blockchain technology is known to be more secure because it's difficult to modulate the original data. However, when we applied the blockchain technology to the web for the implementation of this paper, the blockchain itself is not the DB but the raw data that is guaranteed to be integrity. Even though the integrity of the raw data is assured, direct access to the raw data, which is about 150GB in size, takes too much time. Therefore, it's necessary to construct and use meta-data(DB) for the raw data. This means that the

blockchain technology doesn't replace the existing database, but rather requires a higher level database technique.

7. RELATED WORK

Portnoff et al. [9] proposed an automated analysis of cybercriminal markets through natural language processing. As mentioned in the discussion, the cybercriminal market is only accessible through manual exploration because it is difficult to deal with the crawling technology. Manual exploration is not an automated method and there is limit to scalability. Therefore, Portnoff et al. attempted to grasp more precise meaning through natural language analysis. In this paper, the proposal of Portnoff et al. can be used to help identify illegal activities related to cybercriminal. However, getting bitcoin addresses still requires manual intervention, such as facebook tagging, which still has difficulties to overcome.

Current study on bitcoin deanonymization focuses on Analysis of the Transaction Chain(ATC) which is to obtain transactions from public blockchain data to classify bitcoin addresses based on the weakness of bitcoin anonymity, and to relate bitcoin addresses to personal identities [11, 8, 12, 6, 7]. ATC is able to find clue to connect bitcoin addresses with real or virtual user identity information [10]. For the ATC attacks, coin-mixing obfuscates the transaction chain, and separate the corresponding relationship between the input and output of a bitcoin transaction and even hide the amount of transaction [4, 15, 14]. Coin mixing uses the concept of a shared wallet. User sends bitcoins to the addresses of Coin mixer owns. Once a payment has been confirmed, the amount of bitcoins is transferred to the destination address using a different address, Then, it's no longer linked to the sender's address[10].

Related works are considering anonymity and deanonymity of bitcoin itself. And when using coin-mixing, ATC has limitation to track transaction because of no relationship between bitcoin address. Our perspective is when bitcoin is used in trackable or open place, we can monitor illegal bitcoin transaction even though coin-mixing is used, TxMon can extract all bitcoin addresses and investigate them if they have criminal charges. In addition, using coin mixing can be considered to be highly relevant to crime.

8. FUTURE WORK

Optimization is needed in the process of monitoring blacklist. As we mentioned in the evaluation, it's our assumption that the number of addresses does not increase abruptly even if the depth of blacklist is increased. However, in the case of Internet shopping fraud, the number of addresses increases rapidly at 3 depth. It is expected that this is because whitelist that we don't know is in the middle. Therefore, by setting the threshhold in the process of increasing the depth of the blacklist, nodes that increase the number of addresses suddenly do not crawl anymore. This is because of our assumption that there will be many transactions because this node is a whitelist. Such a transaction can be found later in the process of adding a new whitelist. This means that the connection between blacklist and whitelist we are looking for has already occurred.

We also want to enhance the monitoring of illegal transactions by adding the demixing function against the simple mixing attack described in related works. In the process of monitoring illegal transactions, similar transactions are found and considered as mixing candidates, and the mixing candidates are stored in the database and monitored. A simple example of finding a mixing candidate would be a transaction that transfers the same amount of bitcoin and is located in the blocks close to and behind the original block, which means the time similar to the original transaction. Although performance overhead can occur by adding this function, the overall performance will not be effected because only a small number of addresses are monitored through the optimization process described in the previous paragraph.

9. CONCLUSION

By developing web application that uses bitcoin, we explored several aspects related to the blockchain and web application. And we propose a tool for monitoring illegal bitcoin transaction. The strength of TxMon is that it allows real time monitoring. If transaction occurs between blacklist and whitelist regarded as illegal transaction, then the tool generates an alarm to signal that a transaction needs to be tracked. We evaluated 4 whitelists identified on Youtube and 3 blacklists identified by related works. We crawled bitcoin addresses in *blockchain.info* with increasing depth. TxMon crawled whitelist into 3 depths with an average of 1,312 addresses found in 1,430 seconds and 276 addresses for 138 seconds in blacklist. We tested TxMon whether it properly monitors illegal bitcoin transaction in mining pool and latest block. TxMon alarmed illegal transaction in real time and there was no overload occurred.

10. REFERENCES

[1] blockchain. https://en.wikipedia.org/wiki/Blockchain.

[2] blockchain.info. https://blockchain.info.

[3] TxMon's demo for Monitoring Blockchain Mining Pool. https://youtu.be/0Cn4IVaQ6S8.

[4] Joseph Bonneau, Arvind Narayanan, Andrew Miller, Jeremy Clark, Joshua A Kroll, and Edward W Felten. Anonymity for bitcoin with accountable mixes. *Preprint*, 2014.

[5] Bitcoin Core Developers. Bitcoin core. *https://bitcoin.org*.

[6] M. Malte. Anonymity of bitcoin transactions. In *Münster bitcoin conference*, pages 17–18, 2013.

[7] M. Malte, B. Rainer, and B. Dominic. An inquiry into money laundering tools in the bitcoin ecosystem. In *eCrime Researchers Summit (eCRS), 2013*, pages 1–14. IEEE, 2013.

[8] Micha Ober, Stefan Katzenbeisser, and Kay Hamacher. Structure and anonymity of the bitcoin transaction graph. *Future internet*, 5(2):237–250, 2013.

[9] R. Portnoff, S. Afroz, G. Durrett, J. Kummerfeld, T. Berg-Kirkpatrick, D. McCoy, K. Levchenko, and V. Paxson. Tools for automated analysis of cybercriminal markets. In *Proceedings of the 26th International Conference on World Wide Web*, pages 657–666. International World Wide Web Conferences Steering Committee, 2017.

[10] S. QingChun and Y. JianPing. Research on anonymization and de-anonymization in the bitcoin system. *arXiv preprint arXiv:1510.07782*, 2015.

[11] Fergal Reid and Martin Harrigan. An analysis of anonymity in the bitcoin system. In *Security and*

privacy in social networks, pages 197–223. Springer, 2013.

[12] Dorit Ron and Adi Shamir. Quantitative analysis of the full bitcoin transaction graph. In *International Conference on Financial Cryptography and Data Security*, pages 6–24. Springer, 2013.

[13] N. Satoshi. Bitcoin: A peer-to-peer electronic cash system, 2008.

[14] QingChun ShenTu and JianPing Yu. A blind-mixing scheme for bitcoin based on an elliptic curve cryptography blind digital signature algorithm. *arXiv preprint arXiv:1510.05833*, 2015.

[15] Luke Valenta and Brendan Rowan. Blinded, accountable mixes for bitcoin. *exchange*, 24:34.