A Large-scale Empirical Analysis of Ransomware Activities in Bitcoin

KAI WANG, School of Computer Science, Fudan University, China
JUN PANG, Department of Computer Science, University of Luxembourg, Luxembourg
DINGJIE CHEN and YU ZHAO, Software School, Fudan University, China
DAPENG HUANG, School of Computer Science, Fudan University, China
CHEN CHEN and WEILI HAN, Software School, Fudan University, China

Exploiting the anonymous mechanism of Bitcoin, ransomware activities demanding ransom in bitcoins have become rampant in recent years. Several existing studies quantify the impact of ransomware activities, mostly focusing on the amount of ransom. However, victims' reactions in Bitcoin that can well reflect the impact of ransomware activities are somehow largely neglected. Besides, existing studies track ransom transfers at the Bitcoin address level, making it difficult for them to uncover the patterns of ransom transfers from a macro perspective beyond Bitcoin addresses.

In this paper, we conduct a large-scale analysis of ransom payments, ransom transfers, and victim migrations in Bitcoin from 2012 to 2021. First, we develop a fine-grained address clustering method to cluster Bitcoin addresses into users, which enables us to identify more addresses controlled by ransomware criminals. Second, motivated by the fact that Bitcoin activities and their participants already formed stable industries, such as Darknet and Miner, we train a multi-label classification model to identify the industry identifiers of users. Third, we identify ransom payment transactions and then quantify the amount of ransom and the number of victims in 63 ransomware activities. Finally, after we analyze the trajectories of ransom transferred across different industries and track victims' migrations across industries, we find out that in order to obscure the purposes of their transfer trajectories, most ransomware criminals (e.g., operators of Locky and Wannacry) prefer to spread ransom into multiple industries instead of utilizing the services of Bitcoin mixers. Compared with other industries, Investment is highly resilient to ransomware activities in the sense that the number of users in Investment remains relatively stable. Moreover, we also observe that a few victims become active in the Darknet after paying ransom. Our findings in this work can help authorities deeply understand ransomware activities in Bitcoin. While our study focuses on ransomware, our methods are potentially applicable to other cybercriminal activities that have similarly adopted bitcoins as their payments.

CCS Concepts: • **Security and privacy** → *Malware and its mitigation.*

Additional Key Words and Phrases: Bitcoin transactions, Clustering, Ransomware

This article is an extended and revised version of the conference paper titled "Temporal Networks based Industry Identification for Bitcoin Users', which has been published and presented at the 16th International Conference on Wireless Algorithms, Systems, and Applications (WASA 2021), Nanjing, China, 2021: 108-120."
This article has been supported by Natural Science Foundation of China (U1836207), the National Key R&D Program of China (2019YFE0103800), and China Postdoctoral Science Foundation (2020M670998).

Authors' addresses: Kai Wang, School of Computer Science, Fudan University, China; Jun Pang, Department of Computer Science, University of Luxembourg, Luxembourg; Dingjie Chen; Yu Zhao, Software School, Fudan University, China; Dapeng Huang, School of Computer Science, Fudan University, China; Chen Chen; Weili Han, Software School, Fudan University, China.

1559-1131/2021/12-ART7 $15.00
https://doi.org/10.1145/3494557

ACM Reference Format:
Kai Wang, Jun Pang, Dingjie Chen, Yu Zhao, Dapeng Huang, Chen Chen, and Weili Han. 2021. A Large-scale Empirical Analysis of Ransomware Activities in Bitcoin. *ACM Trans. Web* 16, 2, Article 7 (December 2021), 29 pages. https://doi.org/10.1145/3494557

1 INTRODUCTION

Ransomware is a type of malware that prevents victims from accessing their valuable data by encrypting files or locking devices and then demands a ransom payment. Before the emergence of Bitcoin [36], victims were required to pay ransom by a collection of online cash-equivalent payment instruments, such as Paysafecard and MoneyPak. For the ransomware criminals, these payment instruments have two major drawbacks: 1) their limited geographic availability narrows the scope of victims; 2) they are operated by companies that are subject to the local law, which might compel them to track the ransom recipients. To overcome these problems, many criminals require victims to pay ransom through bitcoins, after Bitcoin and the concept of cryptocurrency started gaining popularity in 2011. Bitcoin provides a decentralized and anonymous payment scheme, which is convenient for ransomware criminals to collect ransom from worldwide without exposing their true identities. Due to a large amount of ransom and a wide range of victims, ransomware activities have now become a severe threat to public safety, law enforcement, etc. The Biden administration even launched a ransomware task force and offered up to $10 million reward for the information on cyberattacks [11].

To deeply understand ransomware activities, previous studies [12, 21, 32, 40, 47] utilize publicly available Bitcoin transaction records to analyze ransom payment transactions and track ransom transfers. Paquet-Clouston et al. [40] empirically analyze ransom payment transactions related to 35 ransomware families from 2013 to mid-2017 and find that the amount of ransom payments has a minimum value worth of 12,768,536 USD (22,967.54 bitcoins). Huang et al. [21] track the financial transactions and find that ransomware criminals usually cashed out through BTC-e, a now-defunct Bitcoin exchange. While the previous studies provide many insights about the behaviors of ransomware criminals, victims' reactions in Bitcoin that can well reflect the impact of ransomware activities are largely neglected. Meanwhile, these studies track the ransom transfers only at the Bitcoin address level, making it difficult to uncover the patterns of ransom transfers from a macro perspective beyond Bitcoin addresses.

In the past years, with the development of Bitcoin, various economic activities with similar purposes have gradually formed stable groups (we referred to them as *industries* according to their business purposes in this paper), e.g., Darknet and Miner. Similar to the topic community in the citation network [17], the Bitcoin industry consisting of activities with similar purposes can also be considered as a significant modular structure. Several studies have shown that identifying modular structures and analyzing their interactions can provide a better understanding of the development of various activities [14, 41, 48]. As evidenced by Chen et al. [8], some illegal activities and behaviors in Bitcoin have also been consolidated into some communities, i.e., Bitcoin industries in this paper. Thus, the evolution of industries can well reflect the development of ransomware activities, which provides a better way to analyze the patterns of ransom transfers and victims' reactions in Bitcoin.

In this work, we are motivated to quantify the amount of ransom and the number of victims in 63 ransomware activities from 2012 to 2021 and analyze ransom transfers and victim migrations across various Bitcoin industries. To the best of our knowledge, our study is the first to explore ransomware activities from the industry perspective over such a long period. We design a fine-grained address clustering method to accurately cluster Bitcoin addresses controlled by the same Bitcoin user into users. We use *user* to denote a group of Bitcoin addresses generated through our address clustering method, while a natural user involved in Bitcoin is noted as *Bitcoin user*. Next, we design a method to identify the major industry identifier of users that reflects their activity purposes. The details of this approach are described as follows: We first cluster Bitcoin addresses into users and classify their industry identifiers in the five well-known Bitcoin industries, i.e., Darknet, Exchange, Gambling, Miner and Investment, by training a multi-label classification model with an average accuracy of 92.00%. We observe that a non-negligible proportion (23.35%) of users have multiple industry identifiers in a short period and note them as *multi-identifier users*. Furthermore, to understand the primary purpose of multi-identifier users, we devise a major industry identification method to determine the industry they are primarily engaged in within a given short period. Finally, we propose an industry-based approach to analyze ransom transfers and victim migrations across industries. To explore how criminals transfer ransom and cash it into the real world, we design a money tracking model to capture the trajectories of ransom transferred across industries. As for victims, we construct a user movement model to track victims' migrations across industries, which helps us understand victims' reactions and assess the stability of each industry when influenced by ransomware activities.

Empirical results. Utilizing the improved address clustering method, we find hidden addresses controlled by ransomware criminals and quantify the amount of ransom and the number of victims in 63 ransomware activities from 2012 to 2021. According to the above statistics, we take seven typical ransomware activities as case studies to perform our industry-based analysis. Our empirical results can be summarized as follows. (1) We track over $176 million in ransom payments made by 41,424 victims. (2) We discover that in order to obscure the purposes of their transfer trajectories when laundering money, ransomware criminals prefer to move ransom to multiple industries as participants in a short time period, especially through the Exchange industry rather than relying on the services offered by Bitcoin mixers. (3) We find that the Investment industry attracts Bitcoin users to engage in its activities continuously. Although some participants may leave the industry temporally, they are more likely to return to it after a period of time. This result indicates that the Investment industry is highly resilient against ransomware activities. (4) A few victims subsequently join in the Darknet industry after paying ransom. For instance, 8.82% of *WannaCry* victims carry out transactions with darknet vendors. This finding shows that ransomware criminals could further induce the victims to engage in other illegal activities.

We believe that our results can benefit several stakeholders. For researchers, we provide a general industry-based approach for analyzing illegal activities and recommend them to view Bitcoin as an economic society rather than an online social network. For authorities, our empirical results would help them achieve deep insights into the ransomware activities and adopt suitable regulation policies to reduce their negative impact. For instance, we can suggest authorities guide victims to participate in normal economic activities instead of entering the *Darknet* industry.

Organization. We introduce some background knowledge in Section 2 and discuss our data collection in Section 3. An overview of our approach is presented in Section 4. Next, we develop the method for clustering Bitcoin addresses in Section 5 and the method for identifying users' industry identifiers in Section 6. Based on these results, we perform a large-scale analysis of ransomware activities in Section 7. We interpret our findings and the shortcomings of our approach in Section 8. We discuss related work in Section 9 and conclude the paper in Section 10.

2 BACKGROUND KNOWLEDGE

2.1 Anonymity of Bitcoin

As claimed in its white paper [36], Bitcoin provides an anonymous and trusted payment mechanism for Bitcoin users to complete transactions in an open computing environment. The mechanism offers Bitcoin users two major advantages. First, all transaction data can be confirmed by any

Bitcoin user with integrity. The mechanism allows Bitcoin users to access all historical transaction data and apply a binary hash tree storage structure to locate the target transaction. Second, every Bitcoin user who wants to protect his/her privacy can anonymize transactions using a new Bitcoin address (i.e., one-time address) for each newly launched transaction. These one-time addresses can break the association among addresses held by the same Bitcoin user, thereby protecting Bitcoin users' private information.

Based on different purposes and forms, transactions can be described as several patterns. These transaction patterns often imply potential address associations, helping us identify Bitcoin users behind the anonymous addresses.

Bitcoin transaction patterns. In a typical transaction pattern, a sender sends the balances of his multiple Bitcoin addresses to the recipients and pays additional bitcoins to the miner as a transaction verification reward (i.e., miner fee). Similar to the change mechanism in the banknote payment method, when the number of bitcoins sent by the sender exceeds the sum of the recipients' expectation and the miner fee, the remaining bitcoins in the transaction are called *changes* and will be sent back to the sender. The address pre-defined by the sender to receive changes is called *change address*. In addition to this typical pattern, the following four particular transaction patterns are also considered in our work.

• *Coinbase transaction*: Apart from receiving the miner fee from senders, Bitcoin launches such transactions to reward miners who submit new blocks. The coinbase transaction is the first transaction in each block, which contains only the recipients but not the senders. All these recipients are miners.

• *Mixing transaction*: Under the services of Bitcoin mixers, money transfers between multiple Bitcoin users and their corresponding recipients are packaged into one single transaction. In other words, one mixing transaction completes several remittances at one time. This transaction pattern generated by the Bitcoin mixers, such as *Bitcoin Fog*, typically invalidates the rules in practical de-anonymization mechanisms, thus can be leveraged to protect the identity of bitcoin senders.

• *Peeling chain transaction*: These transactions consist of a single input address as the sender and two output addresses as recipients. Usually, a sender peels off a small number of bitcoins to one recipient and sends the remaining bitcoins to the other recipient. Then, the latter recipient will conduct a new transaction to continue the peeling off behavior. This process can be repeated hundreds or even thousands of times until all the bitcoins have been spent or transferred.

• *Locktime transaction*: Bitcoin supports senders to specify the effective time of a transaction through an optional field *Locktime*. There are two options for the field *Locktime*: 1) at a specific block height, and 2) at a specific timestamp. Generally, a single Bitcoin user has his/her preference to set the effective time of transactions. We call transactions following this pattern as locktime transactions.

Association of Bitcoin addresses. In practice, many Bitcoin users often reuse their Bitcoin addresses in multiple transactions for convenience. This 'reuse' potentially exposes the association of their addresses. For example, only the sender with his/her private key can unlock the balance in the address, thus normally all input addresses for a transaction should belong to the same sender. Once the sender reuses one of these input addresses in other transactions, the reused address will become a key to associate other Bitcoin addresses in his/her transactions. In addition, the literature [25] states that Bitcoin users have transaction preferences when participating in different activities. Therefore, personal behaviors in transactions, particularly the usage of change addresses, may become an important entry point for the practical detection of associated Bitcoin addresses.

Based on the above observations, we consider the effect of these special transaction patterns when performing Bitcoin address clustering in Section 5. In particular, we aim to improve the detection of associated addresses in two transaction patterns: *peeling chain* and *locktime*, which are often overlooked in previous studies.

2.2 Industries in Bitcoin

Many activities in Bitcoin use transactions as the carrier together with bitcoins as the settlement currency. Both the number and the value of transactions in these activities increase and gradually evolve to an industry level [33]. For example, the report [43] states that from December 2014 to April 2017, Bitcoin gambling games have received 3.7 million bitcoins as bets, and their popularity continues to grow. In this paper, we introduce the concept of *industry* in Bitcoin: An industry consists of the activities that provide goods or services for similar purposes and the group of Bitcoin users involved in these activities. Based on this definition, we present five industries in Bitcoin: (1) Darknet, where smuggling or illegal service transactions are traded through bitcoins (e.g., *SilkRoad*); (2) Exchange, where Bitcoin users complete exchange services between fiat currencies and cryptocurrencies (e.g., *Mt.Gox*); (3) Gambling, where bitcoins are used as bets in various gambling games (e.g., *SatoshiDice*); (4) Miner, where multiple miners or mining groups generate new blocks and distribute their rewards through coinbase transactions (e.g., *F2Pool*); (5) Investment, where offering the services of bitcoin returns and management, including bitcoin lending (e.g., *Nexo*), bitcoin faucet (e.g., *Cointiply*) and wallet management (e.g., *Trezor*).

The concept of the Bitcoin industry can be analogous to the industry in macroeconomics [39]. The evolution of the Bitcoin industry can well reflect the development of Bitcoin and provides a fundamental way to understand the large-scale Bitcoin economic society, which helps us deeply understand ransomware activities from a macro perspective.

According to the activity patterns and purposes of Bitcoin users in the industries, we can further describe the industry members with two roles: *organizer* and *participant.* As an organizer, a Bitcoin user provides goods or services for participants. For example, organizers such as drug traffickers have served participants within their respective industries for a long time. Within these industries, we note the specific industry identifiers of organizers as darknet vendors, exchange sites, gambling bankers, miner pool members, and investment merchants, respectively. Correspondingly, we note their participants as darknet customers, exchange buyers, gamblers, individual miners and individual investors. Since Bitcoin users are able to involve in different activities, they may play several different roles in multiple industries.

2.3 Ransomware

Ransomware is a type of malware that infects victims' data or resources and demands ransom to release them. It mainly uses two ways to block victims from accessing their data. The most common one is encrypting files that does not destroy other functions of the device. The other is locking the computer or other devices, which restricts all operations but does not directly encrypt the data stored on the device.

Ransomware has become more and more rampant since Bitcoin came into use in 2009. Bitcoin provides a decentralized and anonymous payment scheme, which encourages ransomware criminals to carry out extensive attacks and get paid safely without worrying about being caught or tracked [33]. As an example, the worldwide ransomware *WannaCry* hacked over 300,000 computers across 150 countries by encrypting files and asking for money to ransom them in 2017 [30], and victims were required to pay \$300 - \$600 in Bitcoin to three hardcoded Bitcoin addresses.

From ransomware spreading to the ransom withdrawal, we conclude that a successful ransomware activity needs to go through five stages: (1) *ransomware spread* where ransomware

Table 1. A data sample in the *Transactions* dataset.

Field		Content
TxHash		1d119180ae631c2a491ca273b9a95e7fa498d3e5ea884e1f0fceba07b171e8de
Input	Address	17AumjzL4hTzmeXb3ifKP3u7jwom3AF7nf
	Amount	1.0 bitcoins
	Prev_Tx	41c06303b88651e80ccd3c834646d544b3c365ea8b319e0cf56f4a076f1edc0e
Output	Address	1Cf4s57ErgQJAibuE3tzcWUPJaBpKb2GAc
	Amount	0.9999 bitcoins
	Spent_Tx	4d76ce6878c5af2c01f3011b1c4eff5e9c35e5df8d7cad920d873706f47654dd
Miner Fee		0.0001 bitcoins
Timestamp		1455774860 (i.e., 2016-02-18 13:54:20)
Locktime		398948

criminals implant malware into victims' devices through system vulnerabilities; (2) *data encryption or device locking*, as a result ransomware prevents victims from accessing their data by encrypting data or locking their devices; (3) *ransom payments*, i.e., some victims pay ransom to criminals in order to regain access to their precious data; (4) *ransom transfer* where criminals usually cover their tracks by transferring ransom money; and (5) *withdrawal* – eventually, the criminals perform a withdrawal operation through the exchange to convert the ransom money into legal tender. We focus on the behaviors of ransomware criminals and victims in Bitcoin; that is, we concentrate on the last three stages.

3 DATASETS

We describe three datasets used in our analysis: *Transactions*, *Entity Identities* and *Ransomware Activities*. The first dataset records Bitcoin transaction data, and the other two datasets contain publicly available labels of Bitcoin addresses from websites and previous studies: the *Entity Identities* dataset stores well-known entities, which helps us map anonymous Bitcoin users to their real-world identities; the *Ransomware Activities* dataset records known ransomware activities, which serves as an entry point to study the threat of ransomware activities to Bitcoin. Below we detail the collection methodology of each dataset.

(1) *Transactions* dataset. We download all raw Bitcoin transaction data from 01/03/2009 to 04/30/2021 and parse the data into address-based transactions. Table 1 shows a sample of parsed transactions. In total, we obtain 815,343,064 unique Bitcoin addresses and 633,648,723 transactions.

(2) *Entity Identities* dataset. We collect and preprocess Bitcoin addresses of known entities from *WalletExplorer* [23] and *Ethonym* [46] where the former is widely used as ground truth in several studies [15, 40]. Each address set of a known entity presents the association among its addresses, which helps us evaluate the performance of the address clustering method in Section 5.

Table 2. Extraction rules for participants in different industries.

Industry		**Participant**
Darknet		Sender
Exchange		Sender and Recipient
Gambling		Sender
Miner		Recipient in coinbase transactions
Investment	lending	Sender and Recipient
	faucet	Recipient

As a preparation step, we first clean Bitcoin addresses in this dataset. We exclude addresses that are duplicated, fail in validation checks, or have never been involved in any transactions. We then investigate the types of goods or services provided by these entities to classify these Bitcoin addresses into five industries precisely. Furthermore, to improve the quality of industry identifier classification, we categorize these addresses into organizers and participants of industries. Based on the service declarations of these entities, we classify their Bitcoin addresses as organizers except for the addresses of wallet management services. To facilitate the management of multiple Bitcoin addresses, the service of wallet management assigns each user a primary Bitcoin address as the public account to receive bitcoins from other users. In other words, the primary Bitcoin address represents the participant who utilizes the wallet management service.

After that, we summarize the extraction rules in Table 2 to identify participants from organizer-related transactions or coinbase transactions to identify participants to enrich the dataset. Due to the different types of services offered in the Service industry, whose service requirements vary in terms of transactions, we further divide the rule of this industry into two situations. The first one is the lending service. Bitcoin users can use the service to borrow bitcoins or as a beneficiary to temporarily lend out their own bitcoins to earn interest. Both of these behaviors mean that either the sender or the recipient of such transactions can be considered as the participants in the service. The second situation is the faucet service. After Bitcoin users accumulating enough advertisement clicks or video viewings on the faucet platform, the platform would pay them a certain number of bitcoins as rewards. Thus the recipients who receive the bitcoins in the transactions are regarded as participants of the service.

As a result, the dataset covers 382 known entities with 21,057,772 unique Bitcoin addresses as organizers together with 130,145,529 unique Bitcoin addresses as participants, accounting for 2.77% and 17.14% of the total number of Bitcoin addresses, respectively. These labels of industry organizer and industry participant are used for training an industry identifiers classifier in Section 6. Table 3 details the numbers of organizers and participants in every industry.

(3) *Ransomware Activities* dataset. This dataset records Bitcoin addresses and transactions involved in ransomware activities. We download ransomware activity data published in previous studies [2, 5, 12, 32, 40]. To enrich the dataset, we further collect and verify ransomware addresses and transactions posted on the forum *BitcoinTalk* SCAM Accusations board [37] and *BitcoinAbuse* website [44] that is a public database of Bitcoin addresses used by ransomware criminals. Similar to the data cleaning process in the *Entity Identities* dataset, we filter out invalid data. In total, this dataset contains 63 ransomware families with 22,717 Bitcoin addresses (called ransomware seed addresses), which serve as an entry point to study the impact of ransomware activities in Bitcoin.

Table 3. Number of organizers and participants in five Bitcoin industries.

Industry	# of Organizers	# of Participants
Darknet	2,332,854	5,657,783
Exchange	9,967,932	87,932,289
Gambling	3,098,500	14,451,596
Miner	38,664	683,704
Investment	5,619,822	21,420,157

More specifically, 15 ransomware families with 20,849 addresses were extracted entirely from previous studies. We collected another 13 ransomware families with 1,048 addresses from previous studies and enriched 766 addresses by crawlers. The other ransomware families with 54 addresses in the dataset were collected from websites – on BitcoinTalk, we crawled the texts related to the ransomware families from 2012 to 2020; and on BitcoinAbuse similarly we crawled text from 2017 to 2020.

4 AN OVERVIEW OF OUR APPROACH

This section introduces our industry-level approach for conducting an in-depth empirical analysis of ransomware activities in Bitcoin. The analysis consists of three steps: address clustering, industry identifier classification, and ransomware activity analysis.

Address clustering. Protected by the anonymous payment mechanism, it is hard to figure out the real intention of a Bitcoin user if we analyze his/her transactions solely based on the independent Bitcoin addresses. Thus, before capturing the industry identifiers of Bitcoin users, we develop a novel Bitcoin address clustering method to mine more associated addresses into users in Section 5. If several addresses belong to the same Bitcoin user, they should be clustered into one group. Through the association among addresses, we transform address-based transactions into user-based transactions.

Industry identifier classification. Based on the user-based transactions, we then classify dynamic industry identifiers through users' activity patterns in Section 6. Since Bitcoin users can conduct activities in various industries during different periods, we train a multi-label classification model to classify their industry identifiers of different periods based on several temporal networks. As a result, some users may have multiple industry identifiers from the classification model. Next, we devise a method to determine the major industry of multi-identifier users. With this step, we are able to reproduce major activity trajectories of users across multiple industries.

Ransomware activity analysis. Based on detected industry information of users and collected ransomware activity data, we first quantify the amount of ransom and the number of victims. Then, we propose a money tracking model and a user movement model to explore how the criminals transfer ransom across Bitcoin industries and how the victims in different industries react to the ransomware activities in Section 7.

5 BITCOIN ADDRESS CLUSTERING

The existing address clustering methods mine associated addresses by analyzing the payment behaviors in transactions, e.g., how to pay bitcoins and receive changes. One direct idea, as mentioned in the studies [25, 33, 47], assumes that all input addresses used for one specific bitcoin payment transaction should belong to one Bitcoin user. We note this idea as *MI* (Multiple Input). Due to the over-clustering problem caused by mixing transactions, many researchers put forward another idea [45] that excluding misleading mixing transactions before applying *MI*. We call the improved idea as *MX*, where *X* denotes *mixing transactions*. Meanwhile, some researchers analyze payment behaviors and extract transaction preferences to help uncover potential associations from change addresses.

In addition to the clustering results of input addresses, researchers have also focused on individual transaction behavior, including how to receive changes and pay bitcoins. Such behavior well reflects the potential relationship between Bitcoin addresses. Therefore, they have empirically proposed the following heuristic rules to identify change addresses or associated Bitcoin addresses:

1) New address rule (*NA*) [3, 33, 47]: In a two-output transaction, if one of the output addresses is a new Bitcoin address, then the new address is regarded as the change address of the inputs.
2) Decimal point rule (*DP*) [3, 25]: When a transaction has at least two outputs, if the receiving amount of one output address is three decimal points more than that of other addresses, the output address is considered as the change address of the inputs. This is based on the assumption that Bitcoin users are unlikely to send amounts to other users in cognitively-difficult amounts with a high number of decimal digits.
3) Special transaction rule (*SP*) [25]: The addresses in two consecutive transactions with the same transaction pattern belong to one user, such as the peeling chain transaction pattern.

However, applying these relatively coarse-grained constraints indiscriminately to different transactions patterns may mistakenly associate unrelated addresses to the same user (see Table 4 for an evaluation of these methods).

Observation of transactions in special patterns. Motivated to mitigate the above defects, we observe and summarize the features of two special transaction patterns, i.e., peeling chain and locktime, to help improve the performance of address clustering. (1) The peeling chain pattern is common in transactions, with about 43.11% of transactions matching this pattern. Moreover, 83.82% of its output addresses are the one-time addresses merely used for peeling off bitcoins within two consecutive peeling chain transactions. Namely, these addresses have only appeared in the peeling chain pattern. We argue that the combination of new addresses and the number of their receiving bitcoins can help to mine addresses association in peeling chain transactions. (2) With respect to locktime pattern transactions, we notice that a Bitcoin user is very likely to launch their locktime transactions in the same effective way. Furthermore, in 89.19% of these locktime transactions, all output bitcoins have been spent for subsequent payment purposes. Since the Bitcoin users entirely determine the effective time and the output status of such transactions, we consider these behaviors can help describe their personal preferences.

These identified features serve as entry points for detecting address association in these two special transaction patterns. We design a series of experiments to develop an accurate address clustering method and compare its performance with the existing methods. The evaluation processes are detailed later.

Our method. Our address clustering method consists of three parts. First, we apply *MX* as the basis and eliminate the interference in two types of mixing transactions[1]: CoinJoinMess [22] and JoinMarket [25]. To improve the performance of address clustering, we propose the following two additional heuristic rules.

Heuristic rule 1. Determining change addresses owned by the sender in peeling chain transactions. For one peeling chain transaction, we consider an output address with the following features as a change address of the sender: (1) the number of bitcoins received by the address is larger than that

[1]Excluding mixing transactions is only applied to the *MX* method, instead of directly excluding all the addresses involved in mixing transactions from the clustering results. These addresses may be associated with other users through normal transactions or be recorded as isolated users.

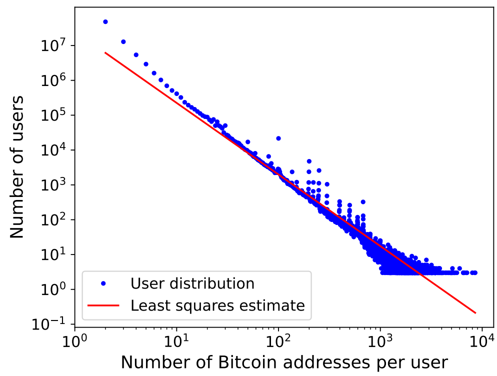


Fig. 1. User distribution follows Zipf's law.

of the other output address; (2) the number of bitcoins received by the address has three decimal points more than that of the other output address; (3) the address is a new Bitcoin address.

Heuristic rule 2. Determining associated Bitcoin addresses in locktime transactions. The input addresses of two consecutive transactions will belong to the same user if each transaction has the following features: (1) all the outputs of the transaction have been spent; (2) these two transactions specify the effective time exactly in the same way, i.e., a specific block number or a specific timestamp.

Analysis of clustering results. Since updating transaction data can dynamically modify the results of address clustering, we apply our address clustering method to the transaction data as of 04/30/2021, which is the focus of our empirical analysis in this paper. As a result, we group Bitcoin addresses into 389,240,195 users. Fig. 1 shows the distribution of addresses owned by per user. We notice that about 82.12% of users have one single address (called *isolated users*), and 16.28% of users have 2-10 Bitcoin addresses. Especially, 0.04% of users have more than 100 Bitcoin addresses. Due to the anonymous payment mechanism, it is expected to generate such a large number of isolated users through our address clustering method. By excluding outliers (i.e., we group users by the number of addresses they hold and filter out the group with less than three users), we plot the distribution of users and apply it with linear regression. We calculate the coefficient of determination R^2 as 0.95, which indicates the regression line with a high correlation with the distribution points. Based on these analyses, the user-address distribution in Bitcoin largely follows Zipf's law.[2] In addition, as we mentioned in Section 2, only a specific Bitcoin user with private keys can consume the balances in his Bitcoin addresses. Therefore, we use these users to represent Bitcoin users in our analysis.

Method evaluation. We evaluate our method by answering the research question: can our address clustering method uncover more potential associated addresses than baseline methods? In order to assess the performance of our clustering method, we select three existing methods [3, 25, 47] as baseline methods for comparison. We use the association of addresses group held by 382 entities of the *Entity Identities* dataset to evaluate the quality of our address clustering method.

We measure clustering results from two aspects. First, we evaluate the number of identified entities, including the number of entities successfully identified (indicator N) and the number of

[2]Zipf's law is an empirical law that reveals the inversely proportional relationship between the rank of the word and its frequency in natural language utterances.

Table 4. Evaluation of address clustering methods.

Method	N	E	P	R	WP	WR
MI + NA	336	154	0.15	0.02	0.07	0.03
MX + NA + DP	339	96	0.43	0.09	0.18	0.13
MX + SP	355	37	0.80	0.60	0.28	0.20
Our method	**366**	**17**	**0.94**	**0.96**	**0.31**	**0.31**

entities incorrectly clustered into the same user (indicator E). Second, we assess the quality of addresses contained by each identified user through four indicators: *Precise* (P), (R), *Weighted Precise* (*WP*) and *Weighted Recall* (*WR*). The first two indicators are commonly used in the literature [7], while the last two indicators are newly introduced in our study. When some users are evaluated with the same value of precision or recall, a user with a larger number of addresses typically contains more information. That is, the user with a larger number of addresses can better describe its mapped Bitcoin user. Inspired by this, we consider the number of addresses per user as a weight to propose the latter two indicators.

$$P = \frac{\sum_{i=1}^{m} |O_i|}{\sum_{i=1}^{m} |S_i|}, \tag{1}$$

$$R = \frac{\sum_{i=1}^{m} |O_i|}{\sum_{i=1}^{m} |E_i|} \tag{2}$$

$$WP = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} \frac{|o_{ij}|}{|S_i|}, \tag{3}$$

$$WR = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} \frac{|o_{ij}|}{|E_i|} \tag{4}$$

$$\text{where } O_i = \bigcup_{j=1}^{m} o_{ij},\ o_{ij} = E_i \cap c_{ij},\ S_i = \bigcup_{j=1}^{n} c_{ij},\ w_{ij} = \frac{|c_{ij}|}{|S_i|}.$$

Equation 1 - Equation 4 introduce these four unified indicators, where m denotes the total number of entity and E_i denotes ith entity. Each E_i has n clusters and c_{ij} is its jth cluster. Based on these, we record the total clusters of E_i as O_i, the overlap between E_i and c_{ij} as o_{ij}, and the total overlap between them are calculated as O_i.

After filtering out isolated users, we assess the quality of the remaining users through the evaluation dataset. The results with two decimal places are summarized in Table 4. Note that the evaluation dataset also contains some transaction data of 2021 when compared to the dataset used in our previous work [20]. From Table 4, we can see that the new transaction data of 2021 has little impact on the clustering results, as the measurement results are almost the same as reported in [20]. In Table 4, we observe that our method can cluster more (95.81% of the total) known entities while having fewer over-clustered entities with an average reduction of 20.59%. The result confirms that applying a static method to mine different types of change addresses is somehow unpractical. Those three existing methods are more likely to mistakenly group unrelated addresses that actually belong to different entities into a single user. Instead, our method improves the performance of address clustering based on two special transaction patterns: peeling chain and locktime. In addition, the value of indicators P and R of our method exceed 90.00%, demonstrating better performance than other baseline methods. Compared with the best value of each indicator evaluated in the baseline methods, the indicators *P*, *R*, *WP* and *WR* have increased by 17.50%, 60.00%, 10.71% and 55.00%, respectively. To further verify the effectiveness of our address clustering method, we conduct three additional experiments. We extract data of three time periods from the *Transactions* dataset, i.e., 01/03/2009-01/01/2017, 01/03/2009-01/01/2018 and 01/03/2009-01/01/2019. According to the appearance time of Bitcoin addresses, we generate the corresponding validation dataset for each additional experiment from the entire evaluation dataset. The results remain similar to what presented in Table 4.

Discussion. With our address clustering method, we can capture sophisticated associations among Bitcoin addresses and cluster them into users with higher accuracy and lower over-clustering. It greatly enlarges the address size of known entities in our *Entity Identities* dataset. From this, the industry identifiers of entities in the *Entity Identities* dataset are mapped from the addresses to the corresponding users, including industry organizers or industry participants. These users are the basic units for industry identifier classification in the next section, which can effectively capture the rationality and interpretability of user industry information. And we use the clustering results to enrich our *Ransomware Activities* dataset in Section 7.1 to better portray ransomware activities.

6 INDUSTRY IDENTIFIER CLASSIFICATION

In order to discover the activity purpose of users, we classify industry identifiers of users based on their activity patterns. Active Bitcoin users can change their current activity and participate in another industry, or even perform multiple activities across various industries at the same time. Moreover, most Bitcoin users prefer to focus on one activity in a short period. In other words, Bitcoin users possess dynamic industry identifiers, and their major activity patterns are usually stable within a certain time period.

Motivated by these observations, we design a multi-label classification model to identify users' industry identifiers within a certain period (e.g., one week). Specifically, we construct a directed graph *User-Transaction* to describe the interactions between users, where each node represents a user and each directed edge represents the relationship of the transaction from the sender to the recipient. We record the timestamp and the number of bitcoins received by the recipient as annotations for each edge. Based on the graph, we extract the trading behaviors of users as activity patterns to train the model in Section 6.1. In order to improve the accuracy of industry classification, we use the labels of industry organizers and participants from the *Entity Identities* dataset to refine classification labels during the training of the multi-label model. After that, we focus on multi-identifier users and propose a quantitative method to determine their major industry in Section 6.2.

6.1 Multiple Industry Identifier Classification

Training data. As is well known, some special events usually bring significant impacts on the development of Bitcoin industries. For example, the closure of a famous exchange site *Mt.Gox* severely impacts the volume of transactions in Bitcoin, especially the Exchange industry [10]. Such events may lead to an imbalance in the volume of training data for a particular industry. In order to improve the robustness of our model, for each industry, we select one milestone event from Google Trends [18] and construct the *User-Transaction* graph as a temporal network to extract user activity patterns based on the transactions before and after each event.

Table 5 lists the detail of each temporal network. We select the starting point of each temporal network from three aspects: 1) the number of Bitcoin addresses, 2) the volume of transactions, and 3) the value of transactions. The index 1 and index 2 in the time window name represent the temporal network before and after the event. For example, SatoshiDice1 denotes the temporal

Table 5. Temporal networks with five milestone events.

Event	Temporal Window	Name
Game *SatoshiDice* released.	04/01/2012-04/07/2012 05/22/2012-05/28/2012	SatoshiDice$_1$ SatoshiDice$_2$
Service *Liberty Reserve* unsealed.	04/04/2013-04/10/2013 05/28/2013-06/03/2013	Liverty$_1$ Liverty$_2$
Market *SilkRoad* shut down.	09/18/2013-09/24/2013 10/04/2013-10/10/2013	SilkRoad$_1$ SilkRoad$_2$
Exchange *Mt.Gox* disappeared.	01/02/2014-01/08/2014 02/12/2014-02/18/2014	MtGox$_1$ MtGox$_2$
Miner pool *BTC Guild* announced the closure.	03/02/2015-03/08/2015 03/24/2015-03/30/2015	Guild$_1$ Guild$_2$

Table 6. Comparison of graph embedding algorithms in multi-label classification.

Algorithm	Macro-F1			Micro-F1		
	10%	**20%**	**30%**	**10%**	**20%**	**30%**
DeepWalk	0.62	0.72	0.76	0.90	0.91	0.92
GraphSAGE	**0.70**	**0.75**	**0.77**	**0.90**	**0.91**	**0.93**
LINE	0.33	0.35	0.35	0.81	0.81	0.81
Matrix Factorization	0.30	0.33	0.42	0.80	0.80	0.82
Node2Vec	0.47	0.52	0.62	0.84	0.87	0.88
SDNE	0.46	0.54	0.54	0.79	0.81	0.81

network before the release of *SatoshiDice* game, and SatoshiDice2 denotes the temporal network after the release of *SatoshiDice* game. The time span of the temporal network is a configurable parameter, and we set the parameter as seven days in our study.

Feature extraction. In each temporal network, the proportion of known industry identifier labels is rather limited (accounting for 10%-30% of the total users). In order to extract training features at such a known-label proportion, we test the performance of six representative graph embedding algorithms as discussed in [6], i.e, *GraphSAGE*, *DeepWalk*, *Node2Vec*, *LINE*, *Matrix Factorization*, and *SDNE*.[3] We apply a common one-vs-rest algorithm logistic regression to evaluate the performance of these graph embedding algorithms, randomly sampling 10%, 20% and 30% of the users with known industry identifier labels as the training data and the rest of labeled users as the testing data. To eliminate the contingency of results, we repeat this process ten times and calculate the average of *Macro-F1* and *Micro-F1*. The evaluation results are presented in Table 6.

We observe that the performance of *GraphSAGE* [19] is better than other algorithms and remains relatively stable in different sample proportions. Therefore, we use *GraphSAGE* to extract the features of users in each temporal network. We set its *learning rate* as 0.00001, and use graphsage_mean as the *aggregator*. For the other parameters of *GraphSAGE*, we use their default values.

[3]We are not restricting ourselves to these algorithms, and in the future we plan to apply state-of-the-art methods for graph representation learning.

Table 7. Model evaluation for each temporal network.

Temporal Network	Accuracy	Macro-F1	Micro-F1
SatoshiDice_1	0.92	0.88	0.94
SatoshiDice_2	0.96	0.89	0.97
Liberty_1	0.92	0.90	0.94
Liberty_2	0.91	0.88	0.94
SilkRoad_1	0.89	0.87	0.93
SilkRoad_2	0.90	0.89	0.93
MtGox_1	0.92	0.85	0.94
MtGox_2	0.94	0.87	0.95
Guild_1	0.93	0.87	0.94
Guild_2	0.91	0.88	0.94

Model training. Applying the Multi-Layer Perceptron (MLP) [4], we build a supervised multi-label classification model. We filter out the users who have participated in the transactions less than three times to ensure that the features of the remaining users are valuable to be trained. After that, we split the filtered data into 67% for training and 33% for testing and then adopt 3-fold cross-validation to obtain the best parameters for the model.

Model evaluation. We evaluate our model from three metrics: *Accuracy*, *Macro-F1* and *Micro-F1*. Table 7 list the results. We observe that our model presents relatively high accuracy, with an average of 92.00%. We consider the accuracy of our industry identifier classification is sufficient to conduct an industry-based empirical analysis in Section 7.

Consequently, our model classifies industry identifies for users and identifies 23.35% of them engage in multiple industries within a week. Such a non-negligible proportion of multi-identifier users motivates us to further identify the activity they are mostly involved in during the time period, i.e., to detect their major industry.

6.2 Major Industry Detection

To understand the primary activity purposes of users among these industries, we propose a quantitative method to determine their major industry identifier. Being an industry member, a user is active mainly inside the industry and rarely participates in other activities outside the industry. Therefore, if a user devotes more participation frequency and bitcoin traffic to a specific industry, we determine this industry as his/her major industry. Through the annotations of edges in the *User-Transaction* graph, we extract the participation frequency and the bitcoin traffic of intra-industry transactions as indicators to quantitatively determine the major industry. Based on the information entropy of indicators, we dynamically compute the weight of these indicators. Besides, we consider the probability of industry identifier predicted in Section 6.1 to help assign the major industry of multi-identifier users.

Extracting indicator data. For a user with multiple industry identifiers, we extract his/her transactions involved in a single industry as internal transactions and calculate two indicators from the internal transactions: participation frequency (f) and bitcoin traffic (v). These two indicators describe how often users engage in the activities of their current industry and how many bitcoins are used in each activity. Specifically, the participation frequency (f) denotes time frequency between the current transaction and the recent transaction conducted by the same user in the current industry, i.e., the reciprocal of time span between two consecutive internal transactions of the current industry. The bitcoin traffic (v) is defined as the total number of bitcoins used by the user in the current industry. Sometimes, the sender and recipient of a transaction are both multi-identifier users, so it is difficult to determine whether the current transaction is an internal transaction intuitively. Below, we present two heuristic rules to extract the indicators in this case.

Heuristic rule 3. When a multi-identifier user is a sender of a transaction, and at least one recipient of the transaction has the same industry identifier of the user, we calculate the time frequency between the current transaction and the recently conducted internal transaction by the user as the participation frequency (f) of the industry, and calculate the sum of bitcoins received by such recipients as the bitcoin traffic (v) of the industry.

Heuristic rule 4. When all recipients of a transaction hold at least one same industry identifier, and these same industry identifiers are also held by the multi-identifier sender, we calculate the time frequency between the current transaction and recently conducted internal transaction by the user as the participation frequency (f) of the industry, and calculate the sum of bitcoins received by the multi-identifier user among all recipients as the bitcoin traffic (v) of the industry.

With the above rules, we obtain *sequence F* and *sequence V* as the data sequences of these two indicators, referred to as F and V.

Calculating weights. We apply the entropy weight method (EWM) to calculate the weights of indicator f and indicator v. An indicator with higher entropy contains richer information and should be given more weight for the major industry calculation. We normalize data sequences of these two indicators, compute the entropy (e_F and e_V) and finally obtain the weight w defined by Equation 5.

$$w_j = \frac{1 - e_j}{\sum_{j \in \{F,V\}} (1 - e_j)} \tag{5}$$

Assigning major industry. We jointly assess the major industry of a user from his internal transaction behaviors and the prediction probability of his/her industry identifiers (p_i) in Equation 6. The quantification of user's internal transaction behavior is calculated from the average time frequency (m_{iF}) and the total sum of bitcoin traffic (m_{iV}). After that, we rank each industry by its score (s_i) and determine the industry with the highest score as the major industry.

$$s_i = p_i * \sum_{j \in \{F,V\}} m_{ij} w_j \tag{6}$$

As a result, we are able to identify 74.44% of users' activity purpose and their major industry within a given period. Therefore, we can reproduce the major activity trajectories of these users across the industries and study illegal activities from an industry perspective in the next section. In particular, the detection of major industry captures the representative behaviors of users with multiple industry identifiers, which helps us monitor how the victims react when involved in illegal activities.

7 RANSOMWARE ACTIVITY ANALYSIS

In this section, we analyze statistic trends and the impact of ransomware activities from the industry perspective, based on the results from Section 5 and Section 6. Although our methods for address clustering and industry identifier classification do not have a perfect performance, they still enable us to get a better understanding and reveal deep insights on ransomware activities in Bitcoin.

First, we identify ransom payment transactions of 63 ransomware activities from 2012 to 2021. Then, we quantify the amount of ransom and the number of victims in each ransomware activity.

Based on the above statistics, we choose seven typical ransomware activities, *CryptoLocker*, *CryptoWall*, *Locky*, *Cerber*, *CryptXXX*, *NoobCrypt* and *WannaCry*, as typical examples to analyze the ransom transfer patterns and victim migrations from the industry perspective. These ransomware activities involve different ransom payment requirements, particularly the way of receiving ransom from victims to criminals. More specifically, the criminals of *Locky* generate a new address for every victim as a unique ransom address to collect ransom, while the criminals of *WannaCry* ask multiple victims to pay bitcoins to the same ransom address.

7.1 An Analysis of Ransom Payment Transactions

As mentioned in Section 3, we collect 22,717 Bitcoin addresses of 63 ransomware families in our *Ransomware Activities* dataset and regard them as ransomware seed addresses. To find more Bitcoin addresses controlled by ransomware criminals, we unearth other Bitcoin addresses belonging to the same user as the ransomware seed address according to the address clustering results in Section 5. We find that the number of addresses of a few ransomware activities does not change significantly before and after the address clustering, such as *CryptoLocker* and *Locky*. A major reason is that a portion of addresses we collected for ransomware activities have been expanded by previous studies.

After that, we find some ransomware addresses were involved in other activities before the ransomware activity, which leads to misinterpretations of ransomware activities. To guarantee the reliability of the dataset, we filter out unrelated addresses by determining the starting date of each ransomware activity. We have used the Google trends to gain the date when people start searching for specific ransomware and consider this date as the beginning of a ransomware activity, because victims impacted by ransomware are likely to search online to get some valuable help. Then, for ransomware not found in the Google trends, we filter out transactions far away from the active trading period. As a result, we gain a total of 24,536 ransomware addresses containing ransomware seed addresses in our *Ransomware Activities* dataset and the expanded Bitcoin addresses.

Criminals often use multiple Bitcoin addresses to aggregate and transfer ransom, which causes double counting when calculating the amount of ransom. To address this problem, we divide ransomware addresses into two types based on the usage of ransomware addresses: i) *charge* address is used to receive ransom from victims. ii) *aggregate* address is used to aggregate ransom from multiple *charge* addresses for ransom transferring and laundering. In the actual analysis, we construct a transaction network with only ransomware addresses as nodes based on transaction records and then determine whether the address is a *charge* address or *aggregate* address according to the node's in-degree. More especially, the address nodes with the in-degree larger than 1 are the *aggregate* addresses and the others are considered as *charge* addresses.

Next, we extract victims from transactions that *charge* addresses participating in and precisely estimate the financial impact by the mutual validation of two types of Bitcoin addresses. First, we select transactions where the *charge* address is located in the output and consider input addresses of these transactions as victims. The amount of ransom gained by criminals can be calculated by summing bitcoins transferred from victims to *charge* addresses. Besides, we crawl the historical exchange price of Bitcoin and USD, and use the low price of the day to calculate the amount of ransom of each ransomware activity. Applying the above method, we find that there are 11 ransom activities with the amount of ransom less than $1 in our *Ransomware Activities* dataset. Fig. 2 shows the amount of ransom ($ and BTC) and the numbers of victims in the remaining 52 ransomware families. We find the largest number of victims pay the ransom in the ransomware activities *Locky* while *Zeppelin* receives the largest amount of ransom.

According to the above analysis results, we choose seven ransomware activities with a high amount of ransom and a large number of victims to analyze victims' reactions and ransom transfer

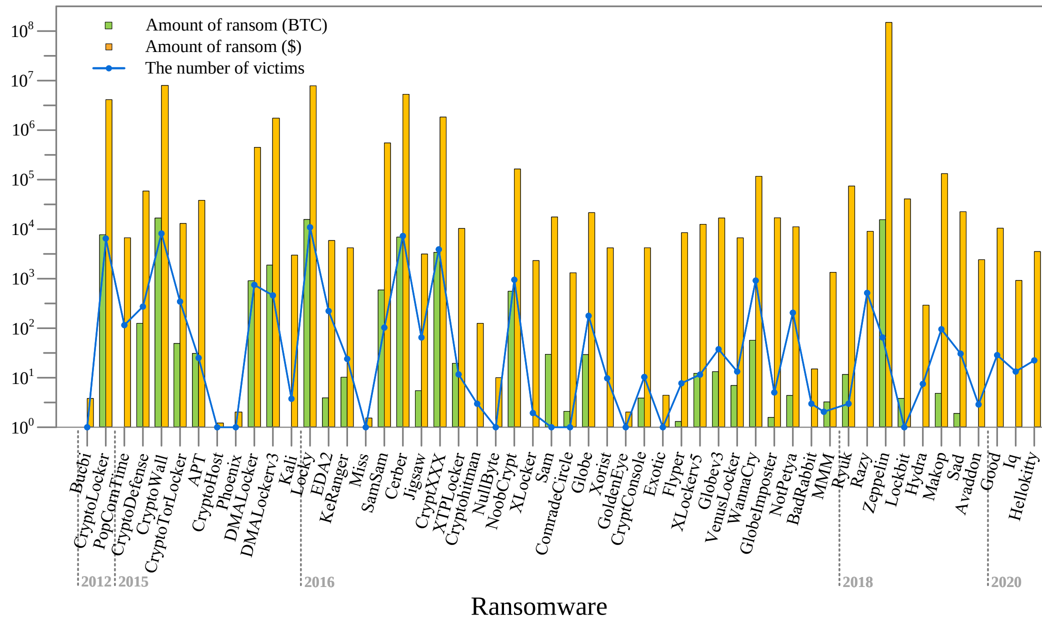


Fig. 2. The amount of ransom and number of victims in each ransomware activity.

patterns of criminals from the Industry perspective. These ransomware activities are *CryptoLocker*, *CryptoWall*, *Locky*, *Cerber*, *CryptXXX*, *NoobCrypt* and *WannaCry*. Although *Zeppelin* receives the largest amount of ransom, it targets healthcare systems and victims are no longer active in Bitcoin after paying ransom.

7.2 Ransom Tracking among Industries

It is crucial to understand the ransom transfer behaviors of criminals when analyzing ransomware activities in Bitcoin. We extract their relevant ransom payment transactions and estimate the total ransom received by each ransomware activity. The amount of ransom varies considerably in different ransomware activities. For example, the criminals of *Locky* receive ransom of 15351.44 bitcoins while *WannaCry* criminals totally obtain 55.80 bitcoins as ransom.[4]

We propose a money tracking model to track the transfer routes and destinations of ransom extorted from victims. The literature [1] introduces three mainstream money tracking algorithms in Bitcoin: *Poison*, *Haircut* and *FIFO*, which apply different strategies to identify money diffusion. The first two algorithms do not find an efficient tracking target from all the received bitcoins of a transaction and directly analyze all the transfer routes of these bitcoins. This may perform additional tracking of the extraneous bitcoins transferred in the transaction, resulting in high time and space complexity. Thus we develop our model based on *FIFO* using industry information to locate transfer routes of ransom.

Our model. In our money tracking model, we regard a transaction launched by the criminals as a source polluted transaction, and the bitcoins sent from the source polluted transaction as the source polluted money. Starting from the source polluted transaction, our model locates the corresponding

[4]The report [28] confirms that despite the large-scale attack in ransomware *WannaCry*, only a small number of victims have paid ransom to criminals.

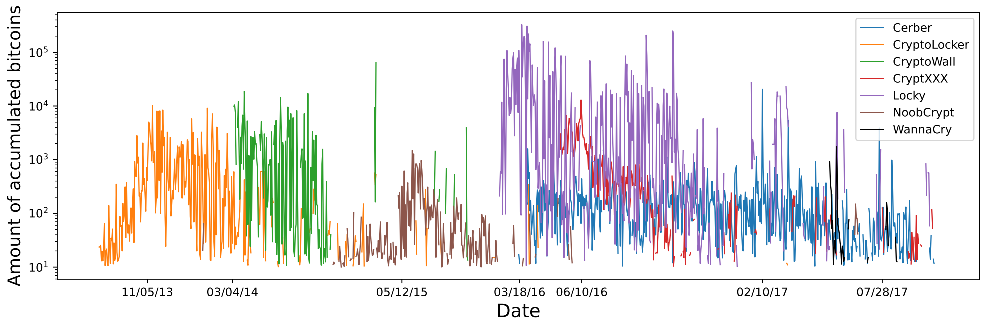


Fig. 3. Ransom transfer periods of six ransomware activities.

output positions through the *Spent_Tx* field recorded in the output. In each tracking step, we treat the polluted transaction A as a current polluted transaction and then extract its next consecutive transactions into a set as next polluted transactions. For each transaction in the set (referred to as transaction B), we explore the polluted state between transaction A and transaction B. Based on the idea of *FIFO*, we calculate the interval (*taint*) and the value (*value*) of the polluted money that flows from transaction A to transaction B. We then calculate the percentage of target polluted money (*ratio*) and identify the receiving industry of the polluted money (*position*). Thus we record the polluted state between transaction A and transaction B: [*transactionA*, *transactionB*, *taint*, *value*, *ratio*, *position*]. When the current tracking step is completed, every transaction of the next polluted transaction set will be performed as the current transaction in the next tracking step.

Applying this model, we track the transfer routes of polluted money which are transferred in a series of pollution transactions. To improve tracking efficiency, the tracking circulation from the source polluted transaction is limited to eight consecutive transactions, i.e., the length of the tracking step (a configurable parameter) is eight. In addition, we enforce several stop-tracking restrictions, such as the amount of polluted money being less than 0.0001 bitcoins, to filter out insignificant transfer branches. Our following analysis results reflect that the time span of tracking eight transactions has met the analysis needs. If we track more transactions, the result is not necessarily more accurate, because the ownership of the ransom may be transferred, and the time consumption will increase exponentially. If we track fewer transactions, it may be difficult for us to track the exact location of the ransom.

By analyzing a large number of transactions, we obtain ransom transfer trajectories of ransomware activities' criminals and study their transfer preferences.

Criminals' ransom transfer preferences. Based on the ransom transfer trajectories, we summarize the transfer preferences of criminals in two aspects: when they prefer to transfer ransom and how they transfer ransom.

Active transfer periods. We estimate the number of bitcoins transferred on a daily basis in seven ransomware activities and plot the distribution curve of bitcoin accumulation in Fig. 3. For example, we observe that the criminals of *Locky* actively transfer ransom from 02/19/2016 to 11/05/2016. The most frequently transfer ransom is carried out around 03/21/2016. The study [21] reports the active periods of ransom payments and concludes that the median holding time span for the ransom is 1.6 days. We find that the duration difference between their ransom payment period and the ransom transfer peak obtained in our analysis actually matches this holding time span, which demonstrates the accuracy of our money tracking model. This finding makes us confident that the following industry-based analysis is effective. Besides, the period of time that *WannaCry* actively transfer

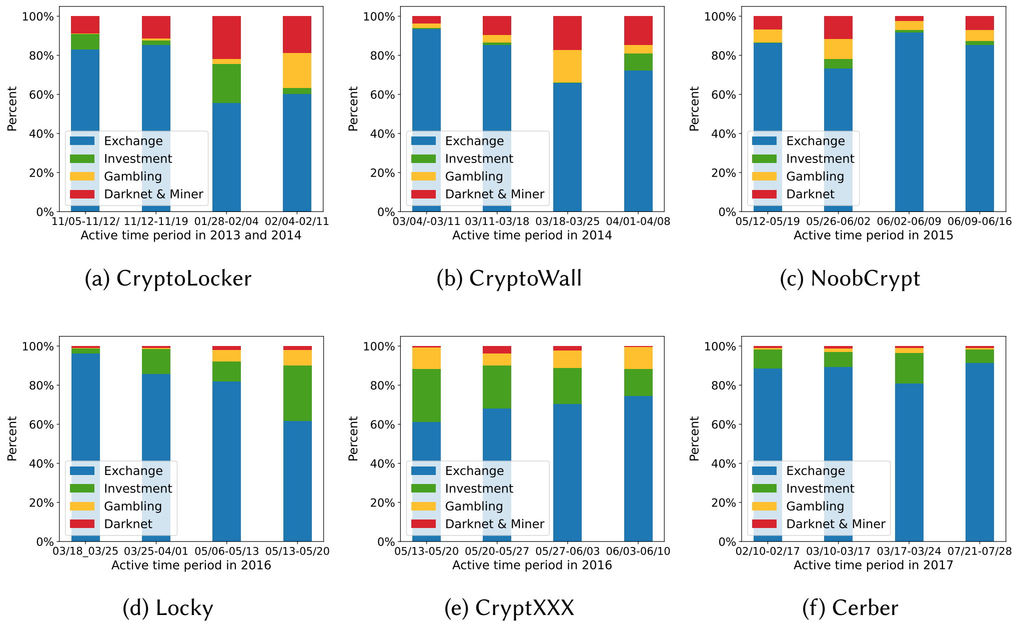


Fig. 4. Distribution of ransom transfers across different industries in different ransomware activities.

ransom is relatively short, so we don't illustrate ransom transfers and victim migrations in the following presentation.

Transfer patterns. With the help of industry identifiers, we reproduce the criminals' transfer trajectories of seven ransomware activities and then summarize their transfer patterns during several active transfer periods.

We find that in order to hide the purpose of ransom transfers and their true transfer destination, criminals prefer to transfer ransom across industries rather than relying on Bitcoin mixers. More specifically, *CryptoLocker* in 2013 uses the service of *Bitcoin Fog* to transfer the most bitcoins among seven ransomware activities, with 65 bitcoins, but only accounting for 0.5% of the total ransom it received. The subsequent ransomware activities use the Bitcoin mixers less and less. Criminals of *CryptoWall* use the services of *Bitcoin Fog* to transfer 2.37 bitcoins in 2014. The criminals of *Locky* use the service of *HelixMixer* and *Bitcoin Fog* to transfer only 0.33 bitcoins in 2016. The small amount of ransom indicates that the criminals no longer use the Bitcoin mixers as their primary way for money transfer. This is most likely due to the strict authentication requirements of these Bitcoin mixers.

Fig. 4 presents the distribution of ransom transfers across different industries in different ransomware activities except *WannaCry*, covering several ransom transfer peaks in Fig. 3. As proportions of ransom move to the Miner industry are at a small value, we combine the ransom from the Darknet industry and the Miner industry in this figure. Through the amount of ransom flowing to various industries, we regard the Exchange industry as the most active industry for ransom transfers. Taking *Locky* as an example, in each period, the Exchange industry diverts over 2,000 bitcoins while the bitcoins flowing into the Miner industry and the Darknet industry is always less than 35 bitcoins.

Because *Locky's* influence is very large, we select the period from 03/25/2016 to 04/01/2016 to analyze it in detail. In this period, 82.10% of ransom flows into the Exchange industry, 16.18% of ransom is received by the Investment industry. Such a large proportion of ransom transferred into the Exchange industry motivates us to further analyze the roles of its users involved in bitcoin remittances. We observe that most (92.43%) of the bitcoins move among Exchange participants (i.e., exchange buyers mentioned in Section 2) while the transactions directly conducted with famous Exchange organizers (i.e., exchange sites mentioned in Section 2) such as *Poloniex.com* and *Cex.io* are less frequent. The transactions among Exchange participants appear to be normal currency exchange activities, obscuring the true intent that it is essentially a ransom transfer. After moving the ransom to another participant, over 36.50% of these participants are no longer involved in transactions and leave Bitcoin. In other words, more than one-third of these participants act as ransom transfer proxies. Ransomware criminals usually do not use *charge* addresses to directly withdraw in the cryptocurrency exchange, but they instead use these proxies to cover up their withdrawal operations. Combined with the ransom transfer analysis of other ransomware activities in Fig. 4, we thus conclude that the preferred transfer pattern for ransomware criminals is to spread ransom across different industries instead of relying on Bitcoin mixers. In particular, the distribution of most bitcoins are completed through normal currency exchange services among the participants in Exchange, thereby alleviating the attention of regulators to these unusual behaviors.

Important transfer destinations. In addition to analyzing the transfer trajectories of ransom, we further pay attention to the engagement of some well-known entities in remittance destinations. We observe that *BTC-e*, *Localbitcoins*, *Bitstamp*, *Satoshi Mines*, *SilkRoad2Market* are the most active entities, where the first three are prevailing exchange sites, the fourth entity acts as a gambling banker and the fifth entity is a darknet vendor. For instance, *CryptoWall* transferred around 2,408.82 bitcoins through *BTC-e*, and *CryptoLocker* transferred around 484.23 bitcoins through *Bitstamp*. The previous studies [21, 40] verify the engagement of these first two exchange sites and state that *BTC-e* is widely applied for money laundering.

As Seunghyeon et al. [31] point out that the main activities of darknet are usually intended to provide illegal services, we infer that the ransom flowing into the Darknet industry are more likely used for other illegal behaviors. To verify our assumption, we continue to track the activity trajectories of criminals. And we detect that *CryptoLocker* and *CryptoWall* criminals participated in the illegal trading in *SilkRoad2Market* and *Locky* criminals participated in the illegal trading in *Nucleus Market*, which is a darknet market for the sale of drugs and other contraband.

7.3 Victim Movement among Industries

Apart from understanding the ransom transfer behaviors of criminals, we further monitor the migrations of victims across various industries to understand how victims react to the ransomware activities and quantify ransomware activities' impact on various industries. We first identify the victims of ransomware activities from ransom payment transactions and then propose a user movement model to describe their migrations across industries.

Ransom payment preference of victims. We detect two interesting phenomena in these ransomware activities. First, we observe that the ratio of ransom payment transactions that specify the effective time through the field of *Locktime* increases yearly. In detail, 0.1% of ransom payment transactions in *CryptoLocker* specify the effective time in 2013. 0.38% of ransom payments transactions in *Locky* specify the effective time in 2016 while 31.12% of the transactions in *WannaCry* specify the effective time in 2017. The change indicates that ransomware criminals learn to require victims to specify the effective time to collect and transfer ransom conveniently. Second, we observe that 80.71% of the victims in all ransomware activities have only one receiving transaction and

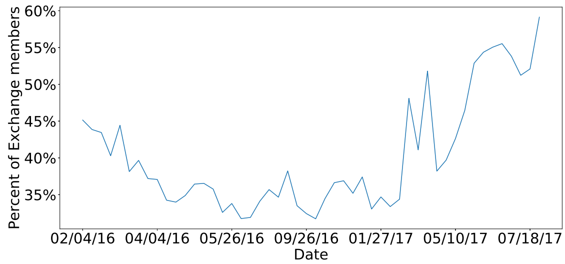


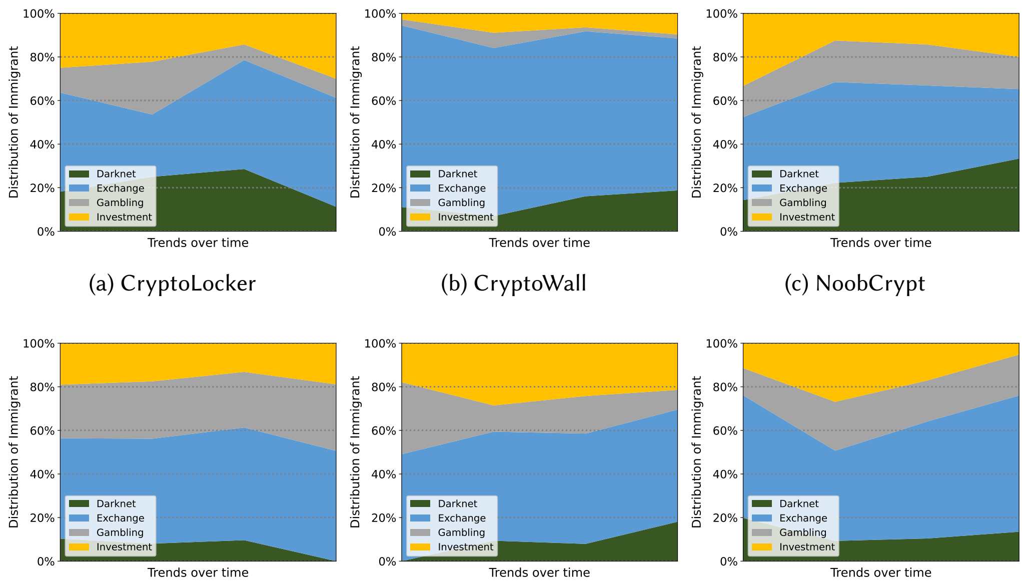
Fig. 5. Proportion of Exchange members among all industry members.

one sending transaction in their historical transactions. More specifically, after the victims were extorted, they enter into Bitcoin to buy the bitcoins as ransom from the only receiving transaction and then send it to the criminals. We consider these victims as one-time victims and the rest as frequent victims. Besides, two active exchange sites *Bitcoin.de* and *Localbitcoins*, who are detected as the victims, have made a total of 25 payments in *Locky* activity, accumulating more than 86 bitcoins. Inspired by the work [21], we speculate that these exchange sites provide victims with proxy payment services to help complete their ransom payments.

Our model. As discussed in Section 6, users can engage in a variety of activities, meaning that users can join an industry as a new member, move to other industries or even leave Bitcoin. To understand the migration of victims in ransomware attacks, we utilize a user movement model to describe their time-varying participation among the industries. Based on the variation in users' industry identifiers over two periods, we propose four user migration modes – (1) *Immigrant*: a user who newly comes into the industry at the latter period, (2) *Emigrant*: a user who leaves /her current industry to join another industry different from these five industries at the latter period, (3) *Migrant*: a user who moves from one industry to another industry, and (4) *Nonimmigrant*: a user who always stays in the same industry. By applying the user movement model to the ransomware victims, we aim to understand how victims react to ransomware activities and the impact of ransomware activities on the entire Bitcoin ecosystem.

Impact of ransomware activity. Based on the ransom transfer peaks and the relative number of searches in Google trends, we focus on victim migrations in different periods, such as the migrations in *CryptoLocker* from 11/15/2013 to 12/17/2013 and the migration in *Locky* from 03/01/2016 to 04/04/2016. We did not identify any victims in the Miner industry through our classification model. Thus, under the entire distribution of victims, we analyze the status of other industries during the ransomware activity while ignoring the impact of potential victims in Miner.

Immigrants and Emigrants. Fig. 6 shows that the distribution of immigrant victims across different industries in different ransomware activities except *WannaCry*. We find most frequent inflows and outflows of victims occur in the Exchange industry. To investigate the reason for such high-frequency movements in Exchange, for these victims, we compare the time they pay ransom with the time of their first or last participation in the transactions. The difference in timing can suggest whether the major activity of victims is solely for the ransom payment. For example, we identify 87.43% of *Locky* victims and 61.78% of *WannaCry* victims temporarily join Bitcoin due to ransomware activities. Moreover, the decreasing proportion between these two ransomware activities indicates that more victims have joined Bitcoin and engaged in exchange activities prior to being extorted by the ransomware in 2017. In other words, fewer victims are forced to enter Bitcoin for ransom payments, as many of them are already regular users of Bitcoin. Inspired by this change in victims, we presume that more users were active in Exchange between 2016 and 2017.



(a) CryptoLocker (b) CryptoWall (c) NoobCrypt

(d) Locky (e) CryptXXX (f) Cerber

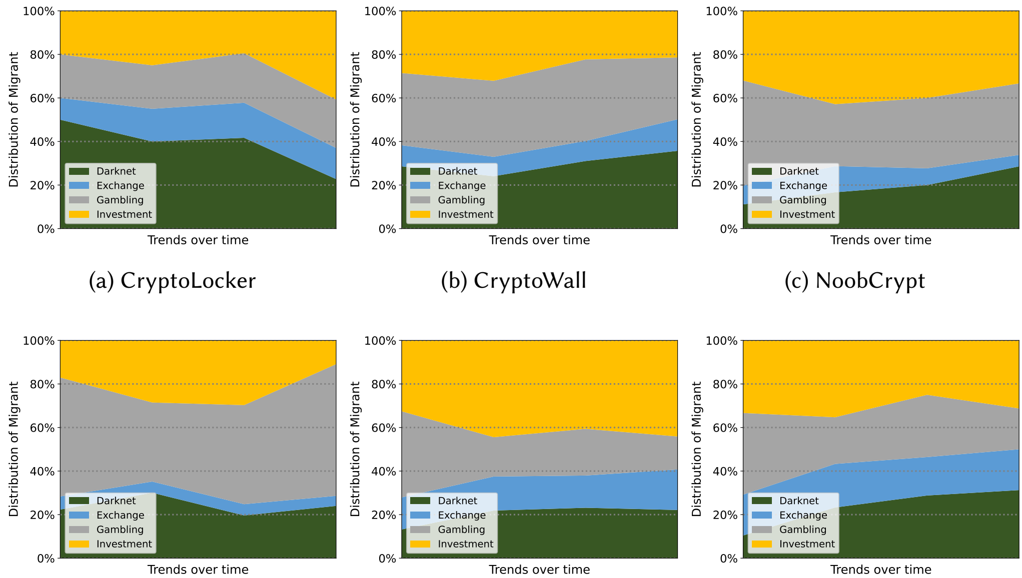
Fig. 6. Distribution of immigrant victims across different industries during the high-risk period.

Fig. 5 presents the proportion of Exchange members among the five industries, demonstrating that the Exchange industry has expanded over this period of time.

Fig. 6 also shows that in all ransomware activities, a certain percentage of victims newly entered the Darknet industry and the Gambling industry after paying ransom. It is worth noticing that this distribution is not calculated based on all the victims, but the victims who newly entered the five industries during the high-risk time period. Combined with the report [9], we speculate that these immigrants are likely to be induced by the criminals to start their illegal activities in the Darknet.

Migrants. Depending on the different magnitude of victims in the several ransomware activities, we focus on the overall trends of migrants in six ransomware activities and explore the specific migration routes of victims in *WannaCry*. In Fig. 7, we present the proportion of migrants across different industries within several periods. The change in area size can reflect the stability of migration across various industries. We clearly see the six ransomware activities has impacted all four industries and drives the members to migrate to other industries. Of these migrants, the Exchange industry holds a small proportion of migrants and has been stable over various periods. However, the other three industries exhibit fluctuating changes during the periods of ransomware activity and can return to their pre-extortion states.

In addition, we investigate the migrant routes of *WannaCry* victims and find that the Investment industry is relatively stable in this ransomware activity. We observe that many victims move from the Investment industry to Exchange. To obtain bitcoins required for the ransom payment, 71.43% of them leave the Investment industry and then trade with well-known exchange sites such as *localbitcoins.com* in Exchange. We infer that the migration route from Investment to Exchange is primarily due to the fact that victims have to purchase bitcoins through exchange sites to pay ransom. After completing the ransom payment, about two-thirds of these victims eventually return to the activities in the Investment industry. Accordingly, we find that the Investment industry



(a) CryptoLocker (b) CryptoWall (c) NoobCrypt

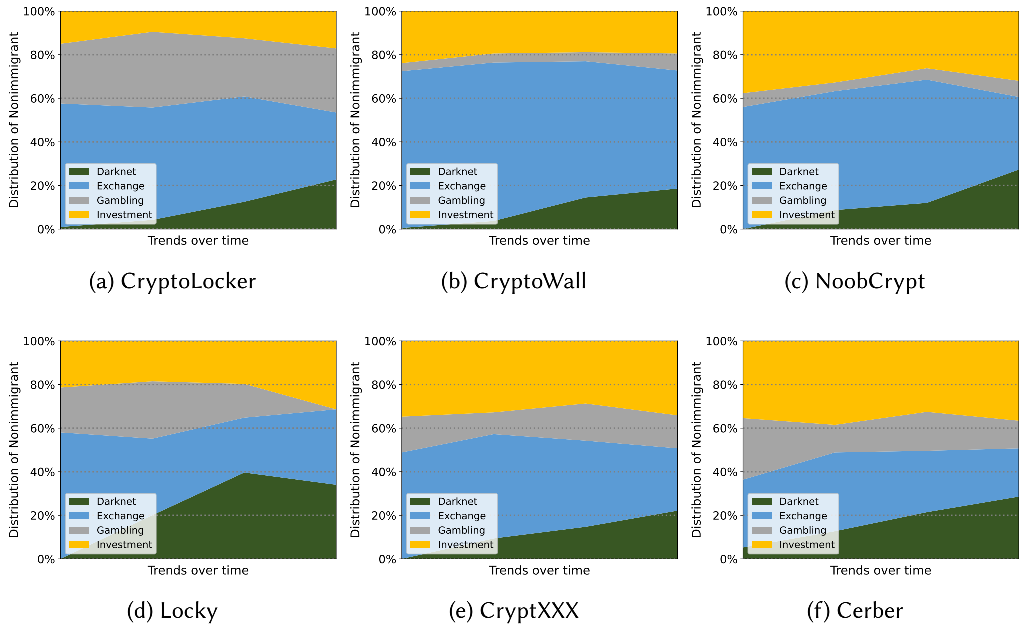
(d) Locky (e) CryptXXX (f) Cerber

Fig. 7. Distribution of migrant victims across different industries during the high-risk period.

appears to be highly resilient, with more than half of the users who left the industry for a short time period would return to the industry and remain active. In other words, the industry presents a strong self-repair ability. To confirm the finding, we trace the industry identifiers of those users before and after the migration period. More than 31.30% of users always stay with Investment. It is likely because the services such as wallet management offered in the Investment industry are particularly attractive to users. Besides, the rest (30.77%) of these victims then engage in the Gambling industry. Betting in the Gambling industry becomes another choice for these victims after they have paid.

Apart from the frequent migration trajectories as described above, it is also surprising that a few victims subsequently participated in the Darknet industry after completing ransom payments. More specifically, 8.82% of these victims migrate from the Exchange industry to Darknet. We observe three specific victims have been involved in 66 transactions with seven well-known darknet vendors, including *PandoraOpenMarket* and *SilkRoad2Market*. All of these transactions are launched by the same user within a very short time period and follow the same pattern that consists of one input address and 740 output addresses. Although these darknet vendors were shut down by regulators when these transactions were launched, we conjecture that the continued trading activities with these darknet vendors are likely to be launched for other illegal purposes. Due to the lack of publicly available data, we cannot find more details about their behaviors in Darknet. However, the finding still makes us hypothesize that the behaviors of criminals can somehow induce the victims to engage in other illegal activities.

Nonimmigrants. We further analyze the nonimmigrants trend of each industry in six ransomware activities (Fig. 8). The proportion of nonimmigrants in Investment remains relatively stable in each ransomware activities, such as an average of 22.74% in *Locky* and 13.54% in *CryptoLocker*. With respect to the Gambling industry, the continuous extortion makes the proportion of nonimmigrants



(a) CryptoLocker (b) CryptoWall (c) NoobCrypt

(d) Locky (e) CryptXXX (f) Cerber

Fig. 8. Distribution of nonimmigrant victims across different industries during the high-risk period.

is small and constantly decreased by 24.53% in *Locky*. These phenomena indicate that the Investment industry is more resilient than the Gambling industry. Interestingly, the Darknet industry presents a rapid growth in all ransomware activities. For example, the Darknet finally accounts for 41.03% of nonimmigrants in the four industries in *Locky*. The growth reflects that many victims remain in the Darknet industry for other illegal purposes. When combined with the finding drawn from *WannaCry* activity, we would suggest that regulatory authorities should highlight the behaviors of victims in other ransomware activities, in addition to focusing on criminals' behaviors, which may help detect other potential *darknet* members.

To summarize, the industry-based analysis in Section 7 provides an effective way to understand ransomware activities in Bitcoin. We tracked over $176 million in ransom payments made by 41,424 victims from 2012 to 2021. Through the large-scale empirical analysis, we detect the ransom transfer patterns of ransomware criminals, analyze victims' migrations and study the impact of ransomware activities on various industries. The findings provide regulators with deep insights into the ransomware activities and advise them to adopt suitable policies to reduce the negative impact of ransomware activities.

8 DISCUSSION

Exploiting the anonymous mechanism of Bitcoin, ransomware activities collect ransom from worldwide without worrying about being tracked, causing substantial monetary losses. An improved understanding of ransomware activities is a key step to identifying new and effective intervention strategies. Based on our analysis results, this section outlines our key findings and their significance.

First, we found that only a few ransomware activities succeeded in collecting ransom payments worth millions, such as *CryptoLocker*, *CryptoWall* and *Locky*. More than half of ransomware activities in our dataset were responsible for less than USD 10,000 of direct financial impacts. Kharraz et al. [29] studied 1,359 samples of 15 ransomware families and Gazet [16] reversed-engineered 15 ransomware samples. They both found that most ransomware families used superficial and flawed techniques to encrypt files. Few of ransomware families had actual destructive capabilities and most of them could be easily defeated. This could explain why only few ransomware families succeeded in generating ransom payments worth millions. However, such observations do not mean that the threat of ransomware activities should be underestimated. As noted by Zhang-Kennedy et al. [49], ransomware activities have severe technological, personal and social impacts on victims.

Second, we found that to hide their transfer trajectories, most ransomware criminals prefer to spread ransom into multiple industries instead of utilizing the services of Bitcoin mixers. Ransomware criminals only transfer a small part of the ransom to the services of Bitcoin mixers, such as 0.5% in *CryptoLocker*. This result is similar to the previous work of Huang et al. [21], which discovered that $541,670 (6.8% of Cerber's total outflows) was sent to BitMixer. There are two main reasons: the cumbersome operation and the lack of trust, as Crawford et al. [13] showed that mixing services far more often fail due to the inability to earn customers than due to law enforcement action.

Third, we observed that a few victims enter into Darknet industry after paying the ransom. Combined with the previous work [34], we speculate that these victims are likely to be induced by criminals to purchase the ransomware in the Darknet. This assumption is somehow confirmed by Kerstens et al. [27] and Jennings et al. [24], who claimed that the victimization experience can produce negative physical, mental, and behavioral outcomes in individuals and some may go on to commit their own crimes. Moreover, we found that Investment is highly resilient to ransomware activities in the sense that the number of users in Investment remains relatively stable.

Although our work has revealed several interesting findings, it also has several limitations. The main limitation is the small number of addresses controlled by ransomware criminals. While the website [42] publishes more than 1,000 kinds of ransomware families it detects from samples of the malware or suspicious files, our work collects relevant Bitcoin addresses of *only* 63 ransomware families. Although ransomware families in our dataset are rampant and have caused substantial financial losses, they only represent a small part of the entire ransomware landscape. Indeed, the more addresses from various ransomware families become available, the more accurate the landscape for ransom payments, ransom transfers and victim migrations will become. Another limitation is the scale and quality of the attribution data available in our *Entitiy Identities* dataset. Without this information, we cannot locate the real-world destination of ransom and victims. Nevertheless, we believe that such data will increasingly become available with the growing popularity of various analytics tools. The last limitation is that some ransomware victims have not paid the ransom. Thus, we cannot measure the indirect impact of ransomware on these victims from ransom payment transactions in Bitcoin.

9 RELATED WORK

Our study is closely related to the literature on Bitcoin address clustering and ransomware activities analysis in Bitcoin.

Bitcoin address clustering. The anonymity property of Bitcoin makes it difficult to determine the ownership of multiple Bitcoin addresses. Several methods are proposed to cluster associated Bitcoin addresses by utilizing heuristics.

Multi-input grouping methods. Cazabet et al. [7] point out that the original multi-input grouping method [33] (i.e., the *MI* method as mentioned in Section 5) has a relatively low recall for users. On the basis of *MI*, Kalodner et al. [25] propose to reduce clustering interference caused by a special type of mixing transactions, i.e., *CoinJoin*. Recently, this new method has been widely applied in address clustering, which we refer to as *MX* in Section 5. However, the improved method only solves excessive clustering introduced by a certain type of mixing transactions without mining the potential association of Bitcoin addresses from these transactions, which somehow reduces the recall of clustering results. Motivated by these problems, we additionally consider the interference of a new type of mixing transactions (i.e., *JoinMarket*) and further detect strong associations of Bitcoin addresses from these two special types of mixing transactions, which improves the performance of address clustering as part of our industry-based analysis approach.

Change address detection methods. A few methods [25, 47] with different patterns are proposed to study the association of output addresses. These patterns have been extended to analyze the anonymity of other cryptocurrencies and help cluster their associated addresses, such as Zcash [26] and Ripple [35]. However, some transactions in Bitcoin may mismatch these patterns and result in incorrect Bitcoin associations. For example, the new address rule (*NA* as mentioned in Section 5) considers the new address in the outputs of a two-output transaction as the change address of the sender. The new output address in the ransom payment transactions of *Locky* is actually the unique ransom address generated by the criminals for every victim, rather than a change address of a victim [21]. To mitigate this problem, we propose a fine-grained address clustering method, which improves both precision and recall a lot in address clustering.

Ransomware activities analysis. Exploiting the anonymous mechanism of Bitcoin, ransomware activities demanding ransom in bitcoins have become rampant in recent years. To deeply understand ransomware activities, many studies utilize publicly available Bitcoin transaction records to analyze ransom payment transactions and track ransom transfers.

BitIodine is a Bitcoin forensic analysis framework that is used to perform a payment analysis for the *CryptoLocker*. Liao et al. [32] perform an expanded analysis of *CryptoLocker* and find evidence that suggests connections to *Bitcoin Fog* and *BTC-e*. Huang et al. [21] use 16 famous ransomware data to describe the development of ransomware activities and the geographic location of the victims. Conti et al. [12] report the financial impact of 20 ransomware from the Bitcoin payment transaction. Paquet-Clouston et al. [40] analyze ransom payment transactions related to 35 ransomware activities and find that the amount of ransom payments has a minimum value worth of 12,768,536 USD (22,967.54 bitcoins). Nerurkar et al. [38] engineer nine features to train the model for segregating 16 different licit-illicit categories of users, such as ransomware operators.

However, all of these works focus on the behaviors of individual addresses or clustered address set. Besides, most previous studies ignore victims' reactions in Bitcoin after paying ransom. In our analysis, we introduce the concept of industry in Bitcoin and perform a large-scale empirical analysis of ransom payments, ransom transfers, and victim migrations from both address and industry perspectives.

10 CONCLUSION AND FUTURE WORK

This paper performed the first large-scale empirical analysis of ransomware activities in Bitcoin over a long period from an industry perspective, which views Bitcoin as an economic society. For our analysis, we have designed a novel and effective address clustering method to mine associated addresses to users, which improves over existing methods on average 17.50% in precision and 60.00% in recall. Based on this result, a multi-label classification model was then designed to identify the industry identifiers of users with the accuracy of 92.00%. In our in-depth study, we have tracked over $176 million in ransom payments made by 41,424 victims from 2012 to 2021 and proposed an industry-based approach to analyze ransom transfer patterns and victims' reactions. Through the industry participation trajectories of users, we observed a popular way of ransom laundering that does not rely on Bitcoin mixers. Besides, we also found out that a few victims became active in Darknet after paying ransom and the Investment industry is highly resilient to ransomware activities. These results showed that our empirical analysis from the industry perspective can offer regulatory authorities a macro-level view to understand ransomware activities in Bitcoin effectively.

In practice, our work has successfully cracked a series of online ransomware activity cases using Bitcoin as a payment method.[5] In the future, we will study more transaction patterns to find more Bitcoin addresses controlled by ransomware criminals. We will further refine the industry classification method to better characterize ransom transfers and victim migrations. We also plan to apply our approach to analyzing other types of illegal activities, e.g., Ponzi scheme.

ACKNOWLEDGMENTS

This paper has been supported by Natural Science Foundation of China (U1836207), the National Key R&D Program of China (2019YFE0103800), and China Postdoctoral Science Foundation (2020M670998). We thank Mr. Shize Chen, Zhijie Fan for their contributions. The corresponding author of this paper is Weili Han.

REFERENCES

[1] Mansoor Ahmed, Ilia Shumailov, and Ross Anderson. 2018. Tendrils of Crime: Visualizing the Diffusion of Stolen Bitcoins. In *Proceedings of the 5th Workshop on Graphical Models for Security*. Springer, 1–12.

[2] Cuneyt Gurcan Akcora, Yitao Li, Yulia R. Gel, and Murat Kantarcioglu. 2020. BitcoinHeist: Topological Data Analysis for Ransomware Prediction on the Bitcoin Blockchain. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*. ijcai.org, 4439–4445.

[3] Susan Athey, Ivo Parashkevov, Vishnu Sarukkai, and Jing Xia. 2016. Bitcoin pricing, adoption, and usage: Theory and evidence. *Stanford institute for Economic Policy Research* 13, 4 (2016), 675–746.

[4] Christopher M Bishop et al. 1995. *Neural networks for pattern recognition.* Oxford University Press.

[5] Stefano Bistarelli, Matteo Parroccini, and Francesco Santini. 2018. Visualizing Bitcoin Flows of Ransomware: WannaCry One Week Later. In *Proceedings of the 2nd Italian Conference on Cyber Security (ITASEC)*. CEUR-WS.org.

[6] Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang. 2018. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering* 30, 9 (2018), 1616–1637.

[7] Rémy Cazabet, Rym Baccour, and Matthieu Latapy. 2017. Tracking Bitcoin Users Activity Using Community Detection on a Network of Weak Signals. In *Proceedings of the 6th Conference on Complex Networks and Their Applications (Complex Networks)*. Springer, 166–177.

[8] Weili Chen, Jun Wu, Zibin Zheng, Chuan Chen, and Yuren Zhou. 2019. Market Manipulation of Bitcoin: Evidence from Mining the Mt. Gox Transaction Network. In *Proceedings of the 38th IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 964–972.

[9] Catalin Cimpanu. 2017. Ransomware is now big business on the dark web and malware developers are cashing in. https://www.zdnet.com/article/ransomware-is-now-big-business-on-the-dark-web-and-malware-developers-are-cashing-in/.

[10] Matt Clinch. 2014. Bitcoin plummets 20% after major exchange halts withdrawals. https://www.cnbc.com/2014/02/07/bitcoin-plummets-20-after-major-exchange-halts-withdrawals.html.

[11] CNN. 2021. US offers up to $10 million reward for information on cyberattacks against critical infrastructure by foreign states. https://www.cnn.com/2021/07/15/politics/us-state-department-reward-cyberattacks/index.html.

[12] Mauro Conti, Ankit Gangwal, and Sushmita Ruj. 2018. On the economic significance of ransomware campaigns: A Bitcoin transactions perspective. *Computers & Security* 79 (2018), 162–189.

[13] Jesse Crawford and Yong Guan. 2020. Knowing your Bitcoin Customer: Money Laundering in the Bitcoin Economy. In *Proceedings of 13th International Conference on Systematic Approaches to Digital Forensic Engineering (SADFE)*. IEEE, 38–45.

[5] In early 2020, by leveraging our approach, we practically supported a network regulatory department in Zhejiang Provance, P. R. China, to combat a series of cyber crimes successfully. With the help of our approach, the department effectively identified and arrested the criminals of these extortion cases, where the criminals maliciously encrypted important documents of organizations, then forced their owners to transfer bitcoins as ransom to specific addresses, as a pre-condition of decrypting these important files.

[14] Jean-Charles Delvenne, Renaud Lambiotte, and Luis EC Rocha. 2015. Diffusion on networked systems is a question of time or structure. *Nature Communications* 6, 1 (2015), 1–10.

[15] Sean Foley, Jonathan R Karlsen, and Tālis J Putniņš. 2019. Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? *The Review of Financial Studies* 32, 5 (2019), 1798–1853.

[16] Alexandre Gazet. 2010. Comparative analysis of various ransomware virii. *Journal in computer virology* 6, 1 (2010), 77–90.

[17] Michelle Girvan and Mark EJ Newman. 2002. Community structure in social and biological networks. *National Academy of Sciences* 99, 12 (2002), 7821–7826.

[18] Google. 2021. Google Trends. https://trends.google.com/trends/.

[19] William L. Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*. Curran Associates, Inc., 1024–1034.

[20] Weili Han, Dingjie Chen, Jun Pang, Kai Wang, Chen Chen, Dapeng Huang, and Zhijie Fan. 2021. Temporal Networks Based Industry Identification for Bitcoin Users. In *Proceedings of the 16th International Conference on Wireless Algorithms, Systems, and Applications (WASA)*. Springer, 108–120.

[21] Danny Yuxing Huang, Maxwell Matthaios Aliapoulios, Vector Guo Li, Luca Invernizzi, Elie Bursztein, Kylie McRoberts, Jonathan Levin, Kirill Levchenko, Alex C. Snoeren, and Damon McCoy. 2018. Tracking Ransomware End-to-end. In *Proceedings of the 39th IEEE Symposium on Security and Privacy (S&P)*. IEEE, 618–631.

[22] Aleš Janda. 2021. CoinJoinMess. https://www.walletexplorer.com/wallet/CoinJoinMess.

[23] Aleš Janda. 2021. WalletExplorer.com. https://www.walletexplorer.com.

[24] Wesley G Jennings, Alex R Piquero, and Jennifer M Reingle. 2012. On the overlap between victimization and offending: A review of the literature. *Aggression and Violent behavior* 17, 1 (2012), 16–26.

[25] Harry A. Kalodner, Malte Möser, Kevin Lee, Steven Goldfeder, Martin Plattner, Alishah Chator, and Arvind Narayanan. 2020. BlockSci: Design and applications of a blockchain analysis platform. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security)*. USENIX Association, 2721–2738.

[26] George Kappos, Haaroon Yousaf, Mary Maller, and Sarah Meiklejohn. 2018. An Empirical Analysis of Anonymity in Zcash. In *Proceedings of the 27th USENIX Security Symposium (USENIX Security)*. USENIX Association, 463–477.

[27] Joyce Kerstens and Jurjen Jansen. 2016. The victim–perpetrator overlap in financial cybercrime: Evidence and reflection on the overlap of youth's on-line victimization and perpetration. *Deviant Behavior* 37, 5 (2016), 585–600.

[28] Arjun Kharpal. 2017. Hackers who infected 200,000 machines have only made $50,000 worth of bitcoin. https://www.cnbcafrica.com/technology/2017/05/16/hackers-made-50000-worth-bitcoin/.

[29] Amin Kharraz, William K. Robertson, Davide Balzarotti, Leyla Bilge, and Engin Kirda. 2015. Cutting the Gordian Knot: A Look Under the Hood of Ransomware Attacks. In *Proceedings of 12th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*. Springer, 3–24.

[30] AO Kaspersky Lab. 2021. WannaCry: Are you safe? https://www.kaspersky.com/blog/wannacry-ransomware/16518/.

[31] Seunghyeon Lee, Changhoon Yoon, Heedo Kang, Yeonkeun Kim, Yongdae Kim, Dongsu Han, Sooel Son, and Seungwon Shin. 2019. Cybercriminal Minds: An investigative study of cryptocurrency abuses in the Dark Web. In *Proceedings of the 26th Conference on Annual Network and Distributed System Security Symposium (NDSS)*. ISOC, 1–15.

[32] Kevin Liao, Ziming Zhao, Adam Doupé, and Gail Joon Ahn. 2016. Behind Closed Doors: Measurement and Analysis of CryptoLocker Ransoms in Bitcoin. In *Proceedings of 2016 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 1–13.

[33] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, and Stefan Savage. 2013. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference (IMC)*. ACM, 127–140.

[34] Per Håkon Meland, Yara Fareed Fahmy Bayoumy, and Guttorm Sindre. 2020. The Ransomware-as-a-Service economy within the darknet. *Computers & Security* 92 (2020), 101762.

[35] Pedro Moreno-Sanchez, Muhammad Bilal Zafar, and Aniket Kate. 2016. Listening to Whispers of Ripple: Linking Wallets and Deanonymizing Transactions in the Ripple Network. *Proceedings on Privacy Enhancing Technologies* 2016, 4 (2016), 436–453.

[36] Satoshi Nakamoto. 2009. Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf

[37] Satoshi Nakamoto. 2009. Bitcoin Forum. https://bitcointalk.org.

[38] Pranav Nerurkar, Sunil Bhirud, Dhiren R. Patel, Romaric Ludinard, Yann Busnel, and Saru Kumari. 2021. Supervised learning model for identifying illegal activities in Bitcoin. *Applied Intelligence* 51, 6 (2021), 3824–3843.

[39] Pierre Omidyar. 2021. The concept of industry. https://en.wikipedia.org/wiki/Industry_(economics).

[40] Masarah Paquet-Clouston, Bernhard Haslhofer, and Benoit Dupont. 2019. Ransomware payments in the Bitcoin ecosystem. *Journal of Cybersecurity* 5, 1 (2019), tyz003.

[41] Marcos G Quiles, Elbert EN Macau, and Nicolás Rubido. 2016. Dynamical detection of network communities. *Scientific Reports* 6, 1 (2016), 1–10.

[42] ID Ransomware. 2021. ID Ransomware. https://id-ransomware.malwarehunterteam.com/.

[43] Jamie Redman. 2017. Bitcoin Gamblers Have Wagered $4.5 Billion in BTC Since 2014. https://news.bitcoin.com/bitcoin-gamblers-wagered-4-5-billion-btc-2014/.

[44] Alan Reed. 2021. BitcoinAbuseDataset. https://www.bitcoinabuse.com/.

[45] Fergal Reid and Martin Harrigan. 2011. An Analysis of Anonymity in the Bitcoin System. In *Proceedings of the 2nd Conference on Privacy, Security, Risk and Trust (PASSAT)*. Springer, 1318–1326.

[46] Paul Salber and Paul Elosegui. 2021. Ethonym. https://ethonym.com.

[47] Michele Spagnuolo, Federico Maggi, and Stefano Zanero. 2014. BitIodine: Extracting Intelligence from the Bitcoin Network. In *Proceedings of the 18th Conference on Financial Cryptography and Data Security (FC)*. Springer, 457–468.

[48] Jiachen Sun, Rui Zhang, Ling Feng, Christopher Monterola, Xiao Ma, Céline Rozenblat, H Eugene Stanley, Boris Podobnik, and Yanqing Hu. 2019. Extreme risk induced by communities in interdependent networks. *Communications Physics* 2, 1 (2019), 1–7.

[49] Leah Zhang-Kennedy, Hala Assal, Jessica N. Rocheleau, Reham Mohamed, Khadija Baig, and Sonia Chiasson. 2018. The aftermath of a crypto-ransomware attack at a large academic institution. In *Proceedings of the 27th USENIX Security Symposium (USENIX Security)*. USENIX Association, 1061–1078.