Detect Your Fingerprint in Your Photographs: Photography-based Multi-Feature Sybil Detection

Yerim Kim
Korea University
Seoul, Republic of Korea
yrkim@isslab.korea.ac.kr

Minjae Kim
Korea University
Seoul, Republic of Korea
mjkim@isslab.korea.ac.kr

Myungjae Chung
Korea University
Seoul, Republic of Korea
qa7028@korea.ac.kr

Junbeom Hur
Korea University
Seoul, Republic of Korea
jbhur@korea.ac.kr

ABSTRACT

A darknet market is an online marketplace typically implemented over Tor, where vendors sell illegal products or criminal services. Due to dramatic growth in the popularity of such markets, there is a recognized need for automatic investigation of the market's ecosystem and identification of anonymous vendors. However, as they often create multiple accounts (or *Sybil* accounts) within or across different marketplaces, detecting Sybil accounts becomes the key to understanding the ecosystem of darknet markets and identifying the actual relationship between the vendors.[1] This study presents a novel Sybil detection method that extracts multiple features of vendors from photographs in a fine-grained level (e.g., image similarity, main category, subcategory, and text data), and reveals the multiple Sybil accounts of them simultaneously. Each feature is extracted from multiple rich sources using an image hash algorithm, Deep Neural Network (DNN) classifier, image restoration, and text recognition tool; and merged using a weighted feature embedding model. The matching score of each vendor is then calculated to identify not only the exact Sybil accounts, but multiple potential accounts suspected of being associated to a single operator. We evaluate the efficacy of our method using real-world datasets from four large darknet markets (i.e., SilkRoad2, Agora, Evolution, Alphabay) from 2014 to 2015. Because of the anonymity of darknet market, we construct the ground-truth of Sybil accounts by randomly splitting the dataset of vendors into two even parts. We used the first set to train the model, and linked the second set to the original vendor in the first set to evaluate performance. Our experimental results demonstrated that the proposed method outperforms the existing photography-based system with an accuracy of 98%, identifying up to 700% more candidate Sybil accounts than prior work [27]. Additionally, our method detects multiple Sybil accounts for 90% of evaluated test cases, presenting a very different picture of darknet marketplace dynamics than methods that can only detect a single Sybil account for each target vendor. Due to its fine-grained multiple feature extraction from photographs, our method can be more generically applied to various darknet markets for Sybil detection regardless of their languages or categories of items.

KEYWORDS

Darkweb, Sybil detection, De-anonymization

1 INTRODUCTION

Over the past decades, there has been a dramatic increase in darknet markets in terms of their volume and activities providing illegal products or criminal services such as drugs, credit cards, pornography, software exploits, and criminal/hacking services [3]. The darknet market is located in the darknet, a subsection of the deep web accessible only through specific browsers such as Tor Browser [1]. Because Tor makes it complicated to track the Internet activities of users by concealing their IP addresses and locations, a number of darknet websites are typically operated anonymously over Tor.

In order to further increase anonymity and avoid tracing of activities of darknet users, the darknet markets frequently transform their IP addresses or domain names [2, 6, 7], and allow only cryptocurrencies like Bitcoin [4] or Ethereum [5] for the anonymous transactions [8–16]. In addition, most of the useful information for user identification is hidden by the darknet market operators or the other additional anonymous technologies [17–23]. Especially, many darknet market vendors create multiple accounts, called *Sybil* accounts, within the same market or across different darknet marketplaces. Sybil accounts refer to the accounts controlled by the same person or group for gaining an advantage such as manipulating reputation in reviewing systems or promoting social network accounts. Hence, it becomes much more challenging to investigate the darknet markets and identify their users. Thus, linking multiple accounts and detecting Sybil accounts of the same vendors play crucial roles in accurately identifying vendors, understanding the actual transactions of illegal activities, and evaluating the exact volume of vendors and illegal items advertised across the different darknet markets. Unfortunately, due to the explosive growth of the number of markets [2], it is not feasible to manually analyze and connect multiple accounts owned by the same vendor in practice. Thus, in order to understand the ecosystem of the real-world darknet markets in the presence of such challenges, developing an

[1] There is a possibility for Sybil detection method to be misused for de-anonymizing individuals seeking to leverage pseudo-anonymous online marketplaces to conduct normal anonymous online commerce or even social media activity. Legitimate individuals may preserve their privacy by considering our findings.

Proceedings on Privacy Enhancing Technologies 2023(1), 244–263

https://doi.org/10.56553/popets-2023-0015

efficient and autonomous Sybil detection method is of a considerably necessity.

Several approaches have been proposed to address this problem: one is stylometry-based approach and the other is photography-based one. In the former [24–26], vendors' writing styles have been utilized; while in the latter [27, 28], vendors' distinct photography style (e.g., the product images that vendors have uploaded) have been utilized to detect the Sybil accounts of the same vendors.

Unfortunately, stylometry-based approaches [24–26] face several limitations. First, since the only available text data in the darknet market is likely to be a product description or vendor profile, which are often short, repetitive, and sometimes even written by using a specific template (see Fig. 1 [62]), the detection accuracy drops significantly when duplicated sentences are removed [27]. Second, it is inherently sensitive to the language of the dataset. Since the current darknet market is becoming increasingly diverse in its languages [59] as shown in Fig. 2, it is becoming more difficult to apply a stylometry-based method to the other markets using different languages. Third, the previous stylometry-based methods [26, 28] only utilized drugs, which constitute only a small portion of various categories of products in the real-world darknet market.

On the other hand, the previous photography-based methods [27, 28] only regarded photography as a single feature. Specifically, they extracted only an image feature from the photographs as a single vector, and just matched replicated photos as a proof of Sybil accounts. Thus, the performance of their detection was limited when duplicated images were removed.

In this paper, we propose a novel Sybil detection method to extract multi-features from photographs and leverage each feature to link different accounts to the same vendor simultaneously. Our method characterizes the vendors as follows. First, the similarity between two images is evaluated using the image hash function [32, 52, 53], which is highly efficient in generating the representation of correlated vendors with equivalent or similar photos [31]. Second, the main and subcategories of images are extracted using the Deep Neural Network (DNN) model [33, 34], and the category distributions of the products of each vendor are produced. These 'inventory' features are difficult for any vendors to copy, who may simply pirate images from another vendor or the Internet [40], enhancing Sybil accounts detection accuracy. Third, text data is extracted from the product images (if there are any) to figure out additional features that cannot be captured using the image hash. However, as most of the darknet market photos have low image quality in practice, high accuracy cannot be achieved by simply applying the text extraction tool to darknet market photos [35]. We thus apply a super-resolution [37] method to improve the performance of text recognition [36, 38]. All of these features are finally merged using a weighted feature embedding model [26]. Then, the possibility of the vendor with the highest likelihood of being the same person (the primary vendor) and those of vendors with lower, yet significant likelihood of being the same person (the candidate vendors) are calculated, and multiple Sybil accounts are linked to each vendor on the basis of their possibility scores.

Our approach is novel in that it can find multiple candidate accounts that are considerably likely to be of the same vendor, as opposed to the previous methods [26–28] that can only determine if a single target account is another Sybil account of a specific vendor. When the most similar account turns out not to be of the same vendor, the previous schemes have to run the model iteratively after removing mismatched vendors until they find the correct one.

Additionally, our method achieves much higher accuracy and coverage by categorizing most of the items currently available in the real-world darknet market, training the DNN-based detection model with the multiple features extracted from them in a fine-grained level, and detecting Sybil accounts automatically across all vendors in the darknet market, compared to the previous approaches [26, 28] that only dealt with drugs and their paraphernalia.

We evaluate our method using real-world datasets from four large darknet markets (i.e., SilkRoad 2, Agora, Evolution, Alphabay [51]). A ground-truth evaluation shows that our method achieved an accuracy of 98%, outperforming the previous photography-based Sybil detection scheme [27] in both accuracy and coverage. Our method covered up to 700% more vendors than the image classification approach using ResNet-50 [34]. In addition, our method could identify multiple Sybil accounts for cases in which a vendor has multiple accounts. Specifically, we could find that 90% of all of the corresponding accounts for vendors have at least two additional Sybil accounts. Further case studies show that our method could also detect previously unknown Sybil accounts in the real-world.

Contributions. In summary, this paper makes the following contributions:

- We propose the first method to fingerprint Sybil accounts of darknet vendors by extracting multi-features from their photographs. Experimental result in SilkRoad2 and Agora darknet markets shows that our method outperforms the existing photography-based schemes with an accuracy of 98%.
- Unlike the previous approaches, our method identifies multiple candidate vendors simultaneously based on their similarity scores, which are evaluated based on the multiple features extracted from the categorized in a fine-grained level.
- Our method has a higher coverage compared to the recent photography-based schemes [27], covering up to 700% more vendors. Additionally, we demonstrate the proposed scheme can be generically applied to various darknet markets such as Evolution and Alphabay regardless of their languages or categories of items, showing high applicability of the proposed method in practice.

Organization. The remainder of the paper proceeds as follows: Section 2 introduces the background and related works. Section 3 shows the darknet dataset we used for our scheme construction and the experiment. Section 4 proposes our Sybil detection method, consisting of feature extraction, feature embedding, and vendor identification. Section 5 and 6 show the experimental results of the performance and applicability of our method. Section 7 discusses the practical implication of our findings, and future research directions. Section 8 concludes the paper.

2 BACKGROUND & RELATED WORKS

In this section, we introduce the background of darknet market and the previous approaches for darknet user data analysis.

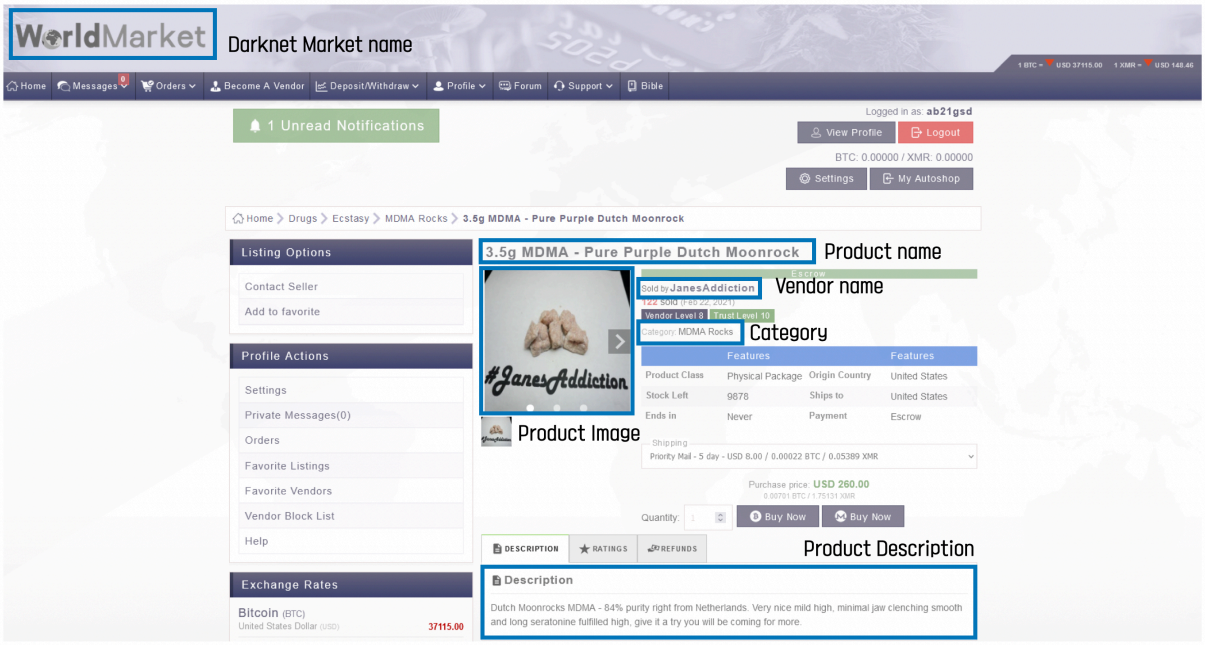


Figure 1: Snapshot of a darknet market.

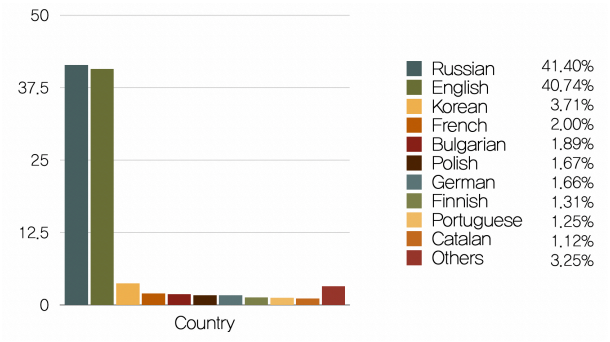


Figure 2: Darknet market share by language region.

2.1 Anonimity of Darknet Market

The darknet market is a specific type of website in the darknet mostly set up by criminals for the purpose of trading illegal goods or stolen items. Frequently traded items include: drugs, credit cards, pornography, software exploits, and criminal/hacking services [3]. With the development of technology for Tor [1], the number of hidden services has continuously increased, and the Tor darknet has had a maximum of more than 120,000 sites, whereas it currently hosts about 80,000, and the number of users was approximately two million in 2019 [2]. In the early days of the darknet market, SilkRoad2, Agora, and Blackbank flourished, where in contrast Alphabay, Vice city, Versus market, etc., are the largest darknet markets nowadays [50].

In order to avoid the traceability of regulatory agencies, the darknet market frequently replaces the address of sites or uses new domain names, and only allows cryptocurrencies such as Bitcoin [4] or Ethereum [5] for its transactions, making it difficult to trace transactions and identify the users involved in them.

2.2 Darknet User Data Analysis

In order to analyze the behaviors of users and identify their hidden relationships in the darknet market, several data analysis methods have been proposed in diverse aspects.

Crawling Darknet User Data. Efficient darknet data crawling is of significant importance as a *priori* technology to monitor the darknet market or forum-based communities. Ghosh et al. [18] developed an automated tool for crawling and indexing content from onion sites into a large-scale data repository with over 100 million pages. Since the darknet sites often have functions to detect and prevent crawling activities, Campobasso et al. [17] recently developed a data crawling tool for semi-automatically learning the structure of darknet forums to overcome the problem.

Finding Hidden Relationships of Users. In order to find the hidden relationship of darknet users, several studies have been proposed to identify key players of the underground forum [19, 20]; while some other works analyzed roles, influence levels, and social relationships of key actors [21]. Sun et al. [22] proposed a method that utilizes uploaded text data from forums to reveal private interactions among users, and Bada et al. [23] explored the prominent use cases of the specific search engine Shodan to highlight hackers' targets and motivations.

Detecting Sybil Accounts. Exact tracing and linking multiple accounts of the same user is a key to the identification of the darknet market's ecosystem by understanding their actual behaviors.

In the darknet market, users, typically vendors, often create multiple accounts within or across different markets. Numerous studies have been proposed to detect these Sybil accounts automatically. The previous Sybil detection methods can be classified into the following two approaches: one is a stylometry-based approach and the other is a photography-based one.

Stylometry-based approaches mainly utilize a technique that captures the distinctive writing style of each user. Some of the existing approaches analyze the stylometry based on datasets revolving around vendor introduction, product description [24, 25], as well as meta-information such as product category and shipping location [26]. The stylometry-based methods suffer from two significant problems when being applied to the real-world darknet market. First, the lengthy text is required to make stylometry analysis effectively. Unlike the underground forums where rich and diverse text dataset exist, the text dataset extracted from the darknet market is only limited to product descriptions or vendor introductions that are short and repetitive, often following a certain template. As demonstrated in [27], the accuracy of Sybil detection using text data becomes dramatically decreased as overlapping data is removed. Second, stylometry-based approaches inherently depend on the language used in the target darknet market. Considering product descriptions are often written in different languages even within a single darknet market in practice, the detection model needs to be trained or generated again when the target language is changed, impeding its applicability.

In order to overcome the limitations of stylometry-based approaches, Wang et al. [27] recently proposed a photography-based analysis scheme using deep learning models [27], while Zhang et al. [28] suggested a hybrid approach that integrates both the stylometry and photography styles using an Attributed Heterogeneous Information Network (AHIN). However, the previous photography-based methods [27, 28] only regarded photography as a single feature. Specifically, they extracted only an image feature from the photographs as a single vector, and just matched replicated photos as proof of Sybil accounts. Thus, the performance of their detection was limited when duplicated images were removed, implying they could hardly capture the unique photography style of the actual vendors in the real world.

Detecting Sybil accounts can be applied to other contexts such as online marketplace (e.g., craigslist, social commerce, online shopping mall) or social media (e.g., Facebook, Twitter, Youtube). Some malicious users create multiple accounts to upload fake ratings or reviews for advertising their brand, and mimic other users' accounts or communities to earn a reputation and upload malicious content [46]. Such malicious users' accounts can be identified by the Sybil detection method. Likewise, results demonstrated in the context of a clandestine marketplace, may generalize to other contexts where analogous de-anonymization strategies could be leveraged against legitimate users of broader online (social) ecosystems and across social platforms.

3 DATASET

We use the public archive of darknet market datasets [51] to analyze the linkability of multiple accounts of the vendors. It contains the daily or weekly crawling data of 89 darknet markets from 2013 to 2015. In this paper, we intensively explore all of its crawling data and finally choose four large markets containing many image data with various product categories: SilkRoad2, Evolution, Agora, and Alphabay.

Data Pre-processing. We generate a data pre-processing tool to extract meaningful information from the crawled dataset in diverse aspects. Specifically, it extracts the ID, product image, product category, and PGP key of each vendor from the dataset of each market. Since most of the market data were scraped differently with distinct formats and structures, we additionally fine-tuned our pre-processing tool appropriately to each market. Table 1 shows the basic statistics of the darknet market dataset used in our experiment.

Market	Unique Vendor	Duplicated Image	Deduplicated Image	Time Span
SilkRoad2	1,188	51,745	32,622	2014/01-2014/11
Evolution	2,930	58,189	28,533	2014/01-2015/03
Agora	2,950	82,527	38,301	2014/01-2015/07
Alphabay	1,472	17,964	10,504	2014/12-2015/07
Total	8,540	210,425	109,960	2014/01-2015/07

Table 1: Summary of the dataset. 'Unique Vendor' indicates the number of vendors with different vendor ID and PGP key; 'Duplicated Image' includes all of the vendor's photos; and 'Deduplicated Image' excludes all of the duplicated photos used for different products.

Ethics of Data Analysis. Our study follows the standard ethical practices for analyzing our datasets [29, 30], and only analyzes the datasets that are made publicly available under the Creative Commons CC0 license [63] as in the previous research [27–30]. First, the dataset only contains publicly available information without any personally identifiable information. Second, our dataset only covers darknet markets that have been taken down by authorities. Third, our dataset does not contain any form of interaction with human subjects. Finally, our research offers useful tools to researchers or law enforcement for tracing, monitoring, and investigating crimes from darknet markets to make the benefits of our research outweigh the potential risks. The analysis follows the 'beneficence' principle in the Belmont report [43] as also claimed in [42, 44].

4 PROPOSED METHOD

This section describes our multi-feature-based Sybil detection method. We first describe the overview of our scheme and then introduce the detailed approaches to extract features from photographs, embed the features into a model, and determine Sybil accounts.

4.1 Model Overview

Let us denote the set of n vendors V as $\{v_1, \cdots, v_n\}$. We define a *target vendor* as the vendor whose multiple accounts we are searching for and *compared vendors* as the set of vendors who are likely to be the same as the target vendor. Each vendor v_i has

a photograph list $P_i = \{p_1, \cdots, p_m\}$, where m is the number of photos that v_i has. Our method extracts four main features such as image similarity, main category, subcategory, and text data from photographs, and embeds the merged features into the vendor identification model. We use an image hash tool to evaluate the similarity of the group of photos that the vendor has uploaded. We calculate the category distribution of the products that each vendor contains. We obtain text data from the photos to figure out detailed features that cannot be extracted from image hashing. All of these features are merged using a weighted feature embedding model. We show the connectivity of vendors with different accounts using final similarity scores and suggest candidate vendors having the meaningful possibility of being the same vendor as well. The overview of our method is shown in Fig. 3. The flow of feature extraction and embedding is represented in Fig. 4.

4.2 Feature Extraction

We describe how to capture multiple features (i.e., image similarity, main category, subcategory, text) from the photographs of each product that a vendor has uploaded.

4.2.1 Image Similarity Feature. We obtain the first feature that is image similarity, using image hash algorithms, which tell us whether two images look nearly identical. Unlike cryptographic hash algorithms (*e.g.*, MD5, SHA), where tiny changes in the image give completely different hashes, image hash algorithms can help generate the representation of correlated vendors with equivalent or similar photos, since they generate similar output hashes when the input images are similar.

We carefully investigated the Python implementations of Average Hash (AHASH), Difference Hash (DHASH), Perception Hash (PHASH), and Wavelet Hash (WHASH) [52] as the candidate image hashing techniques, and compared them to determine which is the most accurate for calculating identical hashes from equivalent or similar images of real-world darknet markets (same-origin-photos similarities). After converting images to image hash values using each hash algorithm, images were grouped by hash and hash type. For each group, we applied the Structural Similarity Index Metrix (SSIM) [53] to each possible pair of images belonging to the same group to quantify the level of hash accuracy for image matching in darknet marketplaces. The SSIM is a metric to determine the similarity between two photos (e.g., the SSIM value will be 1.0 for two identical photos and 0 for entirely different photos). After calculating the SSIM value of each image pair in a group, the results were averaged according to the number of pairs. We also calculated the weighted average SSIM for each of the four hash types. The weighted average SSIM provides each group's SSIM value with a weight determined by the number of image pairs from the group of the same hash value and hash type [31], such that a group with a large number of images more greatly influences the overall weighted averaged SSIM than a group with only a small number of images. Table 2 shows the hash analysis results for each hash technique.

For analyzing the different-origin-photos similarities of each hash function as well as the same-origin-photo similarities, we also examined image pairs with different hash values. Consequently, there were cases where hash values were the same but the images were different; but there were no cases where hash values were different but the images were the same. Therefore, same-origin-photo similarities differed depending on the hash functions, but different-origin-photo similarities were identical for all hash functions (which is zero). Hence, we only compared the same-origin-photo similarities for choosing the hash function.

Hash Type	SilkRoad2		Agora	
	Weighted Avg SSIM	Avg SSIM	Weighted Avg SSIM	Avg SSIM
AHASH	0.887	0.895	0.866	0.898
DHASH	0.929	0.931	0.936	0.942
PHASH	0.917	0.915	0.907	0.915
WHASH	0.738	0.890	0.748	0.904

Table 2: Hash analysis result. SSIM is a metric to determine the similarity between two photos, which quantifies the level of hash accuracy for image matching. The weighted average SSIM provides each group's SSIM value with a weight determined by the number of image pairs.

To obtain the representation for a vendor v_i, we calculated the value of the image hash for all of the photos of the vendor as a list $H_i = \{h_1, \cdots, h_m\}$, where each h_j represents the hash value of image p_j. Following the results of Table 2, DHASH was determined to be the most effective algorithm for image hashing. After the image hashing results were generated, we calculated *Hamming Distance*, which is a well-known method for calculating the difference of the hash value, to compare the similarity of hash results to find the most similar vendor for each v_i [32]. The Hamming distance is a metric to compare the similarity of binary strings, such that the value of Hamming distance will be 0 for two identical strings. For every different value between the two strings, this number is incremented by one. We evaluated the Hamming distance between each image from all compared vendors and target vendors by generating a similarity matrix. For each image from the target vendor, the lowest Hamming distance value is stored when compared to an image from a compared vendor. We then calculate each compared vendor's average Hamming distance value to determine the vendor with the minimum result as a Sybil account.

4.2.2 Main Category Feature. The main category of products is the second feature representing what types of products the vendor mainly sells. The distribution of product category of vendors stands for their identity, because sellers tend to highly specialize in some products in order to dominate the market [30]. However, there are several practical challenges to the achievement of automatic and accurate categorization. First, there are too many different names and types of categories depending on the marketplace in practice. Also, categories are sometimes incorrectly chosen since classes are typically self-selected by vendors. Moreover, there are even crawled image data without category labels. Hence, to create a more uniform embedding, we implemented a DNN classifier that was trained on data from SilkRoad2 and Agora, where ground-truth was available via labeling [30].

The DNN classifier was employed to extract features automatically without manually analyzing the feature list. The major challenge of using the DNN model is that it requires a massive amount of training data to be accurate. In the darknet market, the number of photographs per vendor is not enough to train the model, as

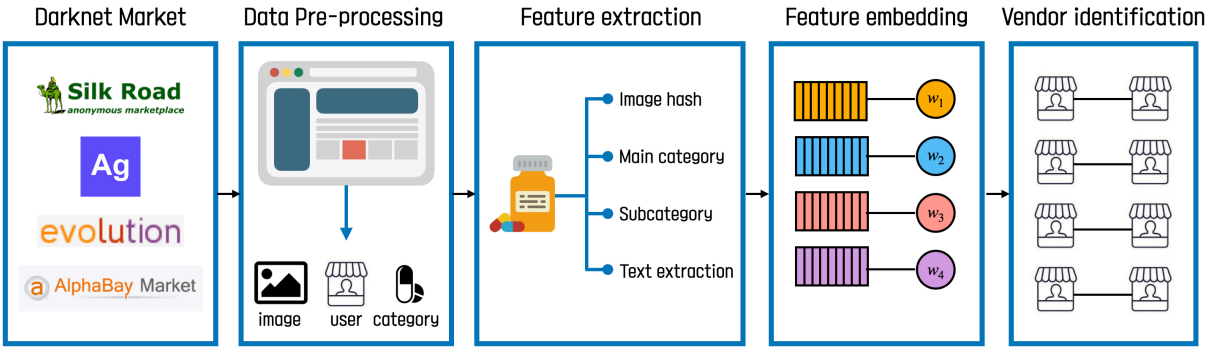


Figure 3: Overview of the proposed method. First, a data is pre-processed to extract meaningful information from the crawled dataset such as the ID, product image, product category, and PGP key of each vendor. Second, four main features such as image similarity, main category, subcategory, and text data from photos are extracted, and similarity scores are calculated between vendors. Third, all of these features are merged using a weighted feature embedding model. Finally, the connectivity of vendors with different accounts is identified using the final similarity scores, and candidate vendors are suggested as well.

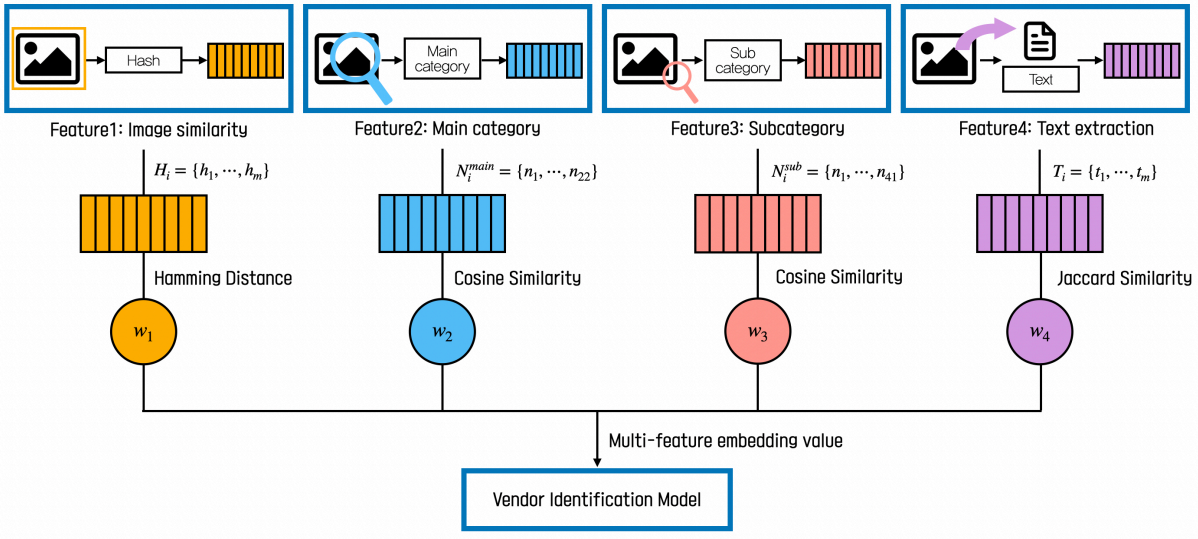


Figure 4: Feature extraction and embedding. After extracting features from a photograph, the similarity scores of each feature are used to calculate the final scores for the vendor by applying weighted feature embedding method which sets weight w_k for each feature's f_k, where f_k represents the similarity score of k^{th} feature.

shown in Table 1. Thus, we adopt transfer learning to pre-train a DNN using a large existing image dataset and fine-tune the last few layers with the darknet market dataset by leveraging the fact that the features of the DNN are more generic in the early layers and more data-precise in the later layers. In the fine-tuning procedure of the neural network, we used open tools such as Pytorch. Fig. 5 shows the model training of image categories.

To train the model, we use ImageNet (14 million images) [33], which is the largest annotated image dataset in the pre-training procedure. We then replace the final softmax layer with a new softmax layer using the darknet dataset, of which class is defined as a set of photographs uploaded by the same vendor. We then fine-tune all layers with back-propagation using the vendor's images by applying a stochastic gradient descent optimizer to minimize the cross-entropy loss function [60]. ResNet-50 [34] is one of the most common selections for generic image classification tasks, so we adopted the ImageNet pre-trained ResNet-50 model based on its outstanding performance in the results of the previous research [27]. We used datasets from SilkRoad2 and Agora, and unified similar categories into one main category by manually observing them. Based on our experiment, the test accuracy of DNN used for main category classification shows around 0.89. In the experiment, we randomly split the dataset with the label for train and test. As a result, as shown in Fig. 6, we defined 22 main categories (*e.g.*, drug, apparel, book, electronics, jewelry), then these were made into a

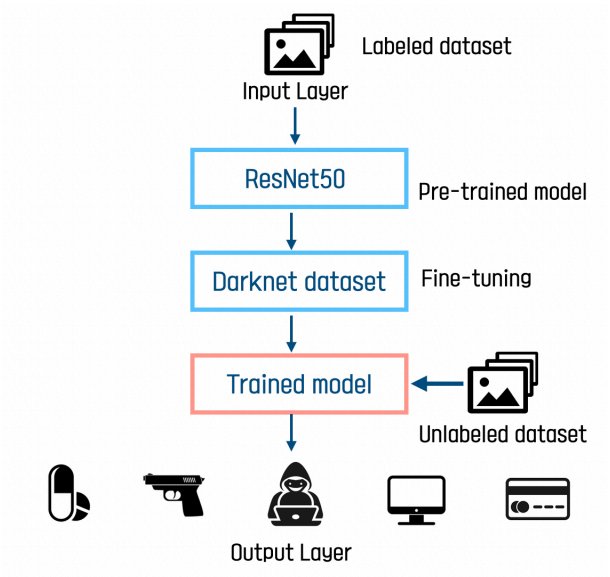


Figure 5: DNN model training of image category.

[Main category]

1	apparel
2	art
3	books
4	chemicals
5	computer
6	custom
7	data
8	digital
9	drug
10	electronics
11	erotica
12	forgeries
13	hardware
14	jewelry
15	lab
16	listings
17	lotteries
18	money
19	other
20	packaging
21	services
22	writing

[Subcategory]

1	barbiturates
2	benzos
3	cannabis
4	dissociatives
5	ecstasy
6	grinders
7	opioids
8	other
9	paper
10	paraphernalia
11	psychedelics
12	precursors
13	prescription
14	RCs
15	scales
16	smoked
17	stashes
18	steriodpeds
19	stimulants
20	weight loss

Figure 6: Definition of main and subcategory.

vector $M = \{m_1, \cdots, m_{22}\}$ for all of the categories. To create category embedding for vendor v_i, we produce a category number list $N_i = \{n_1, \cdots, n_{22}\}$, where each n_j represents the number of products from j^{th} category. We then compute the similarity of category distribution between the target vendor and the compared vendor using *Cosine Similarity*, which measures the similarity between two vectors of an inner product space by the cosine of the angle. The Cosine similarity value becomes closer to 1.0 as the match rate between vectors becomes higher [54]. Cosine similarity is suitable for comparing the category distribution between vendors because it captures the frequency of vector. In addition, Cosine similarity is demonstrated as an effective measure for the redundant dataset [45]. Therefore, it is adopted to measure the main category distribution considering the fact that category datasets can be inherently redundant. In the proposed method, a vendor with the highest Cosine similarity value to the target vendor is determined as a Sybil account for it.

4.2.3 Subcategory Feature. We take the subcategory of products as the third feature to capture the detailed characteristic of vendors' product distribution. In most darknet markets, subcategories are defined for some main categories, especially for drugs; while most of the other main categories do not have (see Fig. 6). Considering the fact that drugs take a significant portion in trading in the darknet market, detecting Sybil accounts only on the basis of the distribution of main categories may degrade the detection accuracy. Therefore, a more fine-grained subcategory information is used (if there is) to enhance performance. Specifically, the DNN model is used to classify the subcategory of darknet market images for uniform embedding, which is trained using the dataset of SilkRoad2 and Agora. The test accuracy of DNN used for subcategory classification shows around 0.90. For drugs, 20 subcategories information (*e.g.*, cannabis, ecstasy, paraphernalia, dissociatives) were involved in the model training; while the same name as the main category was used for the other categories that do not have their subcategories. These were then made into a list $S = \{s_1, \cdots, s_{41}\}$ for 21 main categories (except drug) and 20 subcategories for the drug. We made a category number vector for the whole list and calculated the Cosine similarity between the target vendor and the compared vendor to detect a Sybil account. The approach is precisely the same as that of the main category feature extraction (*c.f.* Section-4.2.2).

4.2.4 Text Feature. Darknet market vendors often put text into images, such as the names of drugs or sellers, as shown in Fig. 7. We extract the recognized text from the photos as the fourth representation of vendors. The critical challenge of applying the text extraction technique to darknet market photos is raised by the image quality, as the resolution of images is often quite low. After detecting the text area within the picture, in order to resolve the problem, we apply a super-resolution technique, which improves the image quality [35]. The text data is then recognized and extracted from the improved image.

When it comes to the implementation of text extraction, we adopted popular open source models for each step: TextFuseNet [36] for detecting text area, SwinIR [37] for applying super-resolution, and SATRN [38] for text extraction, each of which is the state-of-the-art method with high performance [55].

- **TextFuseNet** is a text detection method introduced by Ye et al. [36]. Our scheme adopts TextFuseNet model using Pytorch [56] for text detection. Unlike the previous approaches that only perceive texts based on limited feature representations, TextFuseNet makes use of richer features fused for text detection. The feature representations were perceived at

Figure 7: Text data within darknet market images.

three levels: character, word, and global, while still maintaining their general semantics. Following this, a text representation fusion technique was applied to help achieve robust arbitrary text detection. The model then used a multi-path fusion architecture to collect and merge the texts' features from different levels.

- **SwinIR** is an image restoration method introduced by Liang et al. [37]. We used the repository that contains the official PyTorch implementation of SwinIR [57] to restore the low resolution of darknet images. SwinIR applies transformers, which show improved results on high-level vision tasks compared to the previous methods based on convoluted neural networks. SwinIR consists of three parts: shallow feature extraction, deep feature extraction composed of several residual Swin Transformer blocks, and high-quality image reconstruction.
- **SATRN** is a text recognizing method of arbitrary shapes suggested by Lee et al. [38]. In our scheme, their official model [58] was adopted for text extraction. SATRN utilizes a self-attention mechanism to capture the dependency between word tokens in a sentence in order to describe 2D spatial dependencies of characters in a scene text image. It can recognize texts with arbitrary arrangements and large inter-character spacing by making use of the full-graph propagation of self-attention.

After extracting the text from images as the final step, a list $T_i = \{t_1, \cdots, t_m\}$ for all of the photos of the vendor v_i is created, where each t_j represents the text of image p_j for $1 \leq j \leq m$. The *Jaccard Similarity* [39] is then used to compare a vendor v_i's list T_i with that of the other vendors. The Jaccard similarity is a statistic used for measuring the similarity and diversity between two sets. If two groups share exactly the same components, their Jaccard similarity will be 1.0; conversely, if there are no members in common, it will be 0. We determine the compared vendor with the highest Jaccard similarity to be a Sybil account for the target vendor. Jaccard similarity takes only a unique set of words for each sentence, while Cosine similarity takes total vectors. In practice, many texts extracted from the darknet images (e.g., vendor ID) can be duplicated in a set T_i for each vendor v_i. Unlike Cosine similarity of which values are significantly affected by such redundancy in words, Jaccard similarity is not affected. Therefore, Jaccard similarity is adopted in the text extraction step.

4.3 Feature Embedding

We apply a multi-feature embedding technique [26] to combine each of the extracted features for each vendor. After extracting features from a photograph list P_i of a vendor v_i, the scores of each feature (i.e., Cosine and Jaccard similarities, Hamming distance) are used to calculate the final scores for the vendor. For the Cosine similarity and the Jaccard similarity, the larger value implies the higher similarity; however, for the Hamming distance, the smaller value implies the higher similarity. In order to match the scale and characteristic of different features, we slightly change the formula of the image similarity feature, which depends on Hamming distance value. The equation of image similarity score can be written as

$$\frac{16 - \text{Hamming distance}}{16}, \tag{1}$$

where 16 is the length of the image hash value.

Since not all the features are equally important, we apply weighted feature embedding method [26] by setting weight w_k (range of the weight is a positive number) for each feature's f_k, where f_k represents the similarity score of k^{th} feature. After combining the features, the weight of each feature is initialized with $1/k$ and trained using a stochastic gradient descent optimizer with an initial learning rate 0.01, aiming to minimize the cross-entropy loss function.

For obtaining the aggregated multi-feature representation, we used weighted feature embedding to fuse different features as follows:

- **Category-based embedding value** (*CB*) was created by fusing the main category embedding value (*MC*) and the subcategory embedding value (*SC*).
- **Product-based embedding value** (*PB*) was created by fusing the image similarity embedding value (*IS*) and *CB*.
- **Multi-feature embedding value** (*MF*), which is the final embedding value, was created by fusing *PB* and text embedding value (*TE*).

4.4 Vendor Identification

After embedding all of the features, our method determines if a given pair of vendors is the same individual by evaluating its score. Specifically, the most similar vendor to a target vendor can be found using the final score multiplied by the score for each feature and weight as follow:

$$(\sum_{k=1}^{4} w_k \times {f_k}^{\top})^{\top}, \tag{2}$$

where w_k is the weight of f_k and $\top$ is a transpose of a metric. The vendor with the highest score is evaluated as the same vendor as the target vendor.

One major limitation of the previous Sybil detection schemes in the darknet market is that it was impossible to extract multiple Sybil accounts at the same time with different possibilities, even if they had a high possibility of being so. This study employs a quantitative methodology to capture meaningful candidate vendors to overcome such a limitation faced by the previous studies. The final score is also used to find candidate vendors. We will demonstrate it in the next section.

5 EVALUATION RESULTS

5.1 Methodology

Due to the inherent anonymity of darknet market users, sufficient collection of the actual ground-truth of multiple accounts remains an open problem, which is an inherent difficulty that most of the studies in the literature have in common. In order to overcome the limitation and evaluate the efficacy of our method in the real-world setting, we adopt a data synthesis method for the ground-truth, which is a well-known solution to deal with the lack of ground-truth in the literature [26–28]. Specifically, we construct the ground-truth of Sybil accounts for training our model based on the original dataset from the SilkRoad2 and Agora in which the vendor ID and PGP key are already known. We then randomly split a given vendor's photos into two even parts, used the first set to train the model and linked the second set to the original vendor in the first set. (Later, we will apply our method for identifying the vendor with multiple accounts using the Evolution and Alphabay datasets to prove its versatility to various darknet markets in Section 6.)

Ground-truth Setting. We set only the vendors with the same PGP key as an identical vendor and split the dataset of given vendors into two pseudo pairs for constructing the dataset of Sybil pairs. For testing the feasibility of re-identifying vendors based on their features extracted by our method, we create two versions of the ground-truth dataset to show that our model does not just match the same photos in a naive way. Each product's picture was counted once, and we allowed different products to use the same image. For the duplication version, we consider all of the vendor's photos (potentially including duplicates). For the deduplicated version, we remove all of the duplicated photos that are used for different products by using their base64 values for the duplicate check.

The limitation of the ground-truth setting in our experiment is that several adversarial cases (e.g., adversarial vendors intentionally hiding multiple accounts or impersonating other vendors) were not considered for training, since we split the dataset of identical vendors to construct Sybil pairs to overcome the ground-truth problem. Such adversarial vendors may pose a real-world threat, making it difficult to link multiple Sybil accounts in practice. Hence, we will investigate the cases of adversarial vendors in Section 5.5, and discuss how to improve our method in practical applications in Section 7.

Evaluation Workflow Fig. 8 shows the evaluation workflow of our experiment. We introduce a threshold T_v to define the minimum number of photos for each vendor. For vendors with more than $2 \times T_v$ photos, we randomly split their photos into two even parts. We add the first set to the training dataset and the second set to the testing dataset. For the other vendors with more than T_v images (but less than $2 \times T_v$), we append them to training set as distractors. We set vendors in the testing group as target vendors and vendors in the training group as compared vendors. While testing the model, we capture the most similar training vendor for each testing vendor, and estimate whether the pairs of train and test vendor are correctly matched.

The limitation of threshold setting in our evaluation workflow is that vendors with lower photos than the threshold are consistently excluded. We do not contain low-activity vendors for practical reasons, but this leads to skewed results such that it can be only applied to high-activity vendors. Thus, we chose low threshold to include more vendors in our experiment in Section 5.4.

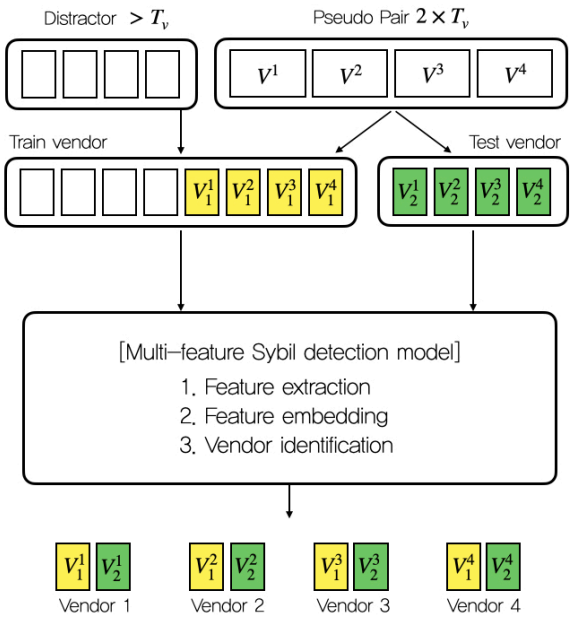


Figure 8: Workflow of the evaluation. The threshold T_v indicates the minimum number of photos for each vendor. For vendors with more than $2 \times T_v$ photos, their photos are randomly split into two even parts. The first and second sets are added to the training dataset and the testing dataset, respectively. For the other vendors with more than T_v images (but less than $2 \times T_v$), they are appended to the training set as distractors.

5.2 Feature Extraction

Table 3 displays the detailed results for the performance of each feature. It also shows the accuracy regarding duplicated and deduplicated datasets and the threshold of T_v. Considering the design choice of threshold T_v, a lower threshold allows us to examine more vendors, but there may not be enough training data for each vendor, vice versa.

As shown in Table 3, across different markets, the matching accuracy of image similarity shows the highest results among the features. Image similarity accuracy is 0.93 or higher when the duplicated images are not removed for different products. After deduplicating the images, the accuracy is still around 0.903-0.971 for SilkRoad2. Unlike SilkRoad2 which mainly deals with the drug category, the accuracy of Agora which contains a variety of products, is lower. The matching accuracy of the text feature had the second-best results, all of which indicate that text data from images can capture the characteristics of a vendor.

The results of category features are not significantly different for duplicated and deduplicated data; sometimes, the results from the deduplicated data are even better than those from the duplicated data. Since we compare the distribution of categories for each vendor using Cosine similarity and not just simply compare the total number of items in each category, similar results were obtained for deduplicated datasets. Also, when duplicated products were removed, exact product types from vendors were identified, so we could better capture each vendor's characteristics.

To evaluate the performance of each feature, we conduct additional analysis to empirically demonstrate the distribution of similarity scores predicted by our method. The density distribution of similarity score was drawn by randomly choosing pairwise comparison samples. Also, the number of pairwise comparison used to plot the Sybil and non-Sybil density functions is equal. Fig. 10 in Appendix shows the density distribution of similarity score for each feature. The more clearly the distribution of Sybil and non-Sybil pairs is separated, the higher the accuracy of each feature is, as can be seen in Table 3, because the features represent each vendor's characteristic. For instance, in the case of the image hash feature that showed the highest accuracy in Table 3, since the mean values of Sybil and non-Sybil are conspicuously different, Sybil accounts can be effectually identified. While the mean of the total distribution which was drawn by randomly choosing all pairwise samples (not considering any Sybil/non-Sybil label) is around 0.5, the mean values of Sybil and non-Sybil distribution are 0.8 and 0.2, respectively.

5.3 Feature Embedding

To train the embedding model's weight, we slightly modify our workflow of Fig. 8. Given a set of vendors having more than $2 \times T_v$ in the training dataset, we randomly put half of the data into the new training set for calculating weight, and the other half into the new testing set. For the other vendors that have more than T_v (but less than $2 \times T_v$), we add them to the training dataset as the distractors. We then train the weight of our feature embedding model and apply the trained weight to our original dataset built on Fig. 8.

Table 4 provides the performance of our model on various metrics using the data from SilkRoad2, Agora, and the two markets combined, when we set $T_v = 20$. Our model manages to outperform most of the other models and shows remarkable improvement on all datasets. A single category feature embedding achieves the lowest performance. Moreover, a single feature of image similarity and text performs better than the category-based feature embedding model.

Furthermore, what stands out in this table is the coverage of the model. Unlike the previous research that mainly dealt with only drugs, most of the other categories currently available in the darknet market are included, and the proposed model shows remarkable outcomes. In addition, it produces high accuracy even when deduplicated data was used. This means that the matching accuracy is high even when there is not enough training data for each vendor, and that our method properly captures the hidden characteristics of vendors, which can hardly be achievable by simply matching identical images.

Single Feature vs Multi-Feature. In order to evaluate the advantage of multi-feature embedding model, we compare the performance between our multi-feature method and the previous approach proposed by Wang et al. [27], which used photographs as a single feature. Specifically, we used a pre-trained ResNet-50 model in both our scheme and Wang et al.'s scheme as a baseline model, because it showed the best result in [27]. The ground-truth evaluation was conducted in the same way as in Fig. 8.

Table 5 shows the comparison results. As shown in the table, accuracy rate of our method is remarkably high especially in deduplicated datasets compared to the single feature case. The single feature model also achieved high accuracy when duplicated images were allowed, but it dropped when duplicated data were removed. This indicates that the single feature model's high accuracy seems to be the result of matching duplicated images, rather than actually capturing vendors' unique photography styles. Furthermore, single feature model is more limited in its coverage. In the comparison results, it can be seen that ResNet-50 model returns the highest accuracy for $T_v = 40$ after removing duplicated images. However, it is still not as accurate as our model with $T_v = 10$. Our model with a lower threshold ($T_v = 10$) outperforms the single feature model with a higher threshold ($T_v = 40$). It implies that our model covers up to 700% more vendors than the previous research and achieves higher matching accuracy even though there may not be enough training data for each vendor. The advantage is more meaningful when duplicated images are removed.

5.4 Vendor Identification

We matched the most similar training vendor for a given testing vendor in the above evaluation. Since not every vendor has a matched Sybil account in practice, we draw a minimal final score threshold T_1 for deciding a Sybil account. Our model will detect a match only if the similarity score between the target vendor and the compared vendor is higher than T_1. We also set the lowest score threshold T_2 for capturing a candidate vendor. Our model also considers the candidate vendor only when the score of the second-best compared vendor is above T_2. Even if we set two threshold scores in this experiment (i.e., for the best and the second-best Sybil accounts detection), it is important to note that our method is more generic such that it can be easily extended to detect more Sybil accounts if needed. Furthermore, our model can also be applied to other marketplaces (e.g., online marketplace, social media) to detect illicit users who manipulate reviews or upload malicious contents, as described in Section 2. When applied to other marketplaces, Sybil account and candidate account can be detected by finding the threshold in the same way.

The threshold T_1 and T_2 were determined considering the trade-off between true positives and false positives. To examine this trade-off, the workflow from Fig. 8 was slightly modified. Given a set of distractors [27], we randomly put half of them in the training set and the other half in the testing set. We generate the Receiver Operating Characteristic (ROC) curves by changing T_1 and T_2. Fig. 9 displays the ROC curve when $T_v = 20$ for distractors. In this paper, we use the elbow point of the ROC curve. The results confirm the

Dup	Market	T_v	Pair	Distractor	Image similarity Accuracy	Main category Accuracy	Subcategory Accuracy	Text Accuracy
Duplicated	SilkRoad2	10	577	220	0.976	0.113	0.329	0.808
		20	348	229	0.986	0.167	0.434	0.865
		40	150	198	1.00	0.207	0.633	0.913
	Agora	10	1031	615	0.947	0.074	0.215	0.683
		20	491	540	0.945	0.126	0.322	0.754
		40	169	322	0.935	0.225	0.609	0.864
Deduplicated	SilkRoad2	10	401	251	0.903	0.122	0.262	0.733
		20	200	201	0.96	0.17	0.39	0.82
		40	69	131	0.971	0.319	0.623	0.87
	Agora	10	433	512	0.663	0.095	0.219	0.506
		20	158	275	0.684	0.177	0.386	0.57
		40	49	109	0.653	0.408	0.612	0.51

Table 3: Vendor matching accuracy of each feature extraction. The accuracy of each feature is evaluated in both duplicated and deduplicated datasets of ground-truth. A lower threshold T_v allows us to examine more vendors, but there may not be enough training data for each vendor, vice versa.

Model	SilkRoad2		Agora		SilkRoad2 & Agora	
	Accuracy	F-Score	Accuracy	F-Score	Accuracy	F-Score
Main category (MC)	0.17	0.140	0.177	0.142	0.212	0.183
Subcategory (SC)	0.39	0.291	0.386	0.283	0.427	0.321
Category-based ($MC + SC$)	0.395	0.293	0.38	0.279	0.433	0.329
Text (TE)	0.82	0.772	0.57	0.469	0.623	0.536
Image similarity (IS)	0.96	0.949	0.684	0.657	0.768	0.724
Product-based ($MC + SC + IS$)	0.975	0.965	0.835	0.796	0.88	0.84
Final model ($MC + SC + IS + TE$)	0.985	0.978	0.861	0.823	0.908	0.872

Table 4: Performance metric of our model ($T_v = 20$ with deduplicated images).

Dup	Market	T_v	Pair	Distractor	Single feature model (Baseline)[27]	Multi-feature model (Ours)
Duplicated	SilkRoad2	10	577	220	0.981	0.988
		20	348	229	0.983	0.991
		40	150	198	0.993	1
	Agora	10	1031	615	0.962	0.97
		20	491	540	0.965	0.971
		40	169	322	0.963	0.959
	SilkRoad2 & Agora	10	1608	835	0.964	0.967
		20	839	769	0.962	0.976
		40	520	319	0.97	0.972
Deduplicated	SilkRoad2	10	401	251	0.88	0.96
		20	200	201	0.955	0.985
		40	69	131	0.957	0.971
	Agora	10	433	512	0.815	0.831
		20	158	275	0.879	0.861
		40	49	109	0.853	0.857
	SilkRoad2 & Agora	10	834	763	0.853	0.867
		20	358	476	0.877	0.908
		40	118	240	0.856	0.924

Table 5: Vendor matching accuracy in the single feature model and our multi-feature model. In order to evaluate the advantage of multi-feature embedding model, the performance of the proposed method is compared to that of Wang et al.'s method [27], which used photographs as a single feature. In the evaluation, a pre-trained ResNet-50 model is used in both methods.

areas under the curves (AUC) are close to 1.0[2], as curves reach the top-left corner of the plots for T_1. Since we could detect most of the Sybil accounts for distractors from T_1, there were not enough cases for T_2 to find a candidate vendor. Although the graph of T_2 may not look statistically significant, the results are suitable for our purposes of finding T_2 to choose a minimum value for a candidate vendor.

[2]A random classifier's AUC would be 0.5, and a higher AUC is better.

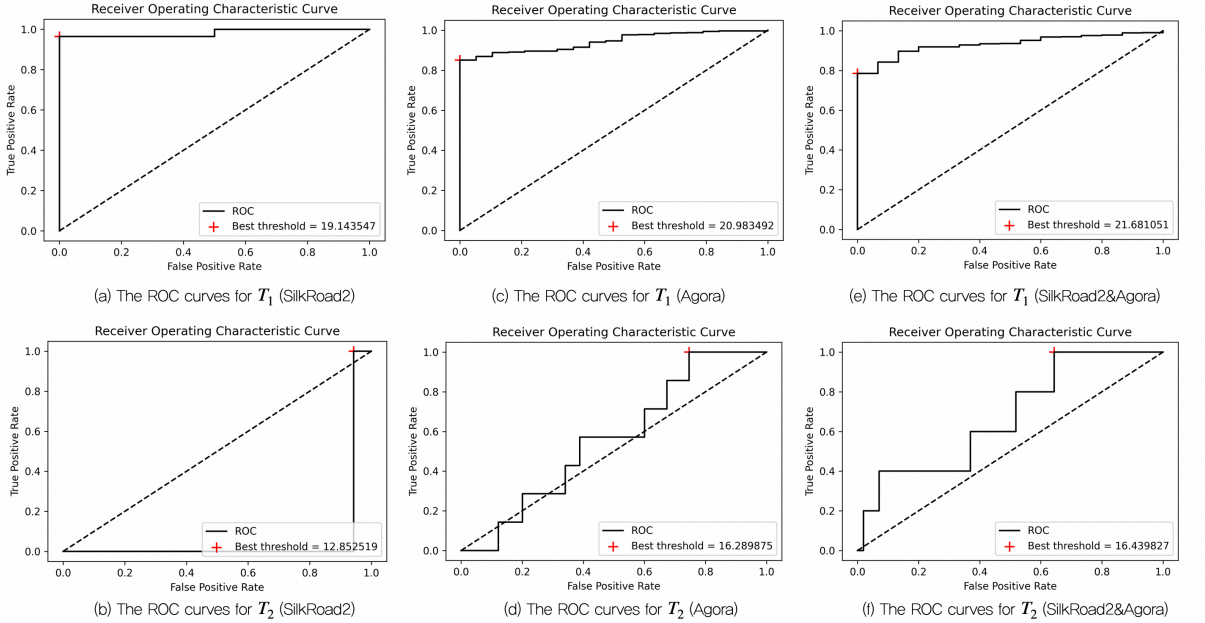


Figure 9: The ROC curves for T_1 and T_2 from distractors ($T_v = 20$, with duplicated images).

Effect of Candidate Vendor. In order to evaluate the effect of candidate vendors to the accuracy of Sybil detection, we measured if candidate vendors contain the correct match meaningfully. We compared the results with and without a candidate vendor. Table 6 displays the experimental results. As shown in the table, the F-score becomes higher when the candidate vendor is considered compared to the case without it in all of the markets. We set the threshold $T_v = 10$ in this table to include more vendors to validate the coverage of our model.

Detection of Multiple Accounts. To examine if our model can detect more than one Sybil account accurately, we slightly adjust the workflow of Fig. 8. For the vendors having more than $3 \times T_v$, we split their photos into three parts as pseudo vendors. We add the first and second parts to the training set and the third to the testing set. For the other vendors having more than T_v (but less than $3 \times T_v$), we add them to the training set. Therefore, each vendor in the testing dataset has two multiple accounts in the training dataset. In this way, for each testing vendor, we identify whether all multiple accounts could be found while also considering a candidate vendor. We set $T_v = 10$ in our simulation to include more vendors.

Table 7 presents the number of multiple accounts that our method detected while including candidate vendors. As shown in the table, our model found almost all Sybil accounts for each pseudo vendor pair, taking a candidate vendor into consideration. In the case of SilkRoad2, we found all of the multiple accounts in 253 pairs (91%) out of 277 pairs with a high precision of about 95%. We also found 134 pairs (55%) out of 242 pairs with a precision of about 77% in the dataset from Agora. However, the limitation of multiple account setting in our experiment is that fewer vendors were represented in the testing set since we included the vendors having more than $3 \times T_v$ photos.

Market	Candidate vendor	Precision	Recall	F-score	Accuracy
SilkRoad2	-	0.976	0.969	0.973	0.948
	O	0.953	1.0	0.976	0.953
Agora	-	0.847	0.956	0.898	0.82
	O	0.853	1.0	0.921	0.855
SilkRoad2 & Agora	-	0.872	0.972	0.92	0.853
	O	0.897	1.0	0.946	0.897

Table 6: Performance of vendor matching with/without candidate vendor ($T_v = 10$, with deduplicated images). 'O' indicates results with candidate vendor; and '-' indicates results without candidate vendor.

Market	Pair	1 Sybil accounts		2 Sybil accounts	
		Match	Precision	Match	Precision
SilkRoad2	277	275 (99%)	0.993	253 (91%)	0.951
Agora	242	214 (88%)	0.911	134 (55%)	0.77
SilkRoad2 & Agora	519	470 (91%)	0.94	307 (59%)	0.841

Table 7: Performance of detecting multiple accounts ($T_v = 10$, with deduplicated images).

5.5 Case Study

We conduct case studies to examine the usage pattern of Sybil accounts of vendors based on the results we observed for deep understanding of the darknet market ecosystem. The case studies

also include an interesting case that achieved a high similarity assessment in our model but was actually not the same vendor in ground-truth. We set $T_v = 20$ and examined the number of accounts indicated in Table 3. Then, we found that the percentage of cases showing a higher similarity than the threshold $T_1 = 20$ but not of the same vendor is less than 1% from within the same market and across different markets. Among them, we manually investigated accounts with high similarity scores (specifically, we analyzed the 10 pairs with the highest similarity score in each market). Table 8 shows the number of candidate Sybil accounts whose similarity score is higher than T_1.

Market	Account	Pair	Candidate Sybil
SilkRoad2	348	121,104	763
Agora	491	241,081	16
SilkRoad2 & Agora	839	703,921	77

Table 8: Summary of dataset for case study ($T_v = 20$, with duplicated images). 'Account' indicates the number of accounts examined; 'Pair' was calculated by comparing each account; 'Candidate Sybil' includes the pairs whose similarity score is higher than T_1.

Sybils within the Same Market. Fig. 11 shows an example of Sybil account ('salt-pepper' and 'salt-pepperSa') from SilkRoad2. They do not use identical photos, but share the same logo on their product images which was captured by text feature extraction in our method. Additionally, our model identified that they sell the same categories of products, that is ecstasy, cannabis, stimulants, and dissociatives. Another example, showing the same pattern as the Sybil pair of Silkroad2, is 'Hedera' and 'HederalExpress' from Agora.

Sybils across Different Markets. Fig. 12 shows two vendors using similar vendor names ('drugsforyou' and 'Drugs4you') but operating on different markets, that is SilkRoad2 and Agora, respectively. Both vendors do not use the exactly same photos but put the same logo, which is 'Drugs4you', into their products, and only sell products in the drug category, especially cannabis. Additionally, 'repaaa' from SilkRoad2 and 'RepAAA' from Agora post similar images, use the same logo, which is 'REPLICAAAA', and sell products only in the apparel category. Each pair is thus identified as Sybil account, demonstrating our model can detect additional Sybil pairs across different markets.

Adversarial Vendors. Some vendors operating Sybil accounts may not want these accounts to be identified for several security or economical reasons. Fig. 13 shows such a case that our method found. Specifically, our model detected Sybil pairs from Agora, that is 'RXChemist' and 'Remedyplus'. They seek to intentionally hide the vendor relationship by sharing few identical photos, using different PGP key and vendor ID, and posting vendor descriptions with dissimilar structure. However, our model captured that they sell the same category of products such as prescription, benzos, and opioids, and upload images of highly similar items. In order to demonstrate the correctness of the result, we conducted a manual analysis of it and confirmed that they were actually of the same vendor, because they sell 214 items with the same name, supply products from the same regions (e.g., US, UK, and Australia), and include a few identical contents in the introduction.

On the other hand, we also verified there are impersonators copying product images from other vendors. For example, in SilkRoad2, 'kimbe' and 'drugstore' were evaluated as the same vendor by our method, since they use identical photos of drugs with slightly different sizes (specifically, 'kimbe' resizes the image of 'drugstore'). After our careful and manual investigation on the case, however, we found that they are not likely to be the Sybil pairs, because 'drugstore' only sells drugs shipping from Croatia, but 'kimbe' supplies various types of products shipping from Vatican. Such impersonating cases may result in false positives in our method. How to handle the impersonating case will be discussed in Section 7.

6 CROSS MARKET ANALYSIS

We apply our method to the other darknet market datasets which were not used for our model training to demonstrate the extensibility of the proposed method. We selected datasets from Evolution and Alphabay for the cross market analysis, since a large number of photos across various categories were included in their datasets. Following the evaluation process in Fig. 8, we randomly split the dataset into two parts for vendors with more than $2 \times T_v$, adding the first half to the set of target vendor and the second to the set of compared vendor. For the other vendors having more than T_v images (but less than $2 \times T_v$), we include them in the set of compared vendors as distractors. We then apply our model and set the thresholds $T_1 = 20$ and $T_2 = 16$, as was trained by both the SilkRoad2 and Agora datasets. Table 9 presents the results of Sybil detection from those darknet markets on various metrics when applying our trained system. In this case, we used all of the photos of vendors without deliberately removing duplicated images. Also, we set $T_v = 10$ to include more vendors to evaluate the coverage of our method. From the table, it can be seen that our model performs well on various darknet markets. Furthermore, the performance of our scheme is improved when considering candidate vendors.

Market	Candidate vendor	Precision	Recall	F-score	Accuracy
Evolution	-	0.939	1.0	0.969	0.94
	O	0.945	1.0	0.972	0.945
Alphabay	-	0.91	1.0	0.953	0.901
	O	0.928	1.0	0.962	0.928
Evolution&Alphabay	-	0.921	0.999	0.959	0.921
	O	0.932	1.0	0.965	0.932

Table 9: Performance of cross market Sybil detection ($T_v = 10$, with duplicated images). We apply our method to Evolution and Alphabay datasets which were not used for training our model to evaluate the extensibility of the proposed method.

7 DISCUSSION

Extension of Multi-account Detection. Our method can detect more than one Sybil account simultaneously according to the similarity score. In our experiment, we set two threshold values (T_1 and

T_2) to find the best and the second-best Sybil accounts and demonstrated their feasibility (*c.f.* Section-5.4). However, it is important to note that our method is more generic such that it can be easily extended to detect more Sybil accounts as the applied system requires. For example, we can also set threshold values T_n for capturing n^{th} candidate vendors, for $n > 2$. Our model then considers the candidate vendor only when the score of the n^{th}-best compared vendor is above T_n. Because there should be an inevitable trade-off between false positives and true positives caused by choice, however, finding the optimal number of candidates would be an important open problem in practice. In contrast to the case of adversarial vendors, it is reasonable to assume that some vendors seek to leverage positive brand association across marketplaces by the inclusion of textual cues in product photographs. Such cases would provide a strong signal to our proposed methodology. Potential examples are discussed in the our highlighted case studies (Section 5.5). As discussed, such cases are difficult to decouple from copy-cat behaviour between vendors. As such performance evaluation remains a complicated and open problem and intentionality on behalf of (Sybil) account operators should not be inferred when considering the association of multiple accounts.

Design Choice of Model. We extract four features, that is image similarity, main category, subcategory, and text data from pictures. However, it is important to note that our model can flexibly embed additional features into the vendor identification model. The DNN feature was also considered as a useful feature in our model, because the previous study in [27] showed that it could capture the unique photography style. According to our experiment, however, we observed that the similarity using a DNN feature becomes 'overconfident' to the most similar vendor with a limited accuracy, which could not be improved even if the other features were combined. Therefore, we determined DNN feature is inappropriate for the multi-feature model, and demonstrated our design choice actually outperforms it as shown in Section 5.3.

Adversarial Vendors. One of the challenging issues in the literature is how to handle adversarial vendors. Some vendors with Sybil accounts in the darknet market actively reveal their brands by using logos or identical product photos, which can be easily captured by our model. However, adversarial vendors trying to hide their Sybil accounts or mimic other accounts make Sybil detection difficult. There may be many real-world cases such as accounts belonging to the same vendor but having different vendor ID and PGP keys for hiding their multiple accounts, or impersonators copying the other vendors' product images. For the former case, our model can effectively detect the hidden Sybil pairs; but, for the latter case, our model may wrongly detect them, incurring false positives, as shown in case studies (*c.f.* Section-5.5). However, we found that even if the impersonators may be able to easily mimic images, they can hardly mimic the categories of all items that are being sold in practice, since it is costly to manipulate. One of the feasible solutions to mitigate this problem is to include more fine-grained category features in our method. It can be achieved by collecting more datasets covering various categories of photos from the other darknet markets, and refining the classification of categories based on them. In addition, the robustness of our method would be improved if the categories of images can be identified more accurately. For example, we can apply a self-supervised representation learning method as an image classification model in our method. Because it does not need labeled dataset for training, it would help to enhance the accuracy of image category identification in the darknet market environment of which datasets are mostly unlabeled. Additionally, our method can also be applied to online marketplaces or social media to detect adversarial users who intentionally hide their accounts or impersonate other accounts for malicious purposes. However, there is also a possibility of misusing Sybil detection approach for de-anonymizing legitimate individuals' multiple accounts. Legitimate individuals may preserve their privacy by considering the result of our research.

8 CONCLUSION

We propose a novel approach for Sybil accounts detection in the darknet markets. It leverages multi-feature extraction from photographs of darknet vendors to fingerprint their Sybil accounts within and across markets. Specifically, our method extracts various features such as image similarity, main category, subcategory, and text data to generate multi-feature embedding at a vendor level. Also, the proposed scheme can find multiple accounts at the same time, which is the first in the Sybil detection literature. For evaluating our model, we construct the ground-truth of Sybil accounts by randomly splitting the dataset of given vendors into two pseudo pairs and including vendors with photos more than threshold. Even if we chose low threshold to include more vendors in our experiment, the results could be skewed to apply to high-activity vendors. We have shown that our model achieves an accuracy of 98%, and its performance significantly outweighs the previous photography-based study leveraging only a single feature. Additionally, our method covers up to 700% more vendors. Furthermore, due to its fine-grained multiple-feature extraction from photographs, our method can be more generically applied to a variety of darknet markets regardless of their languages and item types. Sybil detection in anonymous darknet markets gives a better understanding of the ecosystem by accurately identifying vendors and actual transactions of items among them. In future work, we will investigate how to resolve the adversarial vendor problem in an efficient and accurate manner based on the insight discovered in this study.

ACKNOWLEDGMENTS

This work was supported by IITIP grant funded by the MSIT, Korea (No.2022-0-00411, IITP-2022-2021-0-01810, IITP-2022-2020-0-01819), and Basic Science Research Program through the NRF funded by the Ministry of Education, Korea (NRF-2021R1A6A1A13044830).

REFERENCES

[1] Roger Dingledine, Nick Mathewson, and Paul Syverson. 2004. Tor: The Second-Generation Onion Router. (2004). DOI:https://doi.org/10.21236/ada465464
[2] Hengrui Zhang and Futai Zou. 2020. A Survey of the Dark Web and Dark Market Research. 2020 IEEE 6th International Conference on Computer and Communications (ICCC) (2020). DOI:https://doi.org/10.1109/iccc51575.2020.9345271
[3] Mhd Wesam Al-Nabki, Eduardo Fidalgo, Enrique Alegre, and Laura Fernández-Robles. 2019. ToRank: Identifying the most influential suspicious domains

in the Tor network. Expert Systems with Applications 123, (2019), 212-226. DOI:https://doi.org/10.1016/j.eswa.2019.01.029
[4] Spongebob Squarepants. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. SSRN Electronic Journal (2008). DOI:https://doi.org/10.2139/ssrn.3977007
[5] Wood, Gavin. "Ethereum: A secure decentralised generalised transaction ledger." Ethereum project yellow paper 151.2014 (2014): 1-32.
[6] A. Biryukov, I. Pustogarov, and R. Weinmann. 2013. Trawling for Tor Hidden Services: Detection, Measurement, Deanonymization. 2013 IEEE Symposium on Security and Privacy (2013). DOI:https://doi.org/10.1109/sp.2013.15
[7] Onur Catakoglu, Marco Balduzzi, and Davide Balzarotti. 2017. Attacks landscape in the dark side of the web. Proceedings of the Symposium on Applied Computing (2017). DOI:https://doi.org/10.1145/3019612.3019796
[8] Diana S. Dolliver. 2015. Evaluating drug trafficking on the Tor Network: Silk Road 2, the sequel. International Journal of Drug Policy 26, 11 (2015), 1113-1123. DOI:https://doi.org/10.1016/j.drugpo.2015.01.008
[9] Sean Foley, Jonathan R. Karlsen, and TTlis J. Putniii. 2018. Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed Through Cryptocurrencies? SSRN Electronic Journal (2018). DOI:https://doi.org/10.2139/ssrn.3102645
[10] Husam Al Jawaheri, Mashael Al Sabah, Yazan Boshmaf, and Aiman Erbad. 2020. Deanonymizing Tor hidden service users through Bitcoin transactions analysis. Computers & Security 89, (2020), 101684. DOI:https://doi.org/10.1016/j.cose.2019.101684
[11] Rebecca S. Portnoff, Danny Yuxing Huang, Periwinkle Doerfler, Sadia Afroz, and Damon McCoy. 2017. Backpage and Bitcoin. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2017). DOI:https://doi.org/10.1145/3097983.3098082
[12] Marie Vasek and Tyler Moore. 2019. Analyzing the Bitcoin Ponzi Scheme Ecosystem. Financial Cryptography and Data Security (2019), 101-112. DOI:https://doi.org/10.1007/978-3-662-58820-8_8
[13] Massimo Bartoletti, Barbara Pes, and Sergio Serusi. 2018. Data Mining for Detecting Bitcoin Ponzi Schemes. 2018 Crypto Valley Conference on Blockchain Technology (CVCBT) (2018). DOI:https://doi.org/10.1109/cvcbt.2018.00014
[14] Malte Moser, Rainer Bohme, and Dominic Breuker. 2013. An inquiry into money laundering tools in the Bitcoin ecosystem. 2013 APWG eCrime Researchers Summit (2013). DOI:https://doi.org/10.1109/ecrs.2013.6805780
[15] Thibault De Balthasar and Julio Hernandez-Castro. 2017. An Analysis of Bitcoin Laundry Services. Secure IT Systems (2017), 297-312. DOI:https://doi.org/10.1007/978-3-319-70290-2_18
[16] 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. Proceedings 2019 Network and Distributed System Security Symposium (2019). DOI:https://doi.org/10.14722/ndss.2019.23055
[17] Michele Campobasso, Pavlo Burda, and Luca Allodi. 2019. CARONTE: Crawling Adversarial Resources Over Non-Trusted, High-Profile Environments. 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) (2019). DOI:https://doi.org/10.1109/eurospw.2019.00055
[18] Shalini Ghosh, Ariyam Das, Phil Porras, Vinod Yegneswaran, and Ashish Gehani. 2017. Automated Categorization of Onion Sites for Analyzing the Darkweb Ecosystem. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2017). DOI:https://doi.org/10.1145/3097983.3098193
[19] Yiming Zhang, Yujie Fan, Yanfang Ye, Liang Zhao, and Chuan Shi. 2019. Key Player Identification in Underground Forums over Attributed Heterogeneous Information Network Embedding Framework. Proceedings of the 28th ACM International Conference on Information and Knowledge Management (2019). DOI:https://doi.org/10.1145/3357384.3357876
[20] Jack Hughes, Ben Collier, and Alice Hutchings. 2019. From playing games to committing crimes: A multi-technique approach to predicting key actors on an online gaming forum. 2019 APWG Symposium on Electronic Crime Research (eCrime) (2019). DOI:https://doi.org/10.1109/ecrime47957.2019.9037586
[21] Shin-Ying Huang and Tao Ban. 2019. A Topic-Based Unsupervised Learning Approach for Online Underground Market Exploration. 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (2019). DOI:https://doi.org/10.1109/trustcom/bigdatase.2019.00036
[22] Zhibo Sun, Carlos E. Rubio-Medrano, Ziming Zhao, Tiffany Bao, Adam Doupé, and Gail-Joon Ahn. 2019. Understanding and Predicting Private Interactions in Underground Forums. Proceedings of the Ninth ACM Conference on Data and Application Security and Privacy (2019). DOI:https://doi.org/10.1145/3292006.3300036
[23] Maria Bada and Ildiko Pete. 2020. An exploration of the cybercrime ecosystem around Shodan. 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS) (2020). DOI:https://doi.org/10.1109/iotsms52051.2020.9340224
[24] Prima Chairunnanda, Nam Pham, and Urs Hengartner. 2011. Privacy: Gone with the Typing! Identifying Web Users by Their Typing Patterns. 2011 IEEE Third Int (2011). DOI:https://doi.org/10.1109/passat/socialcom.2011.197
[25] Thanh Nghia Ho and Wee Keong Ng. 2016. Application of Stylometry to DarkWeb Forum User Identification. Information and Communications Security (2016), 173-183. DOI:https://doi.org/10.1007/978-3-319-50011-9_14
[26] Ramnath Kumar, Shweta Yadav, Raminta Daniulaityte, Francois Lamy, Krishnaprasad Thirunarayan, Usha Lokala, and Amit Sheth. 2020. eDarkFind: Unsupervised Multi-view Learning for Sybil Account Detection. Proceedings of The Web Conference 2020 (2020). DOI:https://doi.org/10.1145/3366423.3380263
[27] Xiangwen Wang, Peng Peng, Chun Wang, and Gang Wang. 2018. You Are Your Photographs. Proceedings of the 2018 on Asia Conference on Computer and Communications Security (2018). DOI:https://doi.org/10.1145/3196494.3196529
[28] Yiming Zhang, Qi Xiong, Yujie Fan, Wei Song, Shifu Hou, Yanfang Ye, Xin Li, Liang Zhao, Chuan Shi, and Jiabin Wang. 2019. Your Style Your Identity: Leveraging Writing and Photography Styles for Drug Trafficker Identification in Darknet Markets over Attributed Heterogeneous Information Network. The World Wide Web Conference on - WWW (2019). DOI:https://doi.org/10.1145/3308558.3313537
[29] Nicolas Christin. 2012. Traveling the Silk Road: A Measurement of a Large Anonymous Online Marketplace. (2012). DOI:https://doi.org/10.21236/ada579383
[30] Kyle Soska Carnegie Mellon University et al. 2015. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem: Proceedings of the 24th USENIX Conference on Security Symposium. (August 2015). Retrieved May 18, 2022 from https://dl.acm.org/doi/10.5555/2831143.2831146
[31] Susan Jeziorowski, Muhammad Ismail, and Ambareen Siraj. 2020. Towards Image-Based Dark Vendor Profiling. Proceedings of the Sixth International Workshop on Security and Privacy Analytics (2020). DOI:https://doi.org/10.1145/3375708.3380311
[32] Issam H. Laradji, Lahouari Ghouti, and El-Hebri Khiari. 2013. Perceptual hashing of color images using hypercomplex representations. 2013 IEEE International Conference on Image Processing (2013). DOI:https://doi.org/10.1109/icip.2013.6738907
[33] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision 115, 3 (2015), 211-252. DOI:https://doi.org/10.1007/s11263-015-0816-y
[34] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016). DOI:https://doi.org/10.1109/cvpr.2016.90
[35] Pablo Blanco-Medina, Eduardo Fidalgo, Enrique Alegre, Rocío Alaiz-Rodríguez, Francisco Jáñez-Martino, and Alexandra Bonnici. 2020. Rectification and Super-Resolution Enhancements for Forensic Text Recognition. Sensors 20, 20 (2020), 5850. DOI:https://doi.org/10.3390/s20205850
[36] Jian Ye, Zhe Chen, Juhua Liu, and Bo Du. 2020. TextFuseNet: Scene Text Detection with Richer Fused Features. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (2020). DOI:https://doi.org/10.24963/ijcai.2020/72
[37] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. SwinIR: Image Restoration Using Swin Transformer. 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) (2021). DOI:https://doi.org/10.1109/iccvw54120.2021.00210
[38] Junyeop Lee, Sungrae Park, Jeonghun Baek, Seong Joon Oh, Seonghyeon Kim, and Hwalsuk Lee. 2020. On Recognizing Texts of Arbitrary Shapes with 2D Self-Attention. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2020). DOI:https://doi.org/10.1109/cvprw50498.2020.00281
[39] Paul Jaccard. 1912. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1. New Phytologist 11, 2 (1912), 37-50. DOI:https://doi.org/10.1111/j.1469-8137.1912.tb05611.x
[40] Xiao Hui Tai, Kyle Soska, and Nicolas Christin. 2019. Adversarial Matching of Dark Net Market Vendor Accounts. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2019). DOI:https://doi.org/10.1145/3292500.3330763
[41] Alex Biryukov, Ivan Pustogarov, Fabrice Thill, and Ralf-Philipp Weinmann. 2014. Content and Popularity Analysis of Tor Hidden Services. 2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (2014). DOI:https://doi.org/10.1109/icdcsw.2014.20
[42] DITTRICH, David; BAILEY, Michael; DIETRICH, Sven. Towards community standards for ethical behavior in computer security research. University of Washington, Michael Bailey, University of Michigan, Sven Dietrich, Stevens Institute of Technology, Stevens CS Technical Report 20091, 2009.
[43] Vickie A. Miracle. 2016. The Belmont Report. Dimensions of Critical Care Nursing 35, 4 (2016), 223-228. DOI:https://doi.org/10.1097/dcc.0000000000000186
[44] James Martin and Nicolas Christin. 2016. Ethics in cryptomarket research. International Journal of Drug Policy 35, (2016), 84-91. DOI:https://doi.org/10.1016/j.drugpo.2016.05.006
[45] Sindhuja, B., and Veena Trivedi. "Usage of cosine similarity and term frequency count for textual document clustering." International Journal of Innovative Research in Computer Science & Technology (IJIRCST) 2.5 (2014): 9-12.

[46] Al-Qurishi, M., Al-Rakhami, M., Alamri, A., Alrubaian, M., Rahman, S. and Hossain, M., 2017. Sybil Defense Techniques in Online Social Networks: A Survey. IEEE Access, 5, pp.1200-1219.
[47] James cook. "FBI Arrests Former SpaceX Employee, Alleging He Ran The 'Deep Web' Drug Marketplace Silk Road 2.0." Insider. https://www.businessinsider.com/fbi-silk-road-seized-arrests-2014-11.
[48] Andy Greenberg. "Agora, the Dark Web's Biggest Drug Market, Is Going Offline". Wired. https://www.wired.com/2015/08/agora-dark-webs-biggest-drug-market-going-offline/.
[49] Nicky Woolf. "Bitcoin 'exit scam': deep-web market operators disappear with $12m." The Guardian. https://www.theguardian.com/technology/2015/mar/18/bitcoin-deep-web-evolution-exit-scam-12-million-dollars.
[50] The Big List of Darknet Markets 2021 - Best Darknet Markets - DNStats. Retrieved May 18, 2022 from https://dnstats.net/list-of-darknet-markets/
[51] Gwern Branwen, Nicolas Christin, David Décary-Hétu, Rasmus Munksgaard Andersen, StExo, El Presidente, Anonymous, Daryl Lau, Sohhlz, Delyan Kratunov, Vince Cakic, Van Buskirk, Whom, Michael McKenna, Sigi Goode. "Dark Net Market archives, 2011–2015", 2021-07-12. Web. [access date] /DNM-archives
[52] ImageHash · PyPI. Retrieved May 18, 2022 from https://pypi.org/project/ImageHash/
[53] SSIM: Structural Similarity Index | Imatest. Retrieved May 18, 2022 from https://www.imatest.com/docs/ssim/
[54] Cosine Similarity - an overview | ScienceDirect Topics. Retrieved May 18, 2022 from https://www.sciencedirect.com/topics/computer-science/cosine-similarity
[55] Papers With Code: The latest in Machine Learning. Retrieved May 18, 2022 from https://paperswithcode.com/
[56] ying09/TextFuseNet: A PyTorch implementation of . - GitHub. Retrieved May 18, 2022 from https://github.com/ying09/TextFuseNet
[57] JingyunLiang/SwinIR: SwinIR: Image Restoration Using . - GitHub. Retrieved May 18, 2022 from https://github.com/JingyunLiang/SwinIR
[58] clovaai/SATRN: Official Tensorflow Implementation of . - GitHub. Retrieved May 18, 2022 from https://github.com/clovaai/SATRN
[59] Below the Surface: Exploring the Deep Web. Retrieved May 18, 2022 from https://documents.trendmicro.com/assets/wp/wp_below_the_surface.pdf
[60] CrossEntropyLoss — PyTorch 1.11.0 documentation. Retrieved May 18, 2022 from https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html
[61] 1.11. Ensemble methods — scikit-learn 1.1.0 documentation. Retrieved May 18, 2022 from http://scikit-learn.org/stable/modules/ensemble.html
[62] World Market | DarknetStats. Retrieved May 18, 2022 from https://www.darknetstats.com/world-market/
[63] Creative Commons. 2022. CC0 - Creative Commons. [online] Available at: <https://creativecommons.org/share-your-work/public-domain/cc0/> [Accessed 20 May 2022].

A SIMILARITY SCORE DENSITY DISTRIBUTION

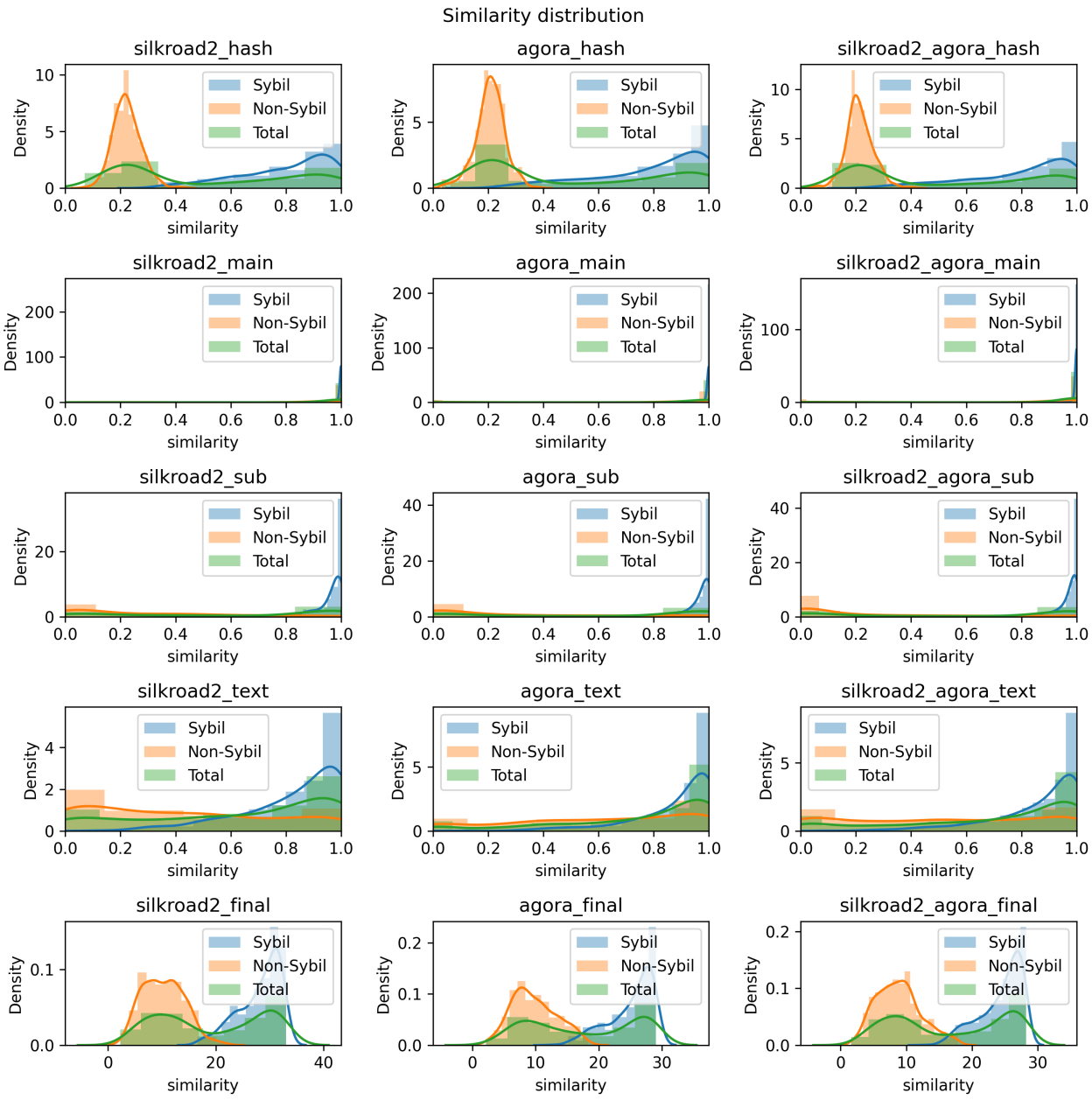


Figure 10: The density distribution of similarity score. It was drawn by randomly choosing pairwise comparison samples. The number of pairwise comparison used to plot the Sybil and non-Sybil density functions is equal. 'Total' density distribution of similarity score was drawn by randomly choosing all pairwise samples not considering any Sybil/non-Sybil label. We used the same dataset in Table 3.

B CASE STUDY

Sybil pair		Vendor 1	Vendor 2
Case1	Market	SilkRoad2	SilkRoad2
	Username	salt–pepper	salt–pepperSa
	Image		
	Category	drug – ecstasy, cannabis, stimulants, dissociatives	drug – ecstasy, cannabis, stimulants, dissociatives
	Text	Saltnpepper	Saltnpepper

Sybil pair		Vendor 1	Vendor 2
Case2	Market	Agora	Agora
	Username	Hedera	HederalExpress
	Image		
	Category	drug – psychedelics, stimulants, dissociative, ecstasy	drug – psychedelics, stimulants, dissociative, ecstasy
	Text	HedEx	HedEx

Figure 11: Case study of Sybil accounts within the same market.

Sybil pair		Vendor 1	Vendor 2
Case1	Market	SilkRoad2	Agora
	Username	drugsforyou	Drugs4you
	Image		
	Category	drug – cannabis	drug – cannabis
	Text	Drugs4you	Drugs4you
Sybil pair		**Vendor 1**	**Vendor 2**
Case2	Market	SilkRoad2	Agora
	Username	repaaa	RepAAA
	Image		
	Category	apparel	apparel
	Text	REPLICAAAA	REPLICAAAA

Figure 12: Case study of Sybil accounts across different markets.

Sybil pair		Vendor 1	Vendor 2
Case1	Market	Agora	Agora
	Username	RXChemist	Remedyplus
	Image		
	Category	drug – prescription, benzos, opioids	drug – prescription, benzos, opioids
	Location	US, UK, Australia	US, UK Australia
	Vendor description	Hello World, WELCOME TO ♛♛♛RXChemist♛♛♛ ★★★★★★★★★★★★★★★★★★★★★★★★★★★★ ♛♛♛About Us♛♛♛ With RXChemist, You have found your personal Online Pharmacy of trust. Not only are we providing our customers worldwide with high quality Brand – and Generic Medication, we will also offer a complete satisfaction guarantee. For the past 1 year we have built a reputation that made us become one of the leaders in the Agora Community, supported by our friendly support. ☺What To Expect When You Order From RXChemist☺ ✔ Quality Product ✔ Lowest Price In Market ✔ Quick Shipping ✔ Professional Service ✔ Friendly Communication ✔ Complete Privacy And Security	**********REMEDYPLUS********** **********WELCOME WELCOME********** JUST HAVE A LOOK AT FEW STEPS TO MAKE YOU ALL CLEAR MYSELF ✔ Professional and reliable service. I am here to serve you all ...for me Quality is my frst priority ✔ Prompt and friendly communication. All my buyers arefree to ask for as many question as they want to ask for ..24 7 SERVICE HAPPY TO SERVE YOU ALL :) ✔ 100% Re–Ship policy For USA,UK,AUS . 99.9% The shipping is smooth but in any case if dont reach to my buyer i will 100% re–ship it ··· ✔ High quality product. I never compromise in quality if any body in any case is not satisfied will be served with the extra pills of legit quality in his next order ✔ Cheapest Price in Darknet. most cheapest price but legit quality will be served !!!!

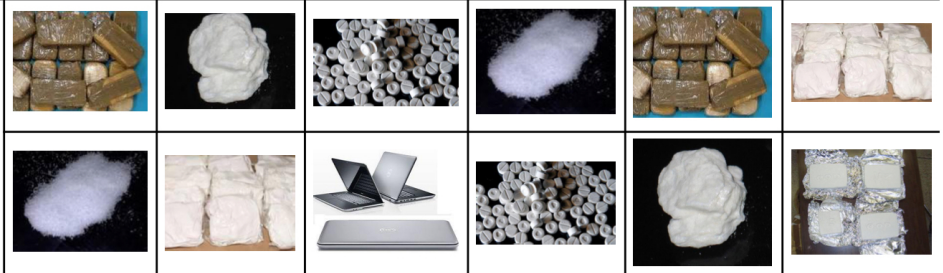
Non–Sybil pair		Vendor 1	Vendor 2
Case2	Market	SilkRoad2	SilkRoad2
	Username	kimbe	drugstore
	Image		
	Category	drug, electronics, jewelry	drug
	Location	Vatican	Croatia

Figure 13: Case study of adversarial vendors.