IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 June 2022, accepted 25 June 2022, date of publication 29 June 2022, date of current version 8 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3187210

RESEARCH ARTICLE

Get off of Chain: Unveiling Dark Web Using Multilayer Bitcoin Address Clustering

MINJAE KIM[1], JINHEE LEE[1], HYUNSOO KWON[2], AND JUNBEOM HUR[1]
[1]Department of Computer Science and Engineering, Korea University, Seoul 02841, South Korea
[2]Samsung Electronics, Suwon-si 16677, South Korea

Corresponding author: Junbeom Hur (jbhur@korea.ac.kr)

This work was supported in part by IITP Grant funded by the Ministry of Science and ICT (MSIT), South Korea, under Grant 2019-0-01697, Grant Institute for Information & communication Technology Planning & evaluation (IITP)-2022-2020-0-01819, and Grant IITP-2021-0-01810; and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, South Korea, under Grant NRF-2021R1A6A1A13044830.

ABSTRACT Bitcoin is the most widely used cryptocurrency for illegal trade in current darknet markets. Owing to the anonymity of its addresses, even though transaction flows are globally visible, Bitcoin clustering remains one of the most challenging and open problems in illegal Bitcoin transaction analysis. In this article, to resolve this problem, we propose a novel *multi-layer heuristic* algorithm for Bitcoin clustering, which leverages on-chain transactions as well as off-chain application data in the real world. For this purpose, we first explored the unique characteristics of darknet market ecosystems including their trading systems. By conducting an in-depth analysis of the data manually collected for 11 months, we found that some darknet market review data disclosed transactions containing Bitcoin value and item delivery information. We then identified unique Bitcoin addresses associated with the disclosed information, owned by the same darknet providers. Based on address ownership, more accurate market clusters could be created, which have not previously been identified by other clustering algorithms. According to our experimental results, approximately 31.68% of the darknet market review data matched real Bitcoin transactions, and 122 hidden clusters associated with Silk Road 4 were found. This indicates that the proposed algorithm can complement existing clustering methods and significantly reduce the false negative rate by up to 91.7%.

INDEX TERMS Address clustering, Bitcoin, blockchain, de-anonymization.

I. INTRODUCTION

Since the advent of Bitcoin in 2008 [1], several blockchain-based cryptocurrencies have been developed, such as Ethereum [2], Monero [3], and Zcash [4]. Because a blockchain is inherently distributed, transparent, and resistant to the modification of data using strong cryptography, blockchain-based cryptocurrencies can work as digital assets while also providing a certain level of anonymity to users. Recently, because of their anonymity, many hidden markets have begun to use cryptocurrencies as a medium of exchange for illegal products and cybercrimes. According to [5], 25.97% of 605.69 million Bitcoin transactions were associated with darknet markets in 2019 involving illegal transactions.

Dark web pages cannot be accessed with surface web browsers such as Chrome [6] and Firefox [7] because they typically adopt anonymous network technologies such as Tor [8], which encrypts the routing path between a client and a server to hide their IP address and the server name used during the connection. Thus, by exploiting such anonymous communications, many illegal activities (e.g., drugs, malware, fraudulent dealings) account for 25% of the Tor dark web [9].

To trace such illegal activities, several heuristic algorithms that examine Bitcoin transactions using clustering have been proposed [10], [11]. Bitcoin clustering refers to a procedure of finding Bitcoin address belonging to the same owner by investigating Bitcoin transaction data recorded in a blockchain. Thus, accurate clustering of Bitcoin

The associate editor coordinating the review of this manuscript and approving it for publication was Nazar Zaki.

addresses is essential for correctly tracing Bitcoin transaction flows among operators of virtual wallets. Unfortunately, such heuristic algorithms will miss a number of addresses that belong to the same owner and cannot cluster them in the real world. The root cause of this limitation is that they utilize only public Bitcoin transaction data in the blockchain, which provides only limited (transparent but anonymized) information regarding the transactions. Without knowledge of ground-truth information, such clustering algorithms produce a high rate of false negatives.

To overcome this fundamental limitation, we propose a *Multi-layer heuristic* algorithm for Bitcoin clustering that leverages not only the Bitcoin transaction data in the blockchain layer (on-chain data), but also user data in the application layer disclosed in the real world (off-chain data), especially information related to the darknet market. The proposed scheme first constructs two independent-looking clusters in the blockchain layer and then combines them by analyzing the linking degree (the linkability) between them in the application layer. Although several previous studies [12], [13] also attempted to utilize off-chain information for Bitcoin clustering, they simply used data associated with existing clusters that have already been identified as a *priori* knowledge. Goldfeder *et al.* [14] linked web cookies to Bitcoin transactions as off-chain data, but the target was confined to only the surface web and excluded the dark web. In a study similar to ours, Schäfer *et al.* [15] recently identified a darknet market vendor's Bitcoin address by investigating the vendor's review. However, this method requires that vendors reuse their Bitcoin addresses and that the market has no built-in addresses, which are not common characteristics of many current darknet markets. In contrast, our study focuses on identifying hidden clusters and revealing the relationships among them that have not been disclosed in previous studies. We accomplish this by leveraging real-world off-chain data from the darknet market, as well as on-chain data.

For this purpose, we explored darknet market ecosystems, manually collected real-world darknet market data from November 2019 to September 2020, and analyzed their trading patterns in depth. Among the various data we collected, we found that review[1] of the darknet market disclosed actual Bitcoin transactions such as the Bitcoin(BTC) value of an item and its date-time shipment information. In addition, we found that the darknet market adopt an escrow system that uses only one-time Bitcoin addresses. Based on this off-chain information in the application layer and the characteristics of the address type provided by the darknet market, we identified unique Bitcoin addresses in the blockchain that are owned by the same darknet operator.

Specifically, the proposed *Multi-layer heuristic* algorithm consists of the following three steps: (1) Find unique matches between Bitcoin addresses (transaction data in the blockchain layer) and those of darknet market review data (in the application layer). (2) The matched addresses obtained in the first step are then grouped as clusters using the existing *Multi-input Heuristic* and *Change Address Heuristic* algorithms [10]. (3) These clusters are grouped into a single cluster, which should be a part of the actual cluster involving the same operator.

To evaluate its efficacy, we implemented a *Multi-layer heuristic* algorithm and measured its clustering accuracy using a review dataset gathered from Silk Road 4 [16]. Our experiments show that approximately 31.68% of illegal transaction information from the Silk Road 4 review data uniquely matched real Bitcoin transactions. Based on the matched transaction information, the proposed heuristic algorithm found 122 hidden clusters, not previously identified. By combining these separate clusters into a single market cluster(that of Silk Road 4), the *Multi-layer heuristic* algorithm could reduce the false negative rate by up to 91.7% using the review data that we collected, demonstrating the efficacy of the proposed method in the real world. The reliability of our heuristic approach was verified by performing a simulation using the actual distribution of Bitcoin transactions, overcoming the absence of ground-truth.

A. CONTRIBUTIONS

Our study provides the following contributions:

- We explored the real-world darknet market ecosystem and found that it typically supports escrow-based transactions and provides one-time Bitcoin addresses to ensure privacy of vendors and buyers.
- We verified that previous Bitcoin clustering algorithms suffer from a fairly high rate of false negatives and investigated their root cause.
- We propose a novel *Multi-layer heuristic* algorithm for Bitcoin clustering that leverages the Bitcoin transactions in the blockchain layer and the one-time Bitcoin addresses in the application layer as off-chain data. Table 1 shows the comparison result among the Bitcoin address clustering methods using off-chain data.
- Through our experiments with Silk Road 4 darknet market data gathered from Nov. 2019 to Sep. 2020, our algorithm revealed a total of 122 hidden clusters for Silk Road 4, and its accuracy was further verified by performing a simulation based on the distribution of actual Bitcoin transactions in the real world.

B. ORGANIZATION

In the remainder of this paper, we first introduce the necessary background of the darknet market and Bitcoin in Section II. In Section III, we review the related work. Then, we present the darknet market transaction analysis in Section IV. Our heuristic, which we call *Multi-layer heuristic*, is introduced in Section V. In Section VI, we present the experimental results of our heuristic and state some limitations. Finally, we provide a conclusion in Section VII.

[1]In the darknet market, users can write reviews on their purchases, and even if they do not write any, reviews such as "No-Feedback" are automatically entered after a certain period of time. These reviews contain information such as the user's ID, price of the product at the time of purchase, and delivery date of the product.

TABLE 1. Comparison of Bitcoin address clustering heuristics using off-chain data.

	[12], [13]	[14]	[15]	Ours
Identifying unknown cluster	X	O	O	O
Off-chain data	Tag	Web cookie	Review (partial)	Review (all)
Clustering target	Already known clusters	Surface web	Darkweb market (vendor)	Darkweb market (escrow)
Clustering with one-time address	-	-	X	O

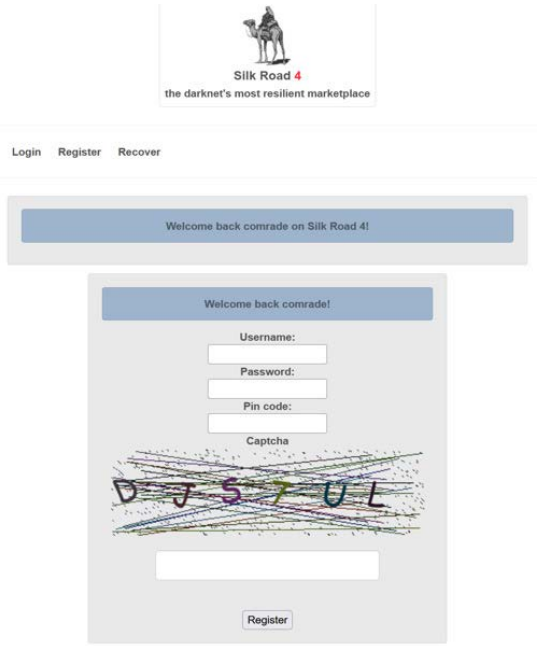


FIGURE 1. Registration Page for Silk Road 4.

II. BACKGROUND

A. TRADING SYSTEM IN DARKNET MARKET

The darknet market is a subset of dark web services designed to provide selling or brokering services involving drugs, counterfeit currency, personal information, fraud, and hacking services, among others. Similar to a conventional e-commerce market, individual vendors list their wares on the darknet market, and then buyers can make a purchase request for goods of interest, typically using Bitcoin to ensure anonymity.

The darknet market adopts several mechanisms to make transactions anonymous in both the customer registration and interaction steps. For example, Fig. 1 shows the registration page for Silk Road 4, which is currently one of the largest darknet markets. As shown in Fig. 1, users are only required to enter an username, password, and PIN code after simple Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) verification, without

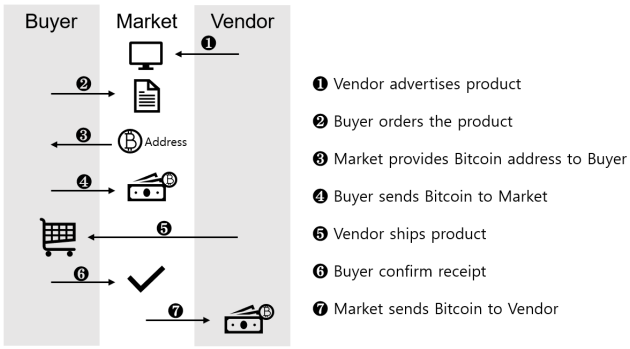


FIGURE 2. Escrow Trading System used by Darknet Markets.

providing any personal information such as e-mail, phone number, or residence information during the membership registration step in the market. When registering in the market, a Pretty Good Privacy (PGP) key is assigned to the vendor's account, which is then used to encrypt every communication between the vendor and buyers. According to a survey in [17], there was a substantial increase in PGP support from vendors, thereby making markets more resilient to illegal transaction analysis.

An escrow trading system is used to provide trust to among buyers and vendors in the darknet market. In escrow-style transactions [18], all transactions are centrally managed by the market so that direct transactions between buyers and vendors are not linked. The escrow trading method is shown in Fig. 2. In this system, registered vendors post illegal items on the market and these become visible to everyone who has subscribed to the market. If a user wishes to buy a product, they make a purchase request through the market post. When a purchase is requested, the darknet market displays the deposit amount along with a Bitcoin address. Once the correct amount has been deposited to the corresponding Bitcoin address by the buyer, the vendor initiates shipment of the items to the buyer. Once the buyer confirms receipt, the transaction is closed. Then, the buyer ID, transaction item, Bitcoin amount, and delivery completion information are left as a review. We observed that escrow systems in darknet markets have recently evolved in terms of issuing Bitcoin addresses for escrow.

Typically, Silk Road 3.1 would issue approximately 70 Bitcoin addresses every month as a pool of addresses, and would then randomly assign them to transactions. However, we found that Silk Road 4 no longer creates a pool of Bitcoin addresses, and instead creates a new address for each transaction. More details are provided in Section IV.

B. BITCOIN ADDRESS

Bitcoin addresses are random sequences of letters, generated by hashing a public key or a set of public keys (called a script) using both SHA-256 and RIPEMD-160, and then encoding it [19]. According to the hashing targets (i.e., a public key or script) and encoding methods (Base 58, or Bech32) [20], [21],

TABLE 2. Bitcoin address types.

Type	Hash target	Encoding	Starting with
PublicKeyHash	Public Key	Base58	"1"
ScriptHash	Script	Base58	"3"
WitnessPublicKeyHash	Public Key	Bech32	"bc1"
WitnessScriptHash	Script	Bech32	"bc1"

FIGURE 3. Bitcoin Transaction.

the address formats are classified into four different types with different starting symbols, as shown in Table 2.

C. BITCOIN TRANSACTIONS AND CLUSTERING

1) BITCOIN TRANSACTIONS

A Bitcoin transaction is the transfer of a Bitcoin value between Bitcoin addresses, which is recorded in a public distributed ledger(blockchain). Transactions are broadcast to the entire Bitcoin network using readily available applications called wallets for verification, and recorded in the blockchain after their validation.

Bitcoin transactions comprise one or more inputs and outputs. When a user sends Bitcoins, the user designates each address and the amount of Bitcoins sent to that address in the output. Fig. 3 shows an example of Bitcoin transactions, where one of the output addresses in Transaction 1 is used as one of the input addresses in Transaction 2, which in turn sends the Bitcoin value(s) to different output addresses.

Blockchain networks are transparent [22]; therefore, Bitcoin transaction flows can be traced by every Bitcoin user in the network from the initial transaction to the latest one. Thus, if we can find addresses belonging to the same wallet, tracing Bitcoin transactions among wallets, and in turn, their owners should also be possible ,[2] which would be useful to trace illegal transactions and identify their operators in the dark web. Therefore, Bitcoin address clustering is the first and most important step for an in-depth transaction analysis of actual Bitcoin owners. However, owing to several Bitcoin mechanisms used to obfuscate transaction flows, such as a stealth address [24], which is used as an additional output for returning any change due (in Bitcoins) back to the payer, Bitcoin clustering remains an open problem in practice.

[2]The link between real identities and wallets is generally unknown and is often referred to as pseudonymity. Although several recent studies have attempted to find linkability between the addresses and identities, such as the real owner's IP address [23], the linkability between them is generally beyond the grasp of Bitcoin clustering itself.

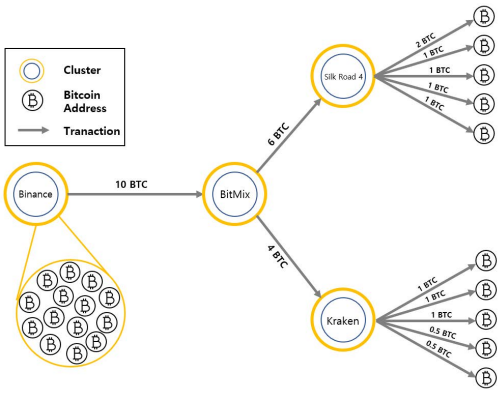


FIGURE 4. Bitcoin Clustering Overview.

2) BITCOIN ADDRESS CLUSTERING

Bitcoin clustering is the task of finding addresses that belong to the same owner (who may have many different wallets). Typically, clustering begins with seed addresses, which are the target addresses. Any other addresses belonging to the same owner are then gathered through clustering algorithms, creating increasingly larger clusters. If the seed addresses (or any other grouped addresses) can be labeled from the known information (e.g., addresses published by some markets or individuals), an entire cluster can be labeled as the same operator, as shown in Fig. 4, which assists in the transparent analysis of operator-to-operator transaction flows.

Unfortunately, there is no guarantee that clustering algorithms will always produce accurate results, mainly because of a lack of knowledge regarding ground-truth information. Therefore, several clustering *heuristics* have been proposed. Each clustering heuristic differs in the method of identifying and classifying the addresses, but all exploit only the mechanisms in which transactions are created by wallet software (i.e., they use only on-chain data for clustering). The heuristic algorithms proposed to date are discussed below.

a: MULTI-INPUT HEURISTIC [10], [11]

If two or more addresses are inputs of the same transaction with one output, all addresses are controlled by the same user. In the early Bitcoin system, when creating a single transaction, all input addresses involving the same transaction were controlled by the same user regardless of the number of outputs. However, with the advent of Coinjoin [25], where multiple users can participate in one transaction,[3] this algorithm has been slightly modified accordingly.

[3]The number of output addresses of transactions created with Coinjoin is equal to the number of participating users.

b: CHANGE ADDRESS HEURISTIC [10], [11]

When a Bitcoin transaction is made, a one-time change address(used to return any BTC left over from a transaction to the buyer), called a shadow address, can be created as one of the output addresses and is controlled by the same user as the input addresses. Thus, the change address is used as a receiving address, and the other output addresses are used as spending addresses. This heuristic exploits the method of handling the change address. However, it is challenging to determine which of the output addresses is actually a change address belonging to the same wallet.

c: CONSUMER HEURISTIC [11]

As opposed to the two previous heuristic algorithms, this heuristic applies only to the clustering of consumer wallets, which by default allow Bitcoins to be sent to only a single address, such as Bitcoin Core, Android Bitcoin [26], or Wallet [27]. Transactions made in consumer wallets always have two or fewer outputs, therefore, the consumer transaction contains either one or no change address. If one of the output addresses contains the address of a unique cluster, the other output address is considered the change address.

d: OPTIMAL CHANGE HEURISTIC [11]

This heuristic is used to identify the change address. It is based on the assumption that wallet software does not insert unnecessary input addresses when creating transactions. Essentially, for a transaction to be valid, the total input value must be greater than the output value (that is, some change is owed to the buyer). If only one of all input value is omitted, the sum of the input is less than the sum of the output values. Therefore, if the minimum input value is less than the value of the change address, the minimum input value can be omitted, so the value of the change address must be less than the minimum value among the input values. Thus, if the transaction has an output that is lower than the minimum input, it is most likely a change address.

Consumer heuristic and *Optimal change heuristic* algorithms suggest ideas for extending the previous two heuristics for further clustering. They are based on users' behaviors, such as sending Bitcoins to a limited number of outputs and not using unnecessary outputs, unlike the previous two heuristics. Unfortunately, these assumptions may not hold in a dark web environment, because dark web users tend to typically carry out the opposite behavior to avoid leaving traces. Therefore, we only consider the *Multi-input heuristic* and *Change address heuristic* algorithms as the primary algorithms for comparison with the proposed algorithm.

III. RELATED WORK

A. BITCOIN CLUSTERING

Meiklejohn *et al.* [10] proposed the first heuristic algorithms for Bitcoin clustering, called the *Multi-input heuristic* and *Change address heuristic* based on Bitcoin transaction data. With the advent of Coinjoin [25], in which the number of output addresses of a transaction is equal to the number of participating users, *Multi-input heuristic* was slightly modified to address the peculiarity of Coinjoin.

Since then, several studies have proposed combining Bitcoin transaction data with web data for Bitcoin clustering. Ermilov *et al.* [12] proposed a Bitcoin clustering method that utilizes public information obtained from the web by leveraging previously identified public tag information. Thus, in an environment where such a *priori* knowledge is not available or given, the clustering method will not work properly, as opposed to our clustering method, which performs Bitcoin clustering without such tag information. Subsequently, Goldfeder *et al.* [14] proposed another approach that allows third-party web trackers to de-anonymize cryptocurrency users by linking their web cookies with blockchain transactions according to their purchasing information. However, their work was conducted only on Bitcoin-accepting surface web sites (not the dark web) and relied on the assumption that the price and time information can be obtained from the information available through cookies.

In a study similar to ours, Schäfer *et al.* [15] recently introduced a method to identify Bitcoin addresses using darknet market's review data. In their method, the market type is first classified into wallet-less or wallet-based markets according to whether the market has its own built-in wallet. Vendor Bitcoin addresses that were reused in the wallet-less market were then identified. Our matching algorithm, however, can be applied more generally regardless of the market type or reuse of a vendor's Bitcoin address. In addition, beyond identifying a vendor's Bitcoin address, we propose a generic clustering heuristic that can group Bitcoin addresses in a market that have not been identified previously, complementing the existing heuristics by mitigating the false negative problem.

B. UNDERSTANDING THE DARK WEB

To understand the target landscape and structure of hidden services in anonymous networks, many studies have analyzed Tor traffic [45]–[47] and related activities [48]–[51]. Biryeukov *et al.* [52], [53] analyzed hidden services hosted through Tor and found that many hidden services were being maintained for illegal trafficking. Recently, Van Wegberg *et al.* [54] observed the increasing commoditization of cybercrime in anonymous online markets, as these lower the entry barriers for aspiring criminals and therefore facilitate cybercrime. To investigate criminal activity in the dark web, they collected more than 500,000 transactions from eight anonymous markets in the dark web over six years from sites such as Silk Road and Alpha Bay, and tracked the evolution of illegal products. They found that these sites generated a large amount of revenue from business-to-business (B2B) and business-to-consumer (B2C) transactions using the dark web, and estimated that the total revenue for cybercriminals in the dark web would exceed $8 million for B2B transactions and $7 million in B2C transactions over the six years, respectively.

Ciancaglini *et al.* [48] analyzed criminal activity in Tor hidden services by classifying features such as language and items, and compared the trade patterns to those of the surface web. In their study, they analyzed how users traded cybercrime commodities in the deep web, and compared the trade patterns to those of the surface web. They then analyzed illegal transactions in the dark web as a follow-up to their previous study [49], and discovered that many anonymous networks in the deep web have become a safe haven for criminal activity in both the cyber and physical realms.

Soska and Christin [50] analyzed the types of products sold on 16 Tor sites between 2013 and 2015. They assessed the effect of adversarial events on the overall size of the economy and provided insights into how vendors were diversifying across marketplaces in the dark web and how security practices were evolving with respect to vendors. They found that vendors are likely to use PGP keys to hide criminal activity by encrypting communications. Barratt *et al.* [55] surveyed the dark web drug marketplace, specifically Silk Road, to investigate the public's awareness of the online illicit drug marketplace in the United Kingdom, Australia, and the US, and to understand to what extent drug purchasers preferred the online drug marketplace. They focused not only on illicit drug traders, but also on normal buyers who purchased generic drugs, and analyzed why they were likely to use the online darknet market. Analysis results indicated that its appeal to drug purchasers was moderated by country-specific deterrents and market characteristics. Foley *et al.* [5] suggested several functions for estimating the scale of illegal activity in darknet markets. They found that approximately one-quarter of Bitcoin users were involved in illegal activities, and estimated that around $76 billion of all illegal activity per year involved Bitcoin, which is close to the scale of the US and European markets for illegal drugs.

C. TRANSACTION ANALYSIS IN DARKNET MARKETS

To understand the structure and illegal transactions within the dark web, Lee *et al.* [56] designed and implemented a tool that performs dark web data crawling. Using the crawler, they collected approximately 27 million dark web pages over 15 months and found 5,440 Bitcoin addresses. Among them, they found that only 85 Bitcoin addresses were actually used for illicit transactions and estimated that these illicit trades amounted to approximately $180 million. This study helped to understand the marketing strategy in the darknet market and the flow of funds associated with illegal transactions. However, their crawling tool gathers too many false addresses (specifically, 98.5% were false addresses), thereby requiring tremendous manual effort to prune false Bitcoin addresses that do not represent actual illicit transactions in the darknet market. Broseus *et al.* [57] investigated the use of dark websites for illegal trade using cryptocurrencies in Canada, and suggested that a more holistic approach should be developed to disclose the cybercrime landscape and hidden structure of darknet market services in the dark web.

There have also been previous studies measuring the volume of illicit trades and cryptocurrency transactions in the dark web. Christin [28] collected and analyzed Silk Road data for six months between 2011 and 2012, and surveyed 24,400 individual items sold on Silk Road. They found that Silk Road was overwhelmingly used in the trade of illegal goods, especially illicit drugs. In addition, the study estimated the total revenue as over $1.2 million per month on Silk Road, which corresponded to approximately $92,000 in commissions for Silk Road operators per month in 2012. Demant *et al.* [58] analyzed Silk Road 2.0 and Agora. They collected data using web crawlers from 2014 to 2015 and examined the demand for traded goods. Based on their analysis, they found that most of the revenue came from B2B trading, and that there was a decrease in volume in both markets during this period. The results indicated that the two darknet markets resembled the traditional drug market in terms of distribution and revenues.

IV. DARKNET MARKET TRANSACTION ANALYSIS

As discussed previously, the primary heuristics depend solely on the on-chain transaction information available in the blockchain layer (particularly the number of inputs and outputs in the Bitcoin transaction). Thus, they may inevitably fail to detect Bitcoin addresses belonging to the same wallet, resulting in a large number of false negatives. One promising approach for reducing false negatives is to extend the information source and leverage the off-chain information of other layers, such as application data. Therefore, to investigate this possibility in the application layer, we collected darknet market data from Silk Road 3.1 and 4 and analyzed the data to understand actual patterns of transactions and their linkability to Bitcoin transactions in practice.

A. SILK ROAD 3.1

In the past decade, several studies [28]–[34] have attempted to analyze the characteristics of transactions conducted in Silk Road 1 and Silk Road 2 in order to disclose their clusters. However, there has been little effort to explore more recent markets, such as Silk Road 3.1 and 4 with respect to transaction patterns. Thus, Bitcoin clusters and their transaction flows in current darknet markets remain behind the veil.

To investigate transactions in Silk Road 3.1, we first attempted to collect Bitcoin addresses publicly revealed on the Silk Road 3.1 escrow system, and found that they are all ScriptHash addresses. Then, we analyzed their relationships from a clustering point of view. From November 25 to December 16, 2019, we collected a total of 75 Bitcoin addresses, including duplicates. After eliminating duplicate addresses, we identified 48 unique addresses. To check the clustering results of the primary heuristics, that is, the *Multi-input heuristic* and *Change address heuristic*, we executed an open-source clustering tool, Blocksci [35] on these 48 addresses.

As a result, of the 48 addresses, 18 were grouped into one cluster, and the other 30 were grouped into a different cluster, which ultimately contained 57 addresses and 72 addresses

in total, respectively. This demonstrates that the primary heuristics produced false negatives in the real-world setting. Considering that the primary heuristics do not produce false positives, we can conclude that a total of 129 addresses that would not be identified using just the data in the blockchain layer can be grouped together as a single cluster.

After further investigating Bitcoin transactions including the 129 addresses identified, we found the following address patterns adopted in Silk Road 3.1:

- The type of Bitcoin address provided by Silk Road 3.1 is ScriptHash, supporting a 1-of-2 multi-signature scheme.
- When a Silk Road 3.1 address is used as one of the input addresses in a transaction, all of the input addresses in the transaction use ScriptHash, and the number of output addresses is always 1.

Unfortunately, while we were investigating the transactions, Silk Road 3.1 was discontinued on Dec. 17, 2019, and Silk Road 4 was implemented on Apr. 1, 2020. Although Silk Road 3.1 was discontinued, we observed convincing evidence of a link between Silk Road 3.1 addresses and Silk Road 4 transactions. Thus, we investigated transactions in Silk Road 4, as discussed in detail in the next section.

B. SILK ROAD 4

To trace hidden transactions and disclose their associated clusters for a specific operator, we must first find several *seed* addresses that are evidently associated with the operator, collect Bitcoin transactions containing these seeds, and construct a cluster using the clustering heuristics. In Silk Road 3.1, this approach was feasible because the escrow system provided constant Bitcoin addresses that were recorded in some Bitcoin transactions in the blockchain, allowing the addresses to be *searchable*.

However, while gathering Bitcoin addresses from Silk Road 4, we found that the escrow system changed the method for address provision. Specifically, Silk Road 4 always provides newly generated addresses per trade that have not been recorded in prior transactions in the blockchain, which makes the obtained addresses *unsearchable*, as opposed to Silk Road 3.1 (however, the address type is still ScriptHash). Hence, we need a different approach to obtain the *seed*, but the use of *searchable* addresses in the Silk Road 4 escrow system using previous transaction history remains visible.

To obtain such Bitcoin addresses from Silk Road 4, we conducted an in-depth analysis of the internal structure and contents of the Silk Road 4 web pages, and noted that 'review' data could be an alternative information source for seeds. As shown in Fig. 5, Silk Road 4 review data contain information regarding user ID, item, Bitcoin value, and shipped date. Because the Bitcoin value and shipped date could potentially be associated with the value and timing of some existing Bitcoin transactions, we collected 606 review data from the Silk Road 4 web pages Apr. 10 to Sep. 30, 2020 in order to explore this possibility.

The Bitcoin value appearing in the review data is the value that the buyer actually paid to the Silk Road 4 escrow system.

FIGURE 5. Silk Road 4 Review Data Example.

Therefore, there must be at least one Bitcoin transaction in a blockchain having the same value as the output. Even if the Bitcoin value is expressed to eight decimal places, which makes it difficult to find multiple transactions in practice, it becomes more likely to find Bitcoin transactions with the same value if the time range is wider. Thus, to find a unique match given the Bitcoin value while avoiding exhaustive searches over all transactions, we narrowed the search space of the transactions by taking advantage of the delivery information recorded in the review data. Specifically, according to the information from Silk Road 4, it normally takes 2-4 days for items to reach the buyer after payment, and the escrow system sends the corresponding Bitcoin amount to the vendor in at most 14 days. In the Silk Road 4 review data, the shipment date is expressed as "*n* days ago". Thus, we searched for the transactions in which the Silk Road 4 escrow received the Bitcoin value and sent it to another address from *n* days prior to the review date to 14 days later. This significantly reduces the search space and increases the possibility of finding a unique match in the real world, as demonstrated in Section III.

C. COMPARISON BETWEEN SILK ROAD 3.1 AND SILK ROAD 4

One open problem in finding the correct match is that we can only utilize the review data of Silk Road 4 without any prior knowledge of the ground truth. Thus, to further increase the possibility of finding a correct match among multiple candidates that would likely include false matches, we first checked whether the patterns of Silk Road 3.1 could also be found in Silk Road 4. If this held, it would assist in finding correct matches by exploiting similar features between the two versions. Thus, we analyzed the Bitcoin address type offered by Silk Road 4 and compared it to that of Silk Road 3.1 as follows.

1) SCRIPTHASH ADDRESS TYPE

We analyzed whether the Bitcoin address type provided by the Silk Road 4 escrow was ScriptHash, as used in Silk

TABLE 3. Transaction analysis results associated with Silk Road 4 review data.

Exact match rate	31.68%
Signature type	1-of-2 multi-signature: 10.48%
	WitnessPublicKeyHash: 89.52%
The number of output address	1-output transaction: 80.73%
	More than 2-output transaction: 19.27%

Road 3.1 and observed that Silk Road 4 only provides Bitcoin addresses that begin with '3', indicating that it also uses ScriptHash address types.

2) SUPPORTING 1-OF-2 MULTI-SIGNATURE

Next, we attempted to determine whether Silk Road 4 still uses SciptHash to support 1-of-2 multi-signatures. To determine Bitcoin addresses use the multi-signature approach, transactions including that Bitcoin addresses are required. However, the Bitcoin addresses provided by Silk Road 4 have never been used in previous transactions, making them unsearchable in the blockchain history, as discussed in Section IV-B. Therefore, we investigated the Bitcoin transactions associated with the review data we collected from Silk Road 4 (particularly the transactions generated within up to 14 days after the delivery date recorded in the review data). If exactly one Bitcoin transaction is found in the search space for a given Bitcoin address, it is most likely the correct Bitcoin transaction related to the review data. Based on our experiment, we found that 31.68% of the addresses in the 606 review data matched exactly one transaction(Table 3). Among them, 10.48% used ScriptHash supporting 1-of-2 multi-signature, and the other 89.52% used ScriptHash nesting WitnessPublicKeyHash. Therefore, we found that Silk Road 4 supports diverse signature types, unlike Silk Road 3.1, and concluded that it is impractical to use the signature type as the sole indicator for identifying Silk Road 4 addresses.

3) 1-OUTPUT TRANSACTION

We also checked whether Silk Road 4 transactions exhibit the 1-output behavior observed in Silk Road 3.1. The results of the analysis are summarized in Table 3. We found that among the transactions associated with the 31.68% addresses that exactly matched one address in the review data, 80.73% have only one output address, whereas the others include two or more output addresses in a transaction. Hence, when it comes to the '1-output transaction' behavior of Silk Road 3.1, we found that Silk Road 4 no longer follows this rule, and thus it should be removed from the set of useful indicators for identifying Silk Road 4 addresses, as is the case for the signature type.

Consequently, the only useful and deterministic pattern regarding Silk Road 4 transactions that we identified is the address format, using ScriptHash. Compared to Silk Road 3.1, it results in a much more restrictive condition for fine-grained clustering approaches. In the presence of such challenges, we describe how our method attempts to overcome the limitations in Section V.

TABLE 4. Review data format by market. ('O' indicates the data is present.).

Market Name	User ID	Item Name	Price Format	Shipped Date
Silk Road 3.1	O	O	BTC value	O
Silk Road 4	O	O	BTC value	O
Apollon Market	O	O	US dollars	O
Agartha Market	O	O	X	O

D. OTHER DARKNET MARKETS

To evaluate whether our method using review data could also be applied to the analysis of other darknet markets, we investigated several other well-known markets such as Apollon [36] and Agartha [37]. The comparison results of the review data format provided by each market are presented in Table 4.

As shown in the Table 4, all markets commonly provide the user ID, item name, and shipped date information in the review. However, the price formats of each market exhibit some differences. Specifically, while Silk Road 3.1 and 4 use the Bitcoin (BTC) value as a currency for products, Apollon uses US dollars, and Agartha does not provide any pricing information. Therefore, Agartha makes it almost impossible to identify and cluster hidden transactions associated with the market solely on its viewable web data. However, Apollon records the exchange rate at the time of a trade, allowing accurate conversion betwwen USD and BTC values. Thus, it is straightforward to apply our heuristic algorithm to the Apollon market using the converted BTC value. Although we investigated only a few darknet markets other than Silk Road 3.1 and 4, it is important to note that our approach can be applied to any darknet market provided that the market uses the same review data format as off-chain data.

V. CLUSTERING HEURISTIC

In this section, we propose a novel heuristic called the *Multi-layer heuristic*. The *Multi-layer heuristic* can complement and improve existing clustering heuristics such as the *Multi-input heuristic* and *Change address heuristic* by taking advantage of the application data as well as the transaction data on the blockchain with full compatibility.

A. MATCHED ADDRESS

Let V be the BTC value of an item in the review data and D be the shipping date/time for the item, where (V, D) represents a single review datum, and $P(V, D)$ represents a web page where the review datum (V, D) is posted. Let A and TX be the Bitcoin address and the transaction, respectively. Finally, we define three functions: $ITX(A) \rightarrow TX$ is a function that returns TXs (transactions) where Bitcoin address A appeared as one of the inputs. $OTX(A) \rightarrow TX$ is a function that returns TXs where Bitcoin address A appeared as one of the outputs, and $Time(TX) \rightarrow datetime$ is a function that returns the date/time when transaction TX was broadcast to the Bitcoin network. $|TX\text{s}|$ representes the number of transactions and $datetime_i \rightarrow datetime_j$ denotes that $datetime_j$ occured later than $datetime_i$.

Definition 1: For each review datum (V, D) of $P(V, D)$, the Bitcoin address A that meets the following conditions is defined as a *matched address*, MA_A.

1) $|ITX(A)| = |OTX(A)| = 1$.
2) $Time(OTX(A)) \rightarrow D \rightarrow Time(ITX(A)) \rightarrow D +$ 14 days.
3) In $ITX(A)$, the input value of A is V BTC.

Intuitively, the first condition implies that if A is the Bitcoin address provided by the Silk Road 4 escrow to the buyer, there must be one $ITX(A)$ for the buyer-to-escrow transaction and one $OTX(A)$ for the escrow-to-vendor transaction. The second and third conditions imply that (1) the buyer-to-escrow transaction for the actual payment corresponding to the review data (specifically, the transaction including the same BTC value as the review data) should exist before the shipping date/time in the review data, and (2) the escrow-to-vendor transaction for the actual payment corresponding to the review data should exist within 14 days of the shipping date/time in the review. As a result, if A is a matched address for a review datum in Silk Road 4, it is the actual Bitcoin address used for the corresponding transaction.

B. MULTI-LAYER HEURISTIC

Let C_i (for a positive integer i) be the Bitcoin cluster generated using *Multi-input heuristic* and *Change address heuristic* algorithms. Subsequently, the proposed *Multi-layer heuristic* algorithm is defined as follows.

Multi-Layer Heuristic

- If C_1 and C_2 contain MA_A and MA_B, respectively, and MA_A and MA_B are from the same $P(V, D)$ for some (V, D), then C_1 and C_2 are controlled by the same user.

An important peculiarity of the above heuristic is that it can find the hidden relationships between existing clusters and can cluster them further by taking advantage of off-chain data obtained from web pages as well as on-chain transaction data, while fully conforming to the existing clustering heuristic algorithms. Owing to this orthogonal property, it can naturally complement existing clustering algorithms by reducing false negatives rather than replacing these algorithms.

Previously, Ermilov *et al.* [12] proposed a Bitcoin clustering method that utilizes a *priori* given tag information (e.g., tags for popular currency exchanges [38], [39], and gambling [40]), that have already been disclosed through web forums or user profiles. If such tag information is provided, users can find change addresses by easily removing the known cluster information [10]. On the other hand, our Bitcoin clustering algorithm leverages off-chain data in the application layer (especially review data in darknet markets), whose association with existing clusters have never been disclosed previously. Unfortunately, from our analysis of Silk Road 4, such a *priori* knowledge can hardly be obtained in current darknet markets. Our clustering algorithm overcomes such practical limitations and increases Bitcoin clustering accuracy by finding hidden relationships among clusters and further linking them. Therefore, the proposed heuristic algorithm should assist in making Bitcoin transactions more transparent and traceable in the real world.

VI. EXPERIMENTAL RESULTS

In this section, we describe the algorithms and implementation details of our proposed multi-layer Bitcoin clustering heuristic. We then present the experimental results using the algorithm.

A. IMPLEMENTATION SETUP

To implement the algorithms for finding matched addresses in the review data and applying Bitcoin clustering, we used a system equipped with an Intel Xeon E5-2620 3.0GHz processor and 256GB RAM. We used the Bitcoin Core [26] to download all Bitcoin blockchain data. Additionally, we used BlockSci [35] for blockchain analysis, which is an open-source software platform that includes a library of useful analytic tools such as identifying special transactions (e.g., CoinJoin) and linking different addresses based on the primary clustering heuristics (*Multi-input heuristic* and *Change address heuristic*). We implemented the algorithms in Python using the Jupyter notebook [41], a web-based interactive development environment.

To find the matched Bitcoin addresses for each review datum automatically, we first represented 606 raw review data points collected in the form of (V, D) tuples. Based on the pre-processed (V, D) tuples, we applied the algorithms for finding matched addresses and Bitcoin clustering as described in Sections VI-B and VI-C, respectively.

B. FINDING MATCHED ADDRESSES

1) ALGORITHM

Algorithm 1 Finding Matched Address

Require: One review data tuple (V, D)
Ensure: Matched address MA
  Obtain blocks B of Bitcoin blockchain data from D to D+14
  In all transactions involving B, Obtain all inputs I
  Initialize matched address set $MA = \emptyset$
  for all $i \in I$ **do**
    $A = i$'s address
    1. Check type of $A = ScriptHash$
    2. Check $|ITX(A)| = |OTX(A)| = 1$
    3. Check $Time(OTX(A)) \rightarrow D$
    $\rightarrow Time(ITX(A)) \rightarrow D + 14$
    4. Check i's value $= V$
    If all above are True, then $MA \cup \{A\} \rightarrow MA$
  end for
  if $|MA| = 1$ **then**
    Return MA
  end if

Algorithm 1 presents the procedure for finding a matched address for a given (V, D) tuple. The algorithm first sets

TABLE 5. Matched Address Rate.

Condition	ScriptHash	ScriphtHash & WitnessPublicKeyHash & 1-output transaction
Rate	31.68%	35.31%

the search space as the period between D and $D + 14$. In the search space, Algorithm 1 attempts to find matched addresses, *MA*s, which meet all of the conditions in Definition 1 using BlockSci. In addition, as described in Section IV-B, the address type provided by Silk Road 4 is ScriptHash; therefore, *MA*s should be ScriptHash-type addresses. If only one candidate satisfies all of the above conditions, it is considered a matched address *MA* for a given (V, D).

Based on our measurements in Section IV-B, 89.52% of Silk Road 4 transactions use the WitnessPublicKeyHash signature method, and 80.73% of these are 1-output transactions. Even if the majority of Silk Road 4 transactions follow this patterns, such information cannot be used as a reliable indicator regarding the unique characteristics of Silk Road 4 because there is a non-negligible quantity of non-compliant cases. Therefore, the algorithm for finding *MA*s does not utilize such signature patterns, and only checks the conditions of Definition 1 and the address type.

2) DATA MATCHING SUCCESS RATE

In this section, we evaluate the success rate of finding matched addresses for the 606 (V, D) tuples gathered using Algorithm 1. We found 31.68% matched addresses (i.e., 192 addresses) for the 606 review data among approximately 0.1 billion Bitcoin addresses. The remaining 68.32% included the cases in which two or more Bitcoin transactions were found with the same BTC as V in the review data. In this case, if more than one transaction is found, it is impossible to distinguish between the correct and false transactions associated with Silk Road 4 without additional conditions or information for identification.

In addition, if we were to include the WitnessPublicKeyHash signature and the 1-output transaction patterns as conditions for finding matched addresses, the matching rate would increase from 31.68% to 35.31%; however, it is evident that the additional 3.63% will include some false positives with non-negligible probability. Therefore, although adding more conditions may be helpful in finding more potentially associated addresses for Silk Road 4, this will inevitably increase false rates as a side effect if the additional condition is not always correct. Considering the fact that improving the accuracy and correctness (e.g., minimizing false positive rates) of the clustering results in Bitcoin clustering analysis is a more important requirement than increasing coverage (e.g., including additional, but potentially false addresses in a cluster) [10], the next Bitcoin clustering heuristic is applied only on the basis of the 31.68% matched addresses we found, even at the expense of losing utility by including some false negatives.

C. CLUSTERING

Algorithm 2 Clustering

Require: Matched Addresses Set $\{MA\}$
Ensure: Review_Cluster RC
  Initialize Review_Cluster $RC = \emptyset$
  for all $MA \in \{MA\}$ **do**
    Get cluster C of MA using blocksci
    $RC \leftarrow RC \cup C$
  end for
  Return RC

1) ALGORITHM

Matched addresses are actual Bitcoin addresses directly used for illicit transactions by the Silk Road 4 escrow and buyers and are controlled by the Silk Road escrow. Therefore, they must be linked as a single cluster of Silk Road 4. If the matched addresses have not been grouped, Algorithm 2, which outlines the proposed Multi-layer heuristic, further links them into a single cluster. Specifically, for the matched addresses included in different clusters (linked via the *Multi-input heuristic* and *Change address heuristic*), the proposed *Multi-layer heuristic* binds them into one cluster (which is the "review_cluster" in Algorithm 2) if they are associated with the same (V, D). Algorithm 2 shows *Multi-layer heuristic* algorithm for Bitcoin clustering.

2) CLUSTERING RESULTS

We evaluated the effect of our *Multi-layer heuristic* algorithm by measuring the distribution of clusters over time from May 14, 2020, the first appearance date of clusters involving *MA*s, to October 8, 2020, the last growth date of clusters. We first classified 133 unique clusters by removing duplicate clusters via the *Multi-input heuristic* and *Change address heuristic* among the clusters containing 192 *MA*s. Fig. 6 shows the numbers of clusters containing *MA*s and their size, x.

In the process, following the previous observations in [10] and [12] and we excluded a remarkably large number of clusters (11 clusters) to reduce false positives. Specifically, clusters with size greater than 10^4 were excluded from the clustering procedure. Among the 11 clusters, we identified that two clusters were already tagged as Binance [38] and Luno [42] clusters, respectively, and the remaining nine were unknown.

Fig. 7 depicts how the remaining 122 clusters changed over time using our *Multi-layer heuristic*. Fig. 7(a) shows the size of the Silk Road 4 cluster, which is the number of addresses controlled by Silk Road 4. In Fig. 7(a), each cluster is represented as an independent block using different colors, and the blue line indicates the Silk Road 4 cluster exactly determined at each time instance. As shown in Fig. 7(a), 122 different clusters that were not grouped by previous clustering heuristics are gradually merged together into a single Silk Road 4 cluster as time goes by.

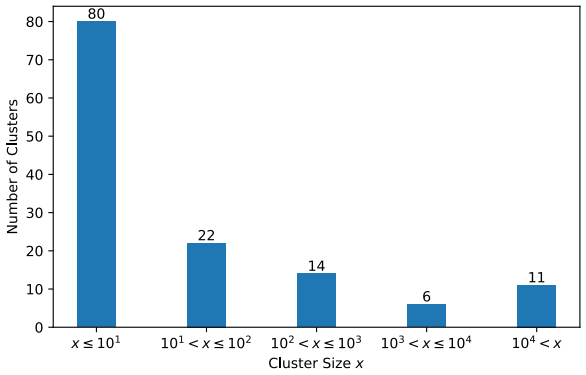


FIGURE 6. Number by Cluster Size.

TABLE 6. Clustering Result of Multi-layer Heuristic.

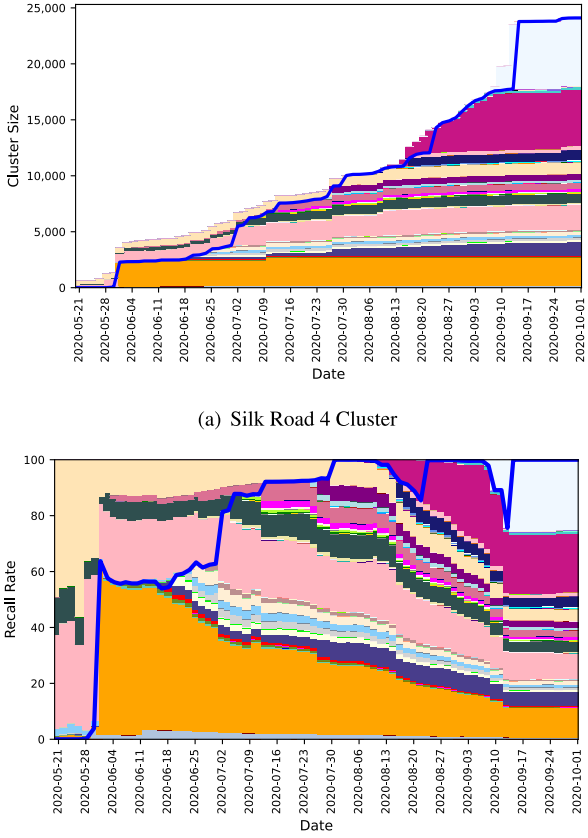
Cluster ID	Number of clusters	
	Previous heuristics	Proposed heuristic
Silk Road 4	0	122
Binance	1	1
Luno.com	1	1
Unknown	131	9

Fig. 7(b) shows the recall rate of the proposed clustering heuristic algorithm during the same period. The blue line represents the recall rate for each time instance. During the first week, the recall rate was almost 0%, because none of the review data were used for clustering. However, as more matched addresses are found using off-chain data, independent-looking clusters are merged into the Silk Road 4 cluster and the recall rate increases significantly. Although the Fig. 7(b) sometimes shows a slight decline in the recall rate when new clusters are found, it eventually increases to almost 100% as time elapses by taking advantage of the off-chain data.

Table 6 shows the clustering results using the proposed heuristic. When applying previous heuristics, 131 of 133 clusters were unknown. However, our *Multi-layer heuristic* could identify 122 out of 131 unknown clusters as Silk Road 4 clusters, which could then be further combined into a single cluster resulting in a reduction of 91.7% of false negatives. The clustering effect of the proposed algorithm is illustrated in Fig. 8. Fig. 8(a) shows the initial status of each transaction associated with 122 independent market clusters before the *Multi-layer heuristic* is applied and Fig. 8(b) shows the status of the same transactions after applying *Multi-layer heuristic*, where all of the 122 independent-looking market clusters in Fig. 8(a) are grouped as a single Silk Road 4 cluster. Therefore, the proposed Bitcoin clustering algorithm can help to analyze the illicit transactions of darknet markets and their hidden flows in a more transparent manner.

D. SIMULATION

We conducted a simulation to evaluate the reliability of our matching algorithm. Because it is infeasible in practice



(a) Silk Road 4 Cluster

(b) Recall Rate

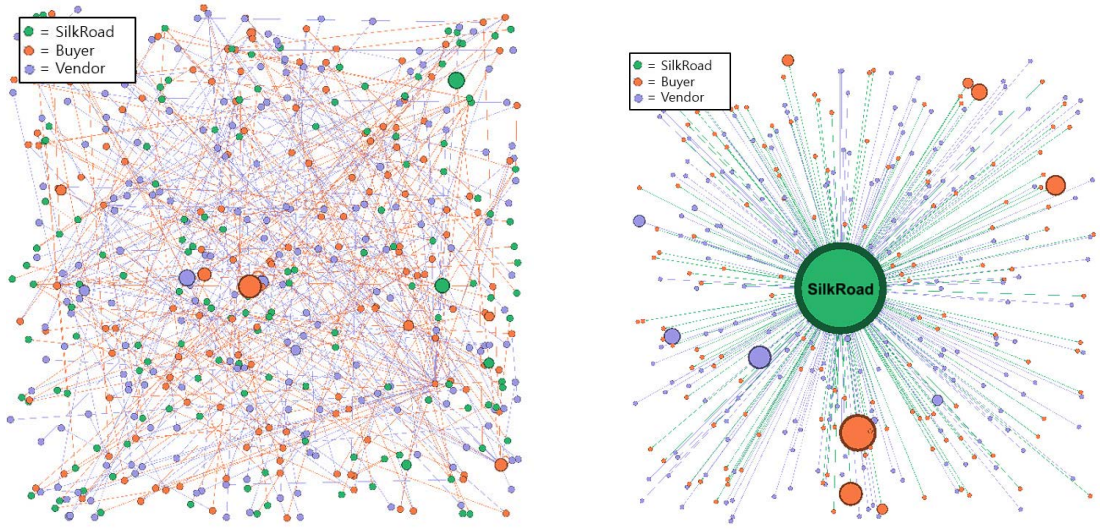
FIGURE 7. Changes of Clusters over Time.

to obtain ground-truth information for Silk Road 4 Bitcoin addresses, our simulation emulates the functionalities of Bitcoin as in a previous study [43] and its distributions in the real world. Our simulation aims to verify that our method can find Silk Road 4 transactions containing *MA*s among actual Bitcoin transactions. Specifically, the simulation proceeds by creating Bitcoin transactions following the actual Bitcoin distribution while hiding the transactions created by the collected review data, and then we determine how many such true positive transactions can be found by our clustering method.

The actual distribution of Bitcoin transactions is determined based on 40,415,437 transactions conducted from May 7, 2020 to September 14, 2020, over which the shipping period of the review data is included. The transactions are then classified into the corresponding types[4] based on the number of inputs and outputs, and then the distribution of the number and values of inputs and outputs for each type is calculated.

To simulate the behaviors of buyers who use Silk Road 4, we assume that they pay all at once (i.e., no partial payments) when purchasing, because our method only considers cases

[4]Each type is described in [44].



(a) Initial Status

(b) Final Status

FIGURE 8. Clustering Effect of Multi-layer Heuristic.

where the BTC in the review data exactly matches that of one input address included in $ITX(A)$ for some address A. In the case of pay-all-at-once, the address would be definitely MA if only one address meets the conditions we set; but in the case of partial payment, it would be almost infeasible to determine it because there is no way to know how many parts are divided. Under this assumption, 100 $OTX(MA)$ and 100 $ITX(MA)$ corresponding to 100 reviews would be concealed in our simulation data.

In the 100 $OTX(MA)$s, the MA addresses used for receiving the BTC value associated with each review are embedded as the ground truth. Because only these 100 $OTX(MA)$ and 100 $ITX(MA)$ can use the MAs as input or output addresses (because these are one-time addresses), other transactions cannot use these MAs.

In the simulation, 670,000 transactions were created, including approximately 1.1 million addresses, and we attempted to find the MAs using Algorithm 1 given in Section VI-B1. To avoid bias in the experiment, we conducted 10 simulations by randomly setting all parameters except for the distribution of the real Bitcoin blockchain. As a result, we found an average of 88.6 MAs and confirmed that they were all true positives (i.e., no false positives) by comparing them to the ground-truth assignations above. We also observed that there was an average of 11.4 addresses we missed (i.e., false negatives) because more than a single address met the conditions used by the algorithm (non-unique MAs) within the search space. Table 7 summarizes the simulation results using our method, resulting in an F1-score of approximately 0.94 based on 100% precision and 89.2% recall rates. It is important to note that our method did not produce any false positives in all simulations, which is a more important requirement for correct Bitcoin transaction analysis than increasing the coverage of clusters

TABLE 7. Simulation results (average of 10 simulations).

True Positive	False Positive	False Negative
88.6	0	11.4
Precision	Recall	F1-score
1.0	0.892	0.939201

(containing potentially false addresses). One reason for the higher match success rates shown in Table 7 compared to Table 5 may be because the simulation was conducted in an idealized environment where there are no partial payments and fewer non-unique MAs than in the real world. Finally, by applying existing heuristics ([10], [11]) to 670,000 transactions, we could confirmed that all 100 ground-truth addresses were included in different clusters, proving that our method can reduce the false negative rates remarkably compared to previous clustering methods. Details regarding the simulation and its source code can be found online.[5]

E. LIMITATIONS AND DISCUSSION

Our study only considered cases where the BTC in the review data exactly matched the BTC of one input address included in $ITX(A)$ for some address A. If the sum of BTCs corresponding to such multiple input addresses included in the $ITX(A)$ becomes uniquely equal to the BTC in the review data, then this must also be the case for matched addresses. Even if finding such combinations in a brute-force manner would work theoretically, it is highly impractical in practice. For example, approximately 200 blocks are currently created per day, and each block contains approximately 2,000 transactions on average in the real world. This means that 400,000 transactions are generated per day; thus, it is impractical to

[5]https://github.com/SREABS/BS

analyze the input addresses of all these transactions, compute every different combination of their BTCs, and find unique combinations for every given BTC in the review data using a brute-force approach. In addition, we observed that even a single Bitcoin transaction sometimes contains a tremendous number of addresses (more than 1,000 input addresses). If we could reduce such a gap between theory and practice and overcome the problem by some means, it is evident that more matched addresses would be found, which would help the proposed clustering heuristic algorithm to discover more relationships among the illicit transactions that have not been identified thus far.

Even if the proposed Bitcoin clustering algorithm can improve the clustering accuracy by disclosing hidden relationships among the addresses using off-chain data, the resulting cluster would also still be a part of the actual cluster. Because it is practically infeasible to acquire the ground-truth information exactly, we have no choice but to evaluate the clustering algorithms based only on the data we gathered. One way to obtain ground-truth information would be to engage in the trading process of the darknet market directly and gather Bitcoin information such as addresses, BTC, shipping information, trading date/time, and so on. However, owing to ethical issues, we could not gather information in such manner, and only observed both the on-chain and off-chain information that is publicly accessible, such as the address and signature patterns of Silk Road escrow systems. If more accurate ground-truth information were to become available, the accuracy of the clustering heuristics would also be improved accordingly. However, because it is practically challenging to obtain hidden ground-truth information in the real world, this remains an open problem that every Bitcoin clustering technique fundamentally confronts.

VII. CONCLUSION

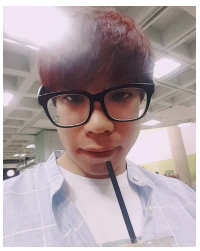
In this study, we propose a multi-layer Bitcoin address clustering method using both blockchain-layer and application-layer information to resolve the false-negative problem associated with existing heuristics. To obtain useful off-chain data from the application layer, we conducted a comprehensive analysis of the data available in the darknet market, primarily focusing on Silk Road 3.1 and 4, and analyzed their unique characteristics such as the address escrow system. Based on the analysis, we found that approximately 31.68% of addresses in the 606 Silk Road 4 review data we collected matched actual Bitcoin transactions, and disclosed a total of 122 hidden clusters of Silk Road 4 which could not be identified using the previous clustering methods. The proposed algorithm can complement various existing clustering methods and significantly reduce the false negative rate by up to 91.7%. To evaluate its accuracy, we performed a simulation following the distribution of actual Bitcoin transactions in the real world. Our method achieved 100% precision and 89.2% recall rates without producing any false positives, demonstrating its potential efficacy in the real world.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized Bus. Rev.*, vol. 21260, pp. 1–9, Oct. 2008.
[2] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, vol. 151, pp. 1–32, Apr. 2014.
[3] S. Noether and A. Mackenzie, "Ring confidential transactions," *Ledger*, vol. 1, pp. 1–18, Dec. 2016.
[4] D. Hopwood, S. Bowe, T. Hornby, and N. Wilcox, "Zcash protocol specification," Version 2018.0-beta-33, Zerocoin Electr. Coin Company, Lakewood, CO, USA, Nov. 2018.
[5] S. Foley, J. R. Karlsen, and T. J. Putniņš, "Sex, drugs, and Bitcoin: How much illegal activity is financed through cryptocurrencies?" *Rev. Financial Stud.*, vol. 32, no. 5, pp. 1798–1853, May 2019, doi: 10.1093/rfs/hhz015.
[6] Google. *Google Chrome*. Accessed: Apr. 16, 2022. [Online]. Available: https://www.google.com/chrome/
[7] Mozilla. *Firefox*. Accessed: Apr. 16, 2022. [Online]. Available: https://www.mozilla.org/en-US/firefox/
[8] Tor. *Tor Browser*. Accessed: Apr. 16, 2022. [Online]. Available: https://www.torproject.org
[9] M. W. A. Nabki, E. Fidalgo, E. Alegre, and I. de Paz, "Classifying illegal activities on Tor network based on web textual contents," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, Valencia, Spain, vol. 1, 2017, pp. 35–43.
[10] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: Characterizing payments among men with no names," in *Proc. Conf. Internet Meas. Conf.*, Blagoevgrad, Bulgaria, Oct. 2013, pp. 127–140.
[11] J. D. Nick, "Data-driven de-anonymization in Bitcoin," M.S. thesis, Distrib. Comput. Group, Comput. Eng. Netw. Lab., ETH Zürich, Zürich, Switzerland, 2015.
[12] D. Ermilov, M. Panov, and Y. Yanovich, "Automatic Bitcoin address clustering," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 461–466, doi: 10.1109/ICMLA.2017.0-118.
[13] Z. Zhang, T. Zhou, and Z. Xie, "BITSCOPE: Scaling Bitcoin address deanonymization using multi-resolution clustering," in *Proc. HICSS*, Hilton Waikoloa Village, HI, USA, 2018, pp. 1–11.
[14] S. Goldfeder, H. Kalodner, D. Reisman, and A. Narayanan, "When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies," 2017, *arXiv:1708.04748*.
[15] J. Schäfer, C. Müller, and G. Armknecht, "If you like me, please don't 'like' me: Inferring vendor Bitcoin addresses from positive reviews," in *Proc. PETS*, vol. 1. Sydney, NSW, Australia, 2022, pp. 440–459.
[16] SilkRoad4. *Silk Road 4*. Accessed: Dec. 13, 2021. [Online]. Available: http://silkroad7rn2puhj.onion/
[17] J. Aldridge and R. Askew, "Delivery dilemmas: How drug cryptomarket users identify and seek to reduce their risk of detection by law enforcement," *Int. J. Drug Policy*, vol. 41, pp. 101–109, Mar. 2017, doi: 10.1016/j.drugpo.2016.10.010.
[18] F. Sabry, W. Labda, A. Erbad, H. A. Jawaheri, and Q. Malluhi, "Anonymity and privacy in Bitcoin escrow trades," in *Proc. 18th ACM Workshop Privacy Electron. Soc. (WPES)*, London, U.K., 2019, pp. 211–220.
[19] S. Delgado-Segura, C. Pérez-Sola, G. Navarro-Arribas, and J. Herrera-Joancomartí, "Analysis of the Bitcoin UTXO set," in *Proc. Int. Conf. Financial Cryptogr. Data Secur.*, 2018, pp. 78–91.
[20] A. Gavin, "Pay to script hash," Bitcoin Improvement Proposal-67, Jan. 2012.
[21] K. Thomas, R. Jean-Pierre, and V. Ruben, "Deterministic pay-to-script-hash multi-signature addresses through public key sorting," Bitcoin Improvement Proposal-67, Feb. 2015.
[22] Blockchain. *Blockchain.com*. Accessed: Jan. 29, 2022. [Online]. Available: https://www.blockchain.com/
[23] P. Koshy, D. Koshy, and P. McDaniel, "An analysis of anonymity in Bitcoin using P2P network traffic," in *Proc. Int. Conf. Financial Cryptogr. Data Secur.*, Barbados, Caribbean, 2014, pp. 469–485.
[24] N. T. Courtois and R. Mercer, "Stealth address and key management techniques in blockchain systems," in *Proc. 3rd Int. Conf. Inf. Syst. Secur. Privacy*, Porto, Portugal, 2017, pp. 559–566.
[25] G. Maxwell. *CoinJoin: Bitcoin Privacy for the Real World*. Accessed: Apr. 28, 2022. [Online]. Available: https://bitcointalk.org/index.php?topic=279249.0
[26] Bitcoin Project. *Bitcoin Core*. Accessed: Jan. 29, 2022. [Online]. Available: https://bitcoin.org/en/bitcoin-core/
[27] Bitcoin Project. *Bitcoin Wallet*. Accessed: Jan. 29, 2022. [Online]. Available: https://bitcoin.org/en/wallets/mobile/android/bitcoinwallet/

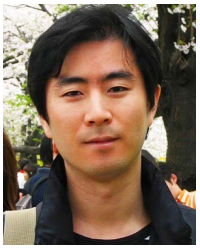
[28] N. Christin, "Traveling the silk road: A measurement analysis of a large anonymous online marketplace," in *Proc. WWW*, Rio de Janeiro, Brazil, 2013, pp. 213–224.
[29] J. Van Buskirk, S. Naicker, R. Bruno, C. Breen, and A. Roxburgh, *Drugs and the Internet*. Drugs and New Technologies (DNeT) Bulletins, 2016, pp. 1–14.
[30] M. Horton-Eddison, "Updating escrow: Demystifying the CDM multisig process," GDPO, SWS UK, Lancaster, U.K., Tech. Rep., 2017.
[31] A. E. Bahrawy, L. Alessandretti, L. Rusnac, D. Goldsmith, A. Teytelboym, and A. Baronchelli, "Collective dynamics of dark web marketplaces," *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, Dec. 2020, Accessed: Apr. 28, 2022, doi: 10.1038/s41598-020-74416-y.
[32] W. Lacson and B. Jones, "The 21st century DarkNet market: Lessons from the fall of silk road," *Int. J. Cyber Criminol.*, vol. 10, no. 1, p. 40, 2016.
[33] M. Horton-Eddison and M. Di Cristofaro, "Hard interventions and innovation in crypto-drug markets: The ESCROW example," *Policy Brief*, vol. 11, pp. 16–27, Oct. 2017.
[34] V. Adewopo, B. Gonen, S. Varlioglu, and M. Ozer, "Plunge into the underworld: A survey on emergence of darknet," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2019, pp. 155–159, doi: 10.1109/CSCI49370.2019.00033.
[35] H. Kalodner, M. Möser, K. Lee, S. Goldfeder, M. Plattner, A. Chator, and A. Narayanan, "BlockSci: Design and applications of a blockchain analysis platform," in *Proc. USENIX Secur.*, 2020, pp. 2721–2738.
[36] *Apollon Market*. Accessed: Sep. 24, 2021. [Online]. Available: http://apollionih4ocqyd.onion
[37] *Apollon Market*. Accessed: Sep. 24, 2021. [Online]. Available: http://agarthaourmnyhq3.onion
[38] *Binance*. Accessed: Jan. 29, 2022. [Online]. Available: https://www.binance.com/en
[39] *Kraken*. Accessed: Jan. 29, 2022. [Online]. Available: https://www.kraken.com/
[40] *Satoshi Dice*. Accessed: Jan. 29, 2022. [Online]. Available: https://satoshidice.com/
[41] *Jupyter*. Accessed: Jan. 29, 2022. [Online]. Available: https://jupyter.org/
[42] *Luno*. Accessed: Jan. 29, 2022. [Online]. Available: https://www.luno.com/
[43] E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, "Evaluating user privacy in Bitcoin," in *Proc. FC*, Okinawa, Japan, 2013, pp. 34–51.
[44] S. Phetsouvanh, A. Datta, and F. Oggier, "Analysis of multi-input multi-output transactions in the Bitcoin network," *Concurrency Comput., Pract. Exper.*, vol. 33, no. 1, p. e5629, Jan. 2021.
[45] S. J. Murdoch and G. Danezis, "Low-cost traffic analysis of Tor," in *Proc. IEEE Symp. Secur. Privacy*, May 2005, pp. 183–195, doi: 10.1109/SP.2005.12.
[46] P. Mittal, A. Khurshid, J. Juen, M. Caesar, and N. Borisov, "Stealthy traffic analysis of low-latency anonymous communication using throughput fingerprinting," in *Proc. 18th ACM Conf. Comput. Commun. Secur. (CCS)*, Toronto, ON, Canada, 2011, pp. 215–226.
[47] A. Kwon, M. A. Sabah, D. Lazar, M. Dacier, and S. Devadas, "Circuit fingerprinting attacks: Passive deanonymization of Tor hidden services," in *Proc. USENIX Secur.*, Washington, DC, USA, 2015, pp. 287–302.
[48] V. Ciancaglini, M. Balduzzi, M. Goncharov, and R. McArdle, "Deepweb and cybercrime," Trend Micro, Tokyo, Japan, Tech. Rep., 2013, pp. 5–6.
[49] V. Ciancaglini, M. Balduzzi, R. McArdle, and M. Rösler, "Below the surface: Exploring the deep web," Trend Micro, Tokyo, Japan, Tech. Rep., 2015, pp. 1–48.
[50] K. Soska and N. Christin, "Measuring the longitudinal evolution of the online anonymous marketplace ecosystem," in *Proc. USENIX Secur.*, Washington, DC, USA, 2015, pp. 33–48.
[51] I. Sanchez-Rola, D. Balzarotti, and I. Santos, "The onions have eyes: A comprehensive structure and privacy analysis of Tor hidden services," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 1251–1260.
[52] A. Biryukov, I. Pustogarov, F. Thill, and R.-P. Weinmann, "Content and popularity analysis of Tor hidden services," in *Proc. IEEE 34th Int. Conf. Distrib. Comput. Syst. Workshops*, Jun. 2014, pp. 188–193, doi: 10.1109/ICDCSW.2014.20.
[53] A. Biryukov, I. Pustogarov, and R. Weinmann, "Trawling for Tor hidden services: Detection, measurement, deanonymization," in *Proc. IEEE Symp. Secur. Privacy*, May 2013, pp. 80–94, doi: 10.1109/SP.2013.15.
[54] R. Van Wegberg, S. Tajalizadehkhoob, K. Soska, U. Akyazi, C. H. Ganan, B. Klievink, and M. Van Eeten, "Plug and prey? Measuring the commoditization of cybercrime via online anonymous markets," in *Proc. USENIX Secur.*, Baltimore, MD, USA, 2018, pp. 1009–1026.
[55] M. J. Barratt, J. A. Ferris, and A. R. Winstock, "Use of silk road, the online drug marketplace, in the United Kingdom, Australia and the United States," *Addiction*, vol. 109, no. 5, pp. 774–783, May 2014, doi: 10.1111/add.12470.
[56] S. Lee, C. Yoon, H. Kang, Y. Kim, Y. Kim, D. Han, S. Son, and S. Shin, "Cybercriminal minds: An investigative study of cryptocurrency abuses in the dark web," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, San Diego, CA, USA, 2019, pp. 1–15.
[57] J. Broséus, D. Rhumorbarbe, C. Mireault, V. Ouellette, F. Crispino, and D. Décary-Hétu, "Studying illicit drug trafficking on darknet markets: Structure and organisation from a Canadian perspective," *Forensic Sci. Int.*, vol. 264, pp. 7–14, Jul. 2016, Accessed: Apr. 28, 2022, doi: 10.1016/j.forsciint.2016.02.045.
[58] J. Demant, R. Munksgaard, and E. Houborg, "Personal use, social supply or redistribution? Cryptomarket demand on silk road 2 and Agora," *Trends Organized Crime*, vol. 21, no. 1, pp. 42–61, Mar. 2018, doi: 10.1007/s12117-016-9281-4.

MINJAE KIM received the B.S. degree in computer science from Korea University, Seoul, South Korea, in 2021, where he is currently pursuing the M.S. degree with the Department of Computer Science and Engineering, College of Informatics. His research interests include safeAI and blockchain.



JINHEE LEE received the B.S. degree in information and communications engineering from Hansung University, Seoul, South Korea, in 2018, and the M.S. degree in Department of computer science and engineering from Korea University, Seoul, in 2021. He is currently researching and developing in cryptography with Dream Security. His research interests include information security and blockchain security.

HYUNSOO KWON received the B.S. degree from Chung-Ang University, Seoul, South Korea, in 2014, and the M.S. and Ph.D. degrees from Korea University, Seoul, in 2016 and 2020, respectively, all in computer science. He is currently a Senior Researcher with Samsung Electronics, South Korea. His research interests include information security, network security, and cloud computing security.



JUNBEOM HUR received the B.S. degree from Korea University, Seoul, South Korea, in 2001, and the M.S. and Ph.D. degrees from KAIST, in 2005 and 2009, respectively, all in computer science. He was a Postdoctoral Researcher with the University of Illinois at Urbana–Champaign, from 2009 to 2011. He was an Assistant Professor with the School of Computer Science and Engineering, Chung-Ang University, South Korea, from 2011 to 2015. He is currently a Professor with the Department of Computer Science and Engineering, Korea University. His research interests include information security, cloud computing security, network security, and applied cryptography.