Photo-based Vendor Re-identification on Darknet Marketplaces using Deep Neural Networks

Xiangwen Wang

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Science and Applications

Gang Wang, Chair
Michel J. Pleimling
Danfeng Yao

April 17, 2018
Blacksburg, Virginia

Keywords: Darknet Market; Sybil Detection; Image Analysis; Stylometry

Photo-based Vendor Re-identification on Darknet Marketplaces using Deep Neural Networks

Xiangwen Wang

(ABSTRACT)

Darknet markets are online services behind Tor where cybercriminals trade illegal goods and stolen datasets. In recent years, security analysts and law enforcement start to investigate the darknet markets to study the cybercriminal networks and predict future incidents. However, vendors in these markets often create multiple accounts (*i.e.*, Sybils), making it challenging to infer the relationships between cybercriminals and identify coordinated crimes. In this thesis, we present a novel approach to link the multiple accounts of the same darknet vendors through photo analytics. The core idea is that darknet vendors often have to take their own product photos to prove the possession of the illegal goods, which can reveal their distinct photography styles. To fingerprint vendors, we construct a series deep neural networks to model the photography styles. We apply transfer learning to the model training, which allows us to accurately fingerprint vendors with a limited number of photos. We evaluate the system using real-world datasets from 3 large darknet markets (7,641 vendors and 197,682 product photos). A ground-truth evaluation shows that the system achieves an accuracy of 97.5%, outperforming existing stylometry-based methods in both accuracy and coverage. In addition, our system identifies previously unknown Sybil accounts within the same markets (23) and across different markets (715 pairs). Further case studies reveal new insights into the coordinated Sybil activities such as price manipulation, buyer scam, and product stocking and reselling.

Photo-based Vendor Re-identification on Darknet Marketplaces using Deep Neural Networks

Xiangwen Wang

(GENERAL AUDIENCE ABSTRACT)

Taking advantage of the high anonymity of darknet, cybercriminals have set up underground trading websites such as darknet markets for trading illegal goods. To understand the relationships between cybercriminals and identify coordinated activities, it is necessary to identify the multiple accounts hold by the same vendor. Apart from manual investigation, previous studies have proposed methods for linking multiple accounts through analyzing the writing styles hidden in the users' online posts, which face key challenges in similar tasks on darknet markets. In this thesis, we propose a novel approach to link multiple identities within the same darknet market or across different markets by analyzing the product photos. We develop a system where a series of deep neural networks (DNNs) are used with transfer learning to extract distinct features from a vendor's photos automatically. Using real-world datasets from darknet markets, we evaluate the proposed system which shows clear advantages over the writing style based system. Further analysis of the results reported by the proposed system reveal new insights into coordinated activities such as price manipulation, buyer scam and product stocking and reselling for those vendors who hold multiple accounts.

To my parents, and my wife Linjun.

Acknowledgments

I wish to express my sincere gratitude to Dr. Gang Wang, my advisor, for his kind and responsive guidance during the development of this work. His constructive feedback, continuous support and patience made this work possible. I would like to extend my gratitude to my committee members Dr. Michel Pleimling and Dr. Danfeng Yao for their valuable comments on this work, and to Peng Peng and Dr. Chun Wang for their contribution in this work. Special thanks to Dr. Michel Pleimling for providing me the opportunity to study in Computer Science.

I would like to express my thankfulness to my parents of their endless care and support throughout the years. I would like to deeply thank my wife Dr. Linjun Li for her constant encouragement and support during my study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cybercrimes, ranging from data theft to ransomware attacks, are posing a serious threat. In the past decade, cybercriminals have evolved rapidly, making it challenging for security researchers and the law enforcement to trace their activities and build proactive defenses [1, 23, 35]. Meanwhile, underground forums, particularly the darknet markets behind Tor [12], are increasingly popular among cybercriminals to *anonymously* trade illegal goods and stolen items (*e.g.*, credit cards, datasets). These platforms thus become the key information source for investigating the cybercrime ecosystem and predicting future incidents [42, 48].

As the key aspect of the investigation, researchers have been seeking to understand the relationships between cybercriminals and identify the stakeholders. Prior works have examined the social networks in underground forums to understand the user interactions [14, 15, 38, 61]. In the darknet markets, however, the key challenge of such investigation is that darknet vendors often maintain multiple accounts (or Sybil accounts) within the same market or across different markets. Without linking these accounts together, analysts might miss key opportunities to reveal the true relationships between cybercriminals and identify coordinated activities.

Unfortunately, due to growing scale of the darknet markets, it is highly labor-intensive to *manually* investigate and link multiple accounts. To solve this problem, existing approaches rely on *stylometry analysis*, which aims to link Sybil accounts based on their writing styles [2, 22]. Stylometry analysis has shown success in fingerprinting underground forum users where users post rich and diverse text, but it faces key challenges to fingerprint vendors in the darknet markets. First, the only available text in the darknet markets are product descriptions, which are short, repetitive, and often follow certain templates. Second, stylometry analysis is sensitive to the language of the content, which is a disadvantage to analyze darknet markets where vendors come from different countries (validated in §5).

In this thesis, we propose a novel approach to link multiple identities in the darknet markets by analyzing the product photos. Our goal is to build reliable fingerprints to re-identify the vendors based on their photos within the same market or even across different markets. This idea is motivated by the fact that darknet vendors often have to take photos for their own products (instead of using stock photos) to prove the possession of the illegal goods or stolen items. Such photos can reflect a vendor's personal style of photography. To build accurate fingerprints, we develop a system where a series of deep neural networks (DNNs) are used to extract distinct features from a vendor's photos automatically. In addition, to fingerprint vendors with relatively fewer photos, we apply *transfer learning* to pre-train the deep neural network with large generic image datasets and fine-tune the model with vendor-specific photos.

We evaluate the proposed system using real-world datasets from 3 large darknet markets (Agora, Evolution, SilkRoad2), which involves 7,641 vendors and 197,682 product photos. We first conduct a "ground-truth" evaluation by splitting a vendor's photos into two random parts and examining how accurately the system can link the two parts back. Our best performing model achieves an accuracy of 97.5% or higher for all three markets. In addition,

we compare our approach with existing stylometry methods that model a vendor's writing styles based on the product descriptions. We demonstrate that image-based approach excels in both accuracy of classification and the coverage of "fingerprint-able" vendors.

To demonstrate the usefulness of the proposed method, we apply our system to detect *previously unknown* Sybil accounts in the wild. Based on manual examinations and external evidence, we confirm that our system detected 715 Sybil pairs across different markets and 23 Sybil account pairs within the same markets. Further case studies reveal new insights into the coordinated activities of Sybil accounts, ranging from price manipulation and buyer scam, to product stocking and reselling, and photo plagiarizing. For example, we identify vendors on Evolution and SilkRoad2 who creates Sybil accounts that only sell a handful of products but at a much lower price. Some of the Sybil vendors are confirmed to have scammed the buyers based on external evidence. In addition, the detected Sybil pairs also reveal the relationships between vendors (*e.g.*, suppliers and retailers) which helps to identify the market stakeholders.

In summary, our contributions are three folds:

- **First**, we present the first system to fingerprint darknet vendors by modeling their unique styles of photography.

- **Second**, we perform ground-truth evaluations on the proposed system. Results show that the photo-based approach outperforms existing stylometry analysis in both accuracy and coverage.

- **Third**, we apply the system to detect previously unknown Sybil accounts in the wild. Extensive analysis of the detected Sybil pairs reveals new insights into the cybercriminal activities within and across darknet markets.

Our study is part of an ongoing effort to develop useful tools to assist the law enforcement and criminal analysts to investigate the cybercriminal networks. Our proposed method can contribute to building profiles of cybercriminals, establishing darknet vendor networks, understanding of darknet vendor reputation systems, and the study of the migration of vendors across different marketplaces. As a future work, we are interested in investigating how Sybils vendors can evade the detection by hiding their personal styles (detailed discussion in §8).

The content of this thesis has been accepted for publication in Proceedings of 2018 ACM Asia Conference on Computer and Communications Security [58].

Chapter 2

Background and Goals

In this chapter, we introduce the background of darknet marketplaces and describe our research goals.

2.1 Tor and Darknet Markets

Tor (short for "The Onion Router") is the most widely used tool for anonymous communications on the Internet [12]. Tor conceals a user's IP and location by redirecting her network traffic through a large overlay network consisting of thousands of relays. Tor not only protects users from network surveillance and censorship but also helps a large number of darknet websites to operate anonymously. Users can access darknet websites through Tor without knowing their actual IP or location. However, the anonymity also creates a challenge for the law enforcement to trace the illegal websites in the darknet [17].

Darknet market is a particular type of trading website in the darknet. Most of the darknet markets are set up by cybercriminals around the world to trade illegal goods (*e.g.*, drugs,

fire weapons), stolen items (*e.g.*, credit cards, password datasets), software exploits, and even criminal/hacking services. Researchers have collected empirical datasets from darknet markets to study the products offered, the revenue and the market dynamics over time [17, 48]. A key difference between the darknet markets and traditional underground forums [1, 15, 23, 27, 35, 42] is that darknet markets are hosted behind Tor, making them difficult to trace and take down.

2.2 User Identities in the Darknet Markets

To study the development of darknet markets, a key challenge is to trace and link user identifies in the markets. Users, particularly the vendors, often create multiple identities (*i.e.*, Sybil accounts) within the same markets or across different markets [2, 22]. The Sybil identities are created either to increase sales or even scam buyers. Due to the strong anonymity of darknet users, it is difficult to effectively link user identities based on traditional IPs or device fingerprints. In addition, given the large number of darknet markets and the user accounts, manual investigation faces key challenges to scale up.

2.3 Stylometry Analysis

Recently, researchers have explored to use *stylometry* to link a user's multiple identities. Stylometry analysis is a standard technique to attribute authorship of anonymous texts by modeling the writing style. The techniques have shown success in re-identifying users in online forums [28, 33, 49] and fingerprinting the programmers of software code [6]. A related work has explored to attribute the authorship based on users' public and private messages posted on underground forums [2].

Directly applying stylometry analysis to darknet markets faces key challenges. First, stylometry analysis requires lengthy text to model a user's writing style. Unlike the rich and diverse text messages available in online forums, the only text on the darknet markets are the *product descriptions* posted by the vendors. The product descriptions are usually short and repetitive (following certain templates). In addition, the product descriptions are often written in different languages by vendors from all over the world, making it difficult to perform stylometry analysis. We have confirmed these challenges in §5.

2.4 Our Goals

In this thesis, we develop novel tools to fingerprint vendors in the darknet marketplaces. The goal is to help investigators to identify and link the multiple identities controlled by the same vendors by analyzing the posted product photos. This idea is motivated by two key intuitions. First, unlike regular e-commerce websites (*e.g.*, Amazon), darknet vendors often need to take pictures of their illegal goods by themselves. Second, photographs can reflect the photographers' unique personal styles [16, 24, 56].

Our exploration contains three key steps: First, we seek to use the product photos posted by vendors to build a distinct profile (or fingerprint) for each vendor. We propose to extract the distinct features from their photos using deep neural networks (§4). Second, we seek to compare (and potentially augment) the photo-based fingerprints with traditional stylometry analysis on product descriptions (§5). Finally, we apply our system in the wild to identify previously unknown Sybils accounts both within the same markets and across different markets (§6). We perform case studies to understand the behavior of Sybil accounts, and demonstrate the usefulness of the tool (§7).

Chapter 3

Data

To examine the possibility of profiling darknet vendors, we leverage the public archive of darknet market datasets [5]. The data archive contains the daily (sometimes weekly) snapshots of the darknet markets crawled by researchers from 2013 to 2015. Each snapshot contains the raw product pages of the respective marketplace. In this thesis, we select 3 largest markets: Agora, Evolution, and SilkRoad2.

For each market, we wrote a customized parser to extract structured data for the product pages. For each product, we obtain the product ID, product description, product image, vendorID, the vendor's pseudo name, and the timestamps when the product was actively listed on the market. Table 3.1 shows the basic statistics. Below, we briefly introduce the background of the 3 markets and validate the data integrity.

SilkRoad2: established in November 2013, SilkRoad2 was the successor of the well-known market SilkRoad (taken down by FBI in October 2013) [11]. Due to the brand attraction of SilkRoad, SilkRoad2 quickly became the largest darknet market in 2014. In February 2014, SilkRoad2 was compromised, losing 2.6 million USD worth bitcoins,

Market	Unique Product	Unique Vendor	Vendor w/Imgs	Image Count	Time Span
Agora	96,821	3,162	2,834	75,979	01/2014–07/2015
Evolution	82,286	4,197	3,635	89,145	01/2014–03/2015
SilkRoad2	32,558	1,332	1,172	32,558	12/2013–11/2014
Total	211,665	8,691	7,641	197,682	12/2013–07/2015

Table 3.1: Basic statistics of the darknet dataset.

which led to a major damage to its reputation [4]. On November 6, 2014, SilkRoad2 was taken down by authorities and its administrator was also arrested.

Evolution: established in January 2014, Evolution was the largest darknet marketplace after the taken down of SilkRoad2. In March 2015, the administrators of Evolution unexpectedly shut down the market and took away all the bitcoins that users deposited to the market, the value of which was estimated to be 11.7 million US dollars [59]. The site then went offline since this "exit scam".

Agora: established in 2013, Agora once became the largest market after the taken down of SilkRoad 2 and the exit scam of Evolution [48]. The market was taken offline by its administrators due to security vulnerabilities in August 2015, and stayed offline since then.

Although all three markets went offline, the dataset provides a unique opportunity to retrospectively study the vendor behavior and inform the investigation of the emerging markets. As shown in Table 3.1, we extracted in total 211,665 products listed by 8,691 vendors from the three markets. 7,641 of the vendors have posted at least one product image (88%). In total, we obtained 197,682 product images. We find that the distribution of the image count per vendor exhibits a long-tail property as shown in Figure 3.1. Note that vendors sometimes use the same image for different products, and thus we display both the total image count and the unique image count (the identical images are identified by MD5 hashes).

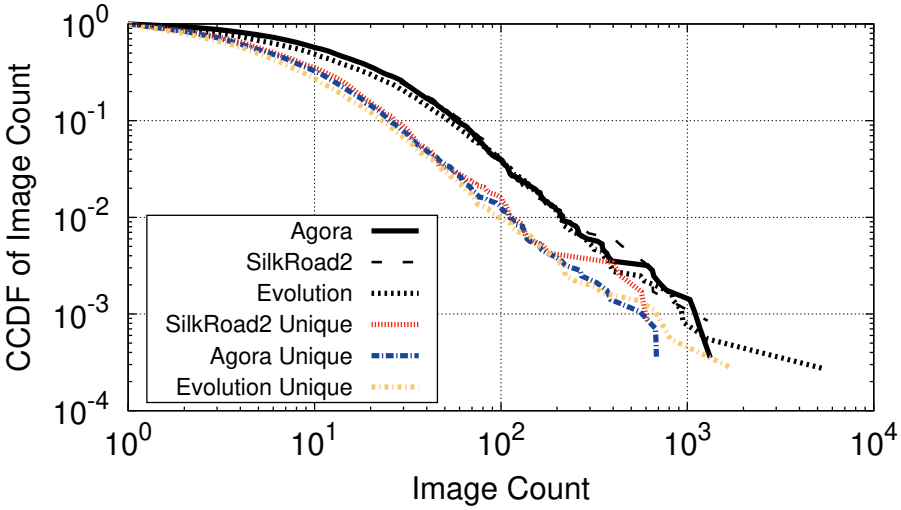


Figure 3.1: Number of product photos per vendor, including the total number and the unique number of photos.

3.1 Validation of Data Integrity

Before using the dataset, we have validated the data integrity. Our concern is that if the crawlers had a major downtime, the data quality would be seriously affected. Without the actual ground-truth, we rely on the statistics reported by related studies and check the over-time consistency of the dataset. First, according to a measurement study, there were about 2200 active vendors on Evolution, 1200 vendors on Agora, and 800 vendors on SilkRoad2 by the end of 2014 [48]. The corresponding numbers in our dataset (2014-2015) are 4197, 3162, and 1332 respectively, which are consistently higher. This is likely due to the growth of the markets. In addition, Figure 3.2 shows the accumulative number of distinct products listed on the markets over time. The curves have smooth upward trends without obvious plateau, indicating a good data integrity.

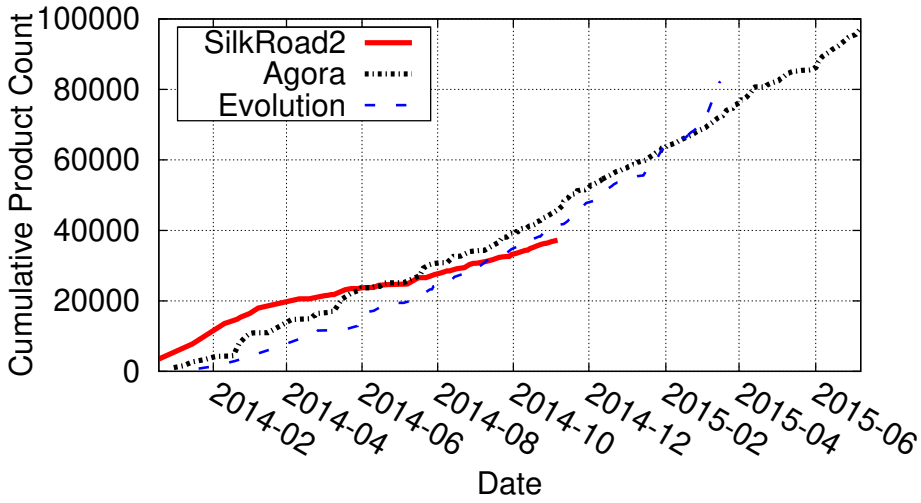


Figure 3.2: Cumulative product count over time.

3.2 Image Metadata

During our data processing, we find that certain images contain the EXIF metadata. When a camera takes a photo, it can add metadata to the photo including the camera information, the timestamp and even the location where the photo is taken. The metadata is tagged following the standard Exchangeable Image File Format (or EXIF). Our analysis shows that darknet markets have realized the problem: Agora and Evolution started to remove the EXIF metadata from all the uploaded photos since June and March of 2014. In total, there are 1,604 vendors who had at least one photo with EXIF metadata, and 112 vendors revealed their *location information* through the metadata. The EXIF metadata only affected a small number of early vendors, and most markets today remove the metadata by default. To this end, our system did not consider the EXIF information (removed from our dataset).

3.3 Ethics of Data Analysis

The darknet datasets in this thesis were originally collected by previous researchers [5] who made the data publicly available under the Creative Common CC0 license. We follow the standard ethical practice to analyze the datasets [11, 48]. First, our analysis only covers darknet markets that have been taken down by authorities. Second, the dataset only contains the publicly available information on the darknet markets (product pages). The dataset does not contain any personally identifiable information. Third, our data analysis is completely passive without any form of interactions with the human subjects. Finally, our research produces useful tools to support researchers and the law enforcement to trace, monitor, and investigate cybercrimes. The benefit of the research significantly outweighs the potential risks.

Chapter 4

Image-based Vendor Fingerprinting

Next, we describe our method to fingerprint darknet market vendors by analyzing their posted photos. In this chapter, we describe our deep-learning based method to building the fingerprints for vendors, and perform ground-truth evaluations using empirical datasets.

4.1 Method and Designs

To fingerprint a vendor based on her photos, we need to identify key features that can uniquely represent the vendor. Related work has explored fingerprinting specific *camera* devices using low-level features, *e.g.*, the unique sensor noise and lens distortions caused by manufacturing imperfection and sensor in-homogeneity [9, 10, 34, 43, 55]. However, previous works on photograph authorship attribution suggested that the high-level features (*e.g.*, object, scene, background, camera angle and other latent photography styles) significantly outperformed low-level features to identify photographers [56]. To this end, we choose high-level features for darknet vendor identification.

To capture the unique features from a vendors' photos, we rely on Deep Neural Networks (DNN) which can extract features automatically without manually crafting the feature list [32]. The key challenge is that deep neural networks, in order to be accurate, requires a massive amount of training data. However, in darknet markets, the number of photos per vendor is limited as shown in Figure 3.1. To this end, we apply *transfer learning* to pre-train a deep neural network using a large existing image dataset (with millions of images) and then fine-tune the last few layers using the smaller darknet dataset. The intuition is that features of the deep neural network are more generic in the early layers and are more dataset-specific in the later layers.

The early layers can be trained using general object photos. For our system, we use the largest annotated image dataset called ImageNet [45] (14 million images) to pre-train a deep neural network. Then we replace the final softmax layer with a new softmax layer which handles the classes in the darknet dataset. Here, a "class" is defined as a set of photos uploaded by the same vendor. Next, we fine-tune the last layers or all layers with back-propagation using the vendors' product photos. The fine-tuning process is implemented using a stochastic gradient descent optimizer with a small initial learning rate, aiming to minimize the cross-entropy loss function. We follow the standard procedures to fine-tune a neural network using toolkits such as TensorFlow and Keras.

To construct the deep neural network, we select 5 popular models for generic image classification tasks. For each model, we re-implement the data feeding module and the prediction module and select the most popular configurations on their respective tasks. The most popular configurations are usually those that lead to the highest accuracy with an acceptable computational overhead. For image pre-processing, we reshape the darknet images to the same sizes of the images that are used in the pre-trained models. We then use the ImageNet utility module in Keras for image preparation.

AlexNet was introduced by Krizhevsky et al. in 2012 [30]. Our code is based on Kratzert's implementation of AlexNet using TensorFlow [29]. The images are reshaped to 227 × 227. The early layers are kept fixed and only the last three layers (fc6, fc7, fc8) of the network are fine-tuned.

Inception models are a series of DNN models introduced by Szegedy et al [53] in 2014–2017. We choose the latest Inception-V4. Our code is based on Yu's implementation [62], where all network layers are fine-tuned. The images are reshaped to 299 × 299.

VGG models were introduced by Simonyan and Zisserman in 2014 [47]. Here we adopted the 19-layer VGG-19 model. The images are reshaped to 224 × 224 (same for ResNet and DenseNet below).

ResNet was introduced by He et al. in 2015 [21]. In our analysis, we adopted ResNet-50 model for its good balance of accuracy and computational overhead.

DenseNet or Densely Connected Convolutional Network was introduced by Huang et al. in 2016 [25]. We adopted DenseNet-121 model for its good performance.

Using the deep neural network model, we train a multi-class classifier where each class represents a vendor in the darknet market. Given an input image, we use the classifier to calculate the probability that the image belongs to a given vendor. Based on the "similarity" of images, we identify pairs of accounts that are likely to be controlled by the same vendor.

4.2 Ground-Truth Evaluation

To evaluate the feasibility of our approach, we first perform a ground-truth evaluation. Due to the high-anonymity of the darknet marketplaces, it is impossible for us to build the actual

ground-truth. One convincing way to build the synthetic ground-truth is through splitting the data of certain vendors. For a given vendor, we randomly split her photos into two even parts. We use the first half to train the classifier and then try link the second half to the original vendor. This evaluation is to examine the feasibility of our approach and help to fine-tune the parameters. Later in §6 and §7, we will apply our method to identify previously unknown multiple-identities controlled by the same vendors in the wild.

4.2.1 Ground-truth Construction

For a given vendor, we evenly split her data into two pseudo vendors. Here we need to introduce a threshold T_r which specifies the minimal number of photos that the vendor has in order to build the fingerprint. We will test different thresholds in our evaluation.

We observe that some vendors use the same photo for different products (based on the product ID). To test the feasibility of re-identifying vendors based on their photo-styles (instead of simply matching the same photos), we create two versions of the ground-truth datasets. For the *duplication* version, we consider all of the vendor's product photos. Each product's photo only counts for once, but we allow different products to use the same photo. For the *non-duplication* version, we intentionally remove the duplicated photos that are used for different products. The duplicated photos are determined by their MD5 hashes.

4.2.2 Evaluation Workflow

Figure 4.1 shows the evaluation workflow. First, for vendors that have more than $2 \times T_r$ photos, we split their photos into two even parts as the pseudo vendors. We add the first part to the training dataset and the second part to the testing dataset. Second, for the other vendors, if their image count $> T_r$, we add them to the training dataset as the "distractors".

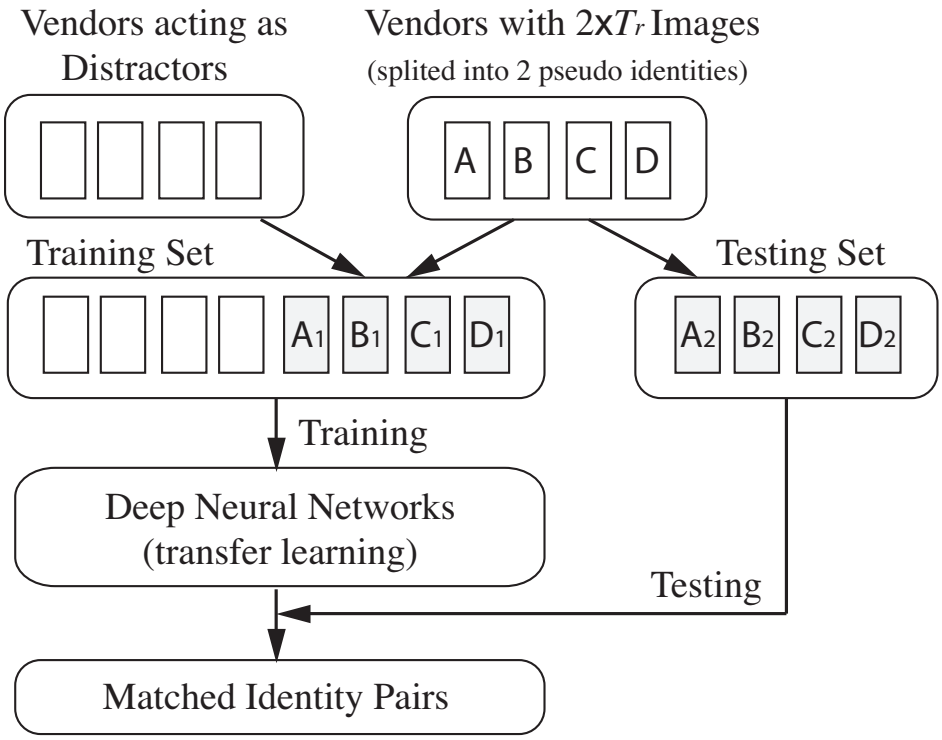
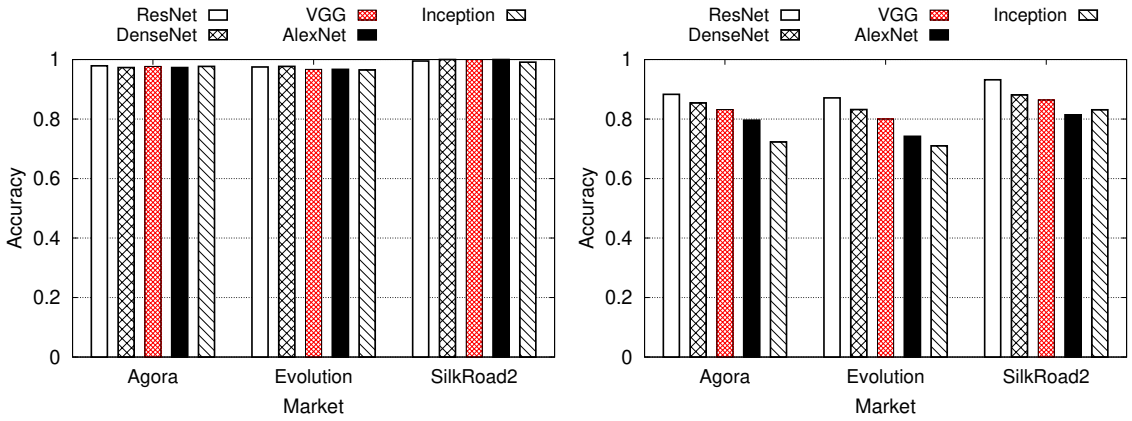


Figure 4.1: Workflow for the ground-truth evaluation.

The number of classes in the training set equals to the number of pseudo pairs plus the number of training distractors shown in Table 4.1. The number of classes in the testing set equals to the number of pseudo pairs. Once we construct the dataset, we then perform transfer learning based on a model pre-trained on ImageNet, and use our training dataset to fine-tune the last layers of the network.

During testing, for each image in the testing dataset, we calculate its probability of belonging to a given vendor in the training set. Then those probabilities are averaged over the images that below to the same vendor, which leads to a *similarity metric* for each "training –testing vendor" pair. In this way, for each testing vendor, we identify the most similar training vendor and examine if the pseudo vendors are correctly paired. We calculate the *accuracy* which is the ratio of the testing vendors that are correctly matched.



(a) With duplicated images (b) Without duplicated images

Figure 4.2: Comparison of different DNN models. $T_r = 20$ for all the settings.

4.3 Evaluation Results

4.3.1 Accuracy

Table 4.1 shows the detailed results for AlexNet and ResNet. Across different markets and parameter settings, the matching accuracy is very high. Consistently, ResNet is more accurate than AlexNet. For all three markets, ResNet has a matching accuracy of 0.975 or higher when we don't intentionally remove duplicated images for different products.

Even after we remove the duplicated images, the matching accuracy is still around 0.871–0.932 for ResNet (for T_r=20). Recall that this is a multi-class classifier with hundreds of classes. An accuracy of 0.871 (for the top-1 matching candidate) is already very high. In practice, analysts may consider the top-K matching candidates (where K is a small number) instead of just the most likely one. The accuracy metric then should measure how likely the top K candidates contain the correct match. For example, applying ResNet (T_r=20) on non-duplicated images returns the top-5 accuracy of 0.964 for Agora, 0.948 for Evolution, and 0.966 for SilkRoad2. The result indicates that the same vendors' photos do carry distinct

Figure 4.3: The ROC curves from the ResNet model ($T_r = 20$, with duplicated images).

personal styles, which can be used to build reliable fingerprints.

Regarding the threshold T_r, a lower threshold allows us to consider more vendors. However, if T_r is too small, then there might not be enough training data for each vendor, which reduces the matching accuracy. For the rest of the thesis, if not otherwise stated, we set the threshold $T_r = 20$.

To compare different DNN models, we present Figure 4.2. Overall, ResNet achieves the best performance. This is not too surprising considering that ResNet is a relatively advanced model for object recognition tasks [7]. However, our performance is not completely aligned with the model performance on object recognition. The InceptionV4 model is the state-of-the-art for ImageNet, but InceptionV4 actually performs the worst on the darknet datasets. Intuitively, the photos posted by vendors are very different from those in ImageNet. ImageNet rarely covers photos of marijuana, cocaine, or stolen credit cards. Overall, the performance differences are not very big between different DNN models. This indicates that our task is not very sensitive to the model selection.

Dupli. Img.	Market	T_r	Pseudo Pairs	Training Distractors	AlexNet Accuracy	ResNet Accuracy
Yes	Agora	10	1020	597	0.969	0.975
		20	480	540	0.973	0.979
		40	161	319	0.950	0.975
	Evolution	10	1093	680	0.952	0.964
		20	519	574	0.967	0.975
		40	197	322	0.990	0.990
	SilkRoad2	10	415	248	0.976	0.980
		20	211	204	1.00	0.995
		40	76	135	0.987	1.00
No	Agora	10	408	518	0.733	0.821
		20	137	271	0.796	0.883
		40	45	92	0.733	0.867
	Evolution	10	443	546	0.626	0.788
		20	155	288	0.742	0.871
		40	47	108	0.830	0.915
	SilkRoad2	10	181	233	0.724	0.873
		20	59	122	0.814	0.932
		40	24	35	0.875	0.958

Table 4.1: Accuracy of ground-truth vendor matching based on image analysis.

4.3.2 True Positives vs. False Positives

In the above evaluation, we always report a match (*i.e.* the most-similar training vendor) for a given testing vendor. However, in practice, not every vendor has a matched Sybil identity. To this end, we will need to draw a minimal probability threshold T_p to declare a match. Our system will report a match only if the similarity score between the testing vendor and the training vendor is higher than T_p.

The threshold T_p determines the trade-off between true positives (the correctly detected vendor pairs) and false positives (the detected vendor pairs that turn out to be false). To examine this trade-off, we slightly modify our workflow of Figure 4.1. Now, given the set of distractors, we randomly put half of the distractors into the training set and the other half into the testing set. By swapping T_p, we generate the ROC (Receiver Operating Charac-

teristic) curves in Figure 4.3. The results again confirm the good performance. The ROC curves all reach the top-left corner of the plots, and the areas under the curves (AUC) are close to 1.0 (a random classifier's AUC would be 0.5 and a higher AUC is better).

In practice, analysts can make their own trade-off between false positives and true positives based on their time budget. If the time allows, the analysts can afford to have some false positives so that they don't miss the actual Sybil identities of a given vendor. In this thesis, we use the ROC curves to pick the threshold T_p based on the elbow point of the curve. The corresponding T_p is about 0.4 when we allow duplicated images. The elbow T_p is 0.2–0.3 if duplicated images are intentionally removed.

Chapter 5

Comparison with Stylometry

Our evaluation so far shows that the image-based approach is effective to fingerprint vendors. Next, we explore to compare our method with existing stylometry approaches, and seek to further improve the accuracy and the coverage of the system. In the following, we briefly introduce the existing stylometry analysis methods and the unique challenges to apply them to the darknet markets. Then we evaluate the number of vendors that stylometry analysis can effectively fingerprint, and the matching accuracy in comparison with the image-based approach.

5.1 Stylometry Analysis

Stylometry analysis aims to attribute the authorship of the anonymous texts by analyzing the writing style. Existing works have explored the feasibility of identifying the authorship of underground forum posts [2] and even computer programs [6]. To this end, the stylometry analysis is a valid comparison baseline for our method. In the darknet markets, a vendor's texts are the product descriptions written by the vendor. However, there are key challenges

for stylometry analysis in darknet markets. First, the product descriptions are usually very short. For example, the median length of Agora's product descriptions is only 118 words. Second, the product descriptions often follow certain templates, and vendors may use the same/similar descriptions for many of their products. Third, most darknet markets are international marketplaces where vendors may use different languages. All these factors pose challenges to extract the unique writing styles of the vendors.

To examine the feasibility of stylometry analysis, we follow the most relevant work [2] and re-implement a similar stylometry classifier. More specifically, given the collection of the text of a vendor, we extract a list of features to model the writing styles. The features include: the percentage of words that start with an upper-case letter, percentage of upper-case letters, average word length, word length histogram, punctuation frequency, stop-word frequency, character unigram, bigram and trigram, Part-of-Speech (POS) unigram, bigram, and trigram, and digit unigram, bigram, and trigram. We used the NLTK library [39] to perform word and sentence tokenization. We applied Stanford Log-linear Part-Of-Speech Tagger [57] to extract the POS features. Considering the high dimensionality of the feature vector (about 100K), we also perform dimension reduction using stochastic singular value decomposition (StochasticSVD) to reduce feature vector size to 1000. Then we use the feature vector to train a logistic regression classifier to make predictions. We refer interested readers to [2] for more details.

5.2 Performance Comparison

Our evaluation focuses on comparing the image-based approach and the stylometry based approach. The goal is to understand whether we can use stylometry analysis to further augment the image-based method. Our evaluation metrics include two key aspects: *accuracy*

(the accuracy to match pseudo identities) and *coverage* (the number of vendors that can be reliably fingerprinted).

Our evaluation follows the same work-flow in Figure 4.1. To generate ground-truth data for stylometry analysis, we again split vendors whose product descriptions with more than $2 \times T'_r$ words. For vendors with more than T'_r words, we add them as the distractors in the training set. Similar to before, we create two versions of the ground-truth datasets, one considers all the product descriptions (one description per product) and allows *duplicated sentences*. The other ground-truth dataset removes the duplicated sentences. The non-duplicated version aims to force the classifiers to learn the writing style instead of matching the exact sentences. In this evaluation, we only consider English text — we have removed Unicode symbols and HTML entities.

5.2.1 Accuracy

Table 5.1 shows that the stylometry analysis can also achieve a high accuracy when we allow the duplicated sentences (0.936–0.990). However, when we remove the duplicated sentences, the accuracy dropped significantly to 0.580 – 0.846. This dramatic accuracy decrease indicates that the previous high accuracy is likely the results of matching the duplicated sentences, instead of actually extracting the generalizable "writing styles". Our result shows that the same method that works well in underground forums [2] has major limitations in darknet markets. Consider that vendors often follow templates to write product descriptions, it is understandable that their personal writing styles are not as strong as the template-free texts in underground forums.

Duplicated Sentences	Market	T'_r	Pseudo Pairs	Training Distractor	Accuracy
Yes	Agora	1500	822	515	0.983
		3000	402	420	0.988
		4500	247	316	0.988
	Evolution	1500	530	404	0.936
		3000	246	284	0.967
		4500	159	179	0.987
	SilkRoad2	1500	338	200	0.970
		3000	169	169	0.988
		4500	99	120	0.990
No	Agora	1500	193	300	**0.694**
		3000	72	121	0.806
		4500	39	63	0.846
	Evolution	1500	162	235	**0.580**
		3000	65	97	0.723
		4500	33	59	0.818
	SilkRoad2	1500	65	126	**0.631**
		3000	25	40	0.800
		4500	12	26	0.833

Table 5.1: Accuracy of ground-truth vendor matching based on stylometry analysis.

5.2.2 Coverage

Stylometry analysis has a more limited coverage. Table 5.2 shows the number of qualified vendors for stylometry analysis and image analysis, given the threshold that produces a comparable accuracy ($T_r = 20$ and $T'_r = 4500$). Note that $T'_r = 4500$ returns the highest accuracy for stylometry analysis, but it is still not as accurate as the image analysis (after removing duplicated images). Meanwhile, the image analysis covers 100%–300% more vendors than the stylometry analysis. The advantage is more significant when duplicated texts or images are removed.

Duplicated texts/images	Market	Image $T_r = 20$	Stylometry $T'_r = 4500$
Yes	Agora	1020	563
	Evolution	1093	338
	SilkRoad2	415	219
No	Agora	408	102
	Evolution	443	92
	SilkRoad2	181	38

Table 5.2: Number of qualified vendors given the thresholds for image analysis ($T_r = 20$ images) and stylometry analysis ($T'_r = 4500$ words).

5.2.3 Run Time

The image analysis also has a shorter runtime by taking advantage of the GPUs. For example, the image analysis for Agora (ResNet, $T_r = 20$, with duplicated images) takes one server 4 hours to finish the whole process including data preparation, training, and testing. The server has one quad-core CPU and one Nvidia GTX 1070 GPU. However, the stylometry analysis on Agora ($T_r = 4500$, with duplicated texts) takes as long as 84 hours to finish (CPU only). In theory, it is possible to re-design the algorithm of [2] to work with GPU, but it would take significant efforts to rewrite the system, particularly the Part-of-Speech tagging algorithm.

In summary, the image-based approach has a clear advantage over stylometry analysis to fingerprint darknet vendors. However, these two techniques are not necessarily competing but can work together to add additional layers of confidence. In the rest of the thesis, we primarily use the image-based approach to detect Sybil identities in the wild, and check the writing style for confirmation during the manual inspection.

Chapter 6

Sybil Identity in the Wild

To demonstrate the usefulness of our system, we apply it to real-world datasets to identify previously unknown Sybil identities in the wild. We focus on two types of Sybil identities. First, we look for vendors who controlled multiple accounts *within the same market*, *i.e.*, intra-market Sybils. Second, we look for vendors who controlled multiple accounts *across different markets*, *i.e.*, inter-market Sybils.

6.1 Detection Method

In the following, we introduce the Sybil detection method, which is based on the image-based approach described in §4.

6.1.1 Inter-Market Sybils

To detect Sybil accounts in different markets, we work on two markets at a time. For market A and B, we use vendors from market A as the training data to build the classifier

and then test on vendors from market B. This step produces the similarity score for any two vendors $S(u_{Ai}, u_{Bj})$ from the two markets. Then, we reverse the order by training on B's data and testing with A's vendors to calculate a new similarity score for the same vendor pair $S(u_{Ai}, u_{Bj})$. The final similarity score between u_{Ai} and u_{Bj} is the average value: $Sim_{u_{Ai},u_{Bj}} = \frac{S(u_{Ai},u_{Bj})+S(u_{Bj},u_{Ai})}{2}$. We set parameters based on the ground-truth evaluation in §4. We focus on vendors with more than $T_r = 20$ photos and set $T_p = 0.4$ as the cut-off threshold for the final similarity score.

6.1.2 Intra-Market Sybils

To detect Sybil accounts in the same market, we again consider vendors who have more than $T_r = 20$ photos. We treat these vendors as the training set to build the classifier. We treat the same set of vendors (with more than 20 photos) as the testing set, and apply the classifier to identify the most similar vendors in the same market. We use $T_p = 0.4$ as the cut-off threshold for the similarity score based on the ground-truth evaluation. Note that this is not a standard machine learning process since the training and testing sets are overlapped. Instead, we are using the multi-class classifier to calculate the "distance" between vendors to identify similar pairs.

For both intra- and inter-market detection, we consider all the photos of a vendor (one photo for each product) without intentionally removing the reused photos. Using the above thresholds, the analysis covers 1,020 vendors in Agora, 1,093 vendors in Evolution and 415 vendors in SilkRoad2 (2,528 vendors in total). We use the most accurate ResNet DNN model for both cases.

6.2 Manual Investigation

To validate the accuracy of the detection, we act as the analysts to manually examine the detected candidate vendor pairs. Our investigation focus on *precision*, which is the ratio of true Sybil pairs out of all the candidate pairs. This analysis does not cover *recall*, given that there is no ground-truth about real-world Sybil accounts. We follow the guidelines below for our manual examination:

For the cross-market pairs, we first check their usernames and alias. If their usernames are identical (case-insensitive), or similar (*e.g.*, with an edit distance <= 1, or one username is the sub-sequence of the other), we label the pair with a "confident Yes".

Then for the rest of the cross-market pairs and all the same-market pairs, we examine the following aspects. (1) We check whether their aliases carry the same or related semantic meanings. For example, ("PeterLusting", "LustingPeter") and ("aaapee911", "evopee911") fall in this category. (2) We check if their photos contain explicit trademarks (or watermarks); We check the background environment of the photos (*e.g.*, photos of the same desk or shelf); (3) We manually read the product descriptions to look for the same/similar shipping information, payment method description and precaution information; (4) We examine the type of products they sell; (5) We check vendor review websites (*e.g.,* the "DarkNetMarkets" section on Reddit) where buyers rate the vendors. Sometimes the buyers who are familiar with the vendor would reveal the vendors' multiple identities. We label the pair as a "confident Yes" if we find strong evidence for either (1), (2), or (5). We label the pair as "probably Yes" if we find some evidence for both (3) and (4). Otherwise, we label the pair as "probably No".

Markets	Candidate Pairs	Confident Yes	Probably Yes	Probably No
Agora-Evolution	402	390	6	6
Agora-SilkRoad2	209	196	5	8
Evolution-SilkRoad2	144	129	5	10
Total	755	715	16	24

Table 6.1: Cross-market Sybil identification result.

(a) Vendor "apoteket" on Agora

(b) Vendor "swecan" on Evolution

Figure 6.1: An example of cross-market Sybil pair identified by our algorithm.

6.3 Sybil Detection Results

In total, we identified 850 candidate sybil pairs and 738 pairs are "confident yes" (87%). Table 6.1–6.2 show the detailed breakdown for Sybil pairs on different markets and those from the same markets.

6.3.1 Sybils on different Markets

Table 6.1 shows the cross-market detection result. The vendor pairs under "confident Yes" take more than 90% of all the candidate pairs. The high matching precision again confirms the usefulness of our method. Some Sybil pairs have the same usernames (484), but many

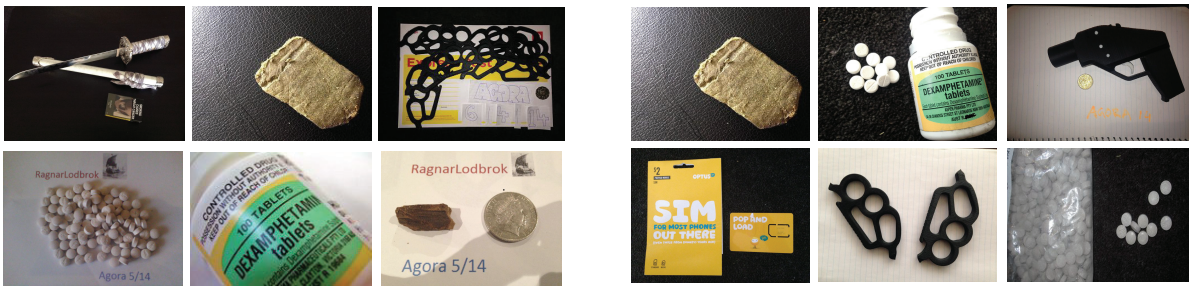
pairs have different names (271). This suggests that our technique cannot be trivially replaced by simply matching usernames. For example, 104 candidate pairs have very different usernames (*i.e.*, edit-distance >=2, and one username is not a substring of the other). More than 60% of these pairs are detected as "confident Yes" (which will be missed by simple name matching). Below, we provide examples of different manual labels, and explain false positives.

Most of the Sybil pairs under "confident Yes" are not difficult to verify. It occurred to us that vendors were not intentionally hiding their multiple accounts on different markets. Instead, some vendors even advertise their other accounts and maintain a consistent trademark for their brand. This explains why some vendors use the same or similar usernames. For example, ("marvel-labs", "Marvel-Labs") and ("GoingPostal", "goingpOZtal"). Some Sybils pairs try to make their usernames *sound similar*, *e.g.*, "Concentrate-Dealer" and "Concentr8-Dealer". Among the confirmed Sybil pairs, some vendors have a unique writing style. For example, for "RastainPeace" (Agora) and "DrugsLover" (Evolution), both accounts like to write sentences ending with the word "seriously".

Figure 6.1 shows randomly selected images from a confirmed Sybil pair. The two accounts share a high image similarity score ($Sim = 0.505$), and we obtain external confirmation from the buyers' online discussions. This vendor has a clear photo-shooting style. He/she likes to take close shots of the piles of the drug pills. The vendor also likes to use the same black tabletop or a piece of thesis as the photo background. We also notice that the vendor has re-shaped/cropped the images before putting them onto different markets. The product description contains both Swedish and English and thus stylometry analysis does not work for this vendor. In fact, the two usernames are somehow connected: "swecan" sounds similar to "Sweden", while "apoteket" means pharmacist in Swedish.

Markets	Candidate Pairs	Confident Yes	Probably Yes	Probably No
Agora	49	14	12	23
Evolution	32	6	7	19
SilkRoad2	14	3	3	8
Total	95	23	22	50

Table 6.2: Intra-market Sybil identification result.



(a) Vendor "RagnarLodbrok" on Agora

(b) Vendor "ReadySetGo" on Agora

Figure 6.2: An example of same-market Sybil pair identified by our algorithm.

6.3.2 Sybils in the Same Market

As shown in Table 6.2, intra-market Sybils are less common compared to the inter-market Sybils. Only 95 pairs are detected and only 23 pairs are "Confident Yes". A possible explanation is darknet markets usually prohibit a vendor from registering multiple accounts in the same market to prevent abuse. In contrary, it is common for a vendor to maintain accounts in different markets to maximize their sales.

Figure 6.2 shows an example Sybil pair from Agora. The two accounts do not have many identical photos, but the styles of the photos have similarities. This vendor likes to place the products on a black table to take photos. In addition, some of the products are the same even though the photos are shot from different angles. Finally, the vendor also likes to place a coin (*e.g.*, a quarter) in the photo to reference the size. Manual analysis also shows that

they have similar writing styles in the product descriptions.

(a) Vendor "streetdreams" (b) Vendor "keydemon"

Figure 6.3: An example of a false positive. The two vendors are incorrectly matched due to the red text in the images. The red text is the username of the respective vendor.

6.3.3 Sybil Pairs of Low Confidence

Our goal is to significantly reduce the scope of manual analysis. Instead of manually checking all the possible vendor pairs, we select the most suspicious ones for human analysts. For example, the above analysis covers 1020 vendors in Agora, 1093 vendors in Evolution and 415 vendors in SilkRoad2 (2528 vendors in total). Intra-market analysis calculate the similarity score for 1,203,637 pairs, and inter-market analysis examines 1,99,1755 pairs. Clearly, the total 3,195,392 pairs are beyond the capacity of manual inspection, and our algorithm has helped the security analysts to narrow down to 850 candidate pairs. This process inevitable introduces false positives. In general, our DNN based approach is designed for object recognition and analyzing the vendors' photo styles. The model is good at identifying similar object shapes and colors, and the background texture, but cannot make sense of the photos like a human analyst.

We have a few pairs under "Probably Yes" (38). For example "Gnarl" (Evolution) and "modalsol" (Evolution) both sell drugs with images of molecular structures. However, through Google image search, we find that they were using stock images from Wikipedia instead of taking their own photos. We cannot guarantee that the two accounts belong to the same vendor. Another example, is "griffin36" and "Cafe_Deluxe". The two accounts use

the same product images, but all the image seem to be stolen from other vendors (based on the visible watermarks on the images).

For the 74 pairs under "Probably No", evidence suggests that they are likely to be different accounts. For example, "subzero!!" (Agora) and "piblz" (Evolution) posted many similar images, but their writing styles are quite different and have different shipping information. In addition, "subzero" always add this sentence "Read our Profile for more about what we do" to the product description while "piblz" never do that. Some of these pairs look like *false positives* caused by the DNN classifier. For example, Figure 6.3 shows the two vendors that are incorrectly matched due to the red text in the images. The red text is the username of each vendor (as the trademark). The deep neural network picked up the red-color area as a feature, but could not tell the difference between the text.

6.3.4 Computation Costs

The whole evaluation process takes 1 day to finish using a single server (one quad-core CPU and one Nvidia GTX 1070 GPU). Although we need to compare the similarity for $N^2/2$ pairs (N is the total number of vendors), the actual complexity is only $O(N)$. This is because deep neural networks allow us to train a multi-class classifier, and thus each vendor only needs to go through the classifier once. In addition, the transfer learning makes the vendor-specific training quicker. The computational overhead is already acceptable, and the performance can be further improved with more powerful machines and optimized algorithms. For example, the similarity comparison can be easily parallelized, as the numbers of vendors and markets increase.

Chapter 7

Case Study

Based on the detected Sybil pairs, we then perform case studies to examine the motivations and behavior patterns of Sybil vendors.

7.1 Price Differences of Sybil Vendors.

We first analyze the Sybil accounts' product prices and examine potential market manipulations. Given a "confirmed" Sybil pair, we match their products from the two accounts based on the *product type* and the listing *time*. Our consideration is that different types of products may have a different price range, and the price is likely to change over time. We set the matching time window as 1 week. For the matched product pairs (A,B), we calculate the normalized price difference as $d(A, B) = \log_e(\frac{P_A}{P_B})$, where P is the product price. A positive (negative) d indicates that product A is more (less) expensive.

Figure 7.1a shows the price difference for *inter-market* Sybil pairs. All three curves are relatively balanced around the $x = 0$ line, indicating that products from the same vendor

are within a similar price range across different markets. For a small portion of products, however, the price differences can be large (*e.g.*, $d = 3$ is equivalent to 20 times more expensive). Comparing the different markets, Evolution's price is relatively lower. This, however, is not necessarily an indication of market manipulation. Even for non-Sybil vendors, Evolution has the lowest product price (median \$74) compared to the other two markets (median \$101 and \$132).

Larger price differences are observed between *intra-market* Sybils. Figure 7.1b compares the two Sybils accounts in same markets. For a given Sybil pair, we first differentiate the bigger account (with more products) and the smaller account (with fewer products). Then we calculate $\log_e(\frac{P_A}{P_B})$ where A represents the smaller account, and B represents the bigger account. For Evolution and SilkRoad2, both curves are heavily skewed to the left, indicating that the smaller accounts tend to sell the products at a significantly cheaper price. A possible explanation is the vendor wants to attract buyers or even to perform scam using the smaller account. In contrary, the Agora line is relatively balanced.

Figure 7.1c further compares the product price of Sybil accounts with that of *other vendors* in the same market. The curves of Evolution and SilkRoad2 are skewed to the left. This suggests that regardless the bigger or smaller accounts, Sybils' product price is substantially lower than that of the rest of the market, which is an indicator of market manipulation.

7.2 Sybil Vendors that Scam Buyers

Certain vendors create multiple accounts in the *same market* just to "scam" the buyers. Sybil vendors may refuse to ship the product after receiving the payment, or switch the original product to a lower-quality one. Using a Sybil account, the vendor does not need to worry about the reputation. Based on the discussions of the buyers (in the "DarkNetMarket"

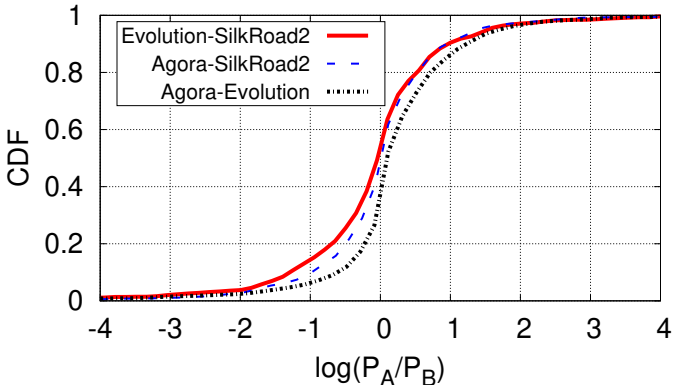
section of Reddit), we confirm that at least 3 of our detected Sybil pairs have involved in scams. For example, "Stratton" and "Montfort" are a detected Sybil pair on Agora. On Reddit, buyers reported that they were scammed by these two accounts. Some buyers even stated that the two accounts followed very similar writing styles when *replying private emails*. We also find that 86.6% (174/201) of their products have a lower price than the matched products of other vendors. This confirms our early intuition that scammers use lower prices to attract buyers.

7.3 Product Stocking and Reselling

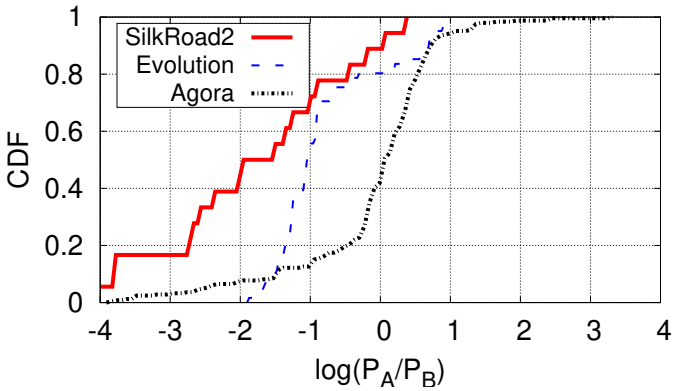
Sybils pairs that are labeled as "Probably No" are not completely useless. Even though they are not the same vendor, most of the detected accounts sell similar products. By analyzing these Sybil pairs, we reveal interesting patterns of *product stocking and reselling*. For example, our model detected two intra-market Sybil pairs on SilkRoad2: ("UGL OZ", "labsdirect") and ("OZAlpha", "labsdirect"). Manual analysis shows that vendor "UGL OZ" mainly sells steroid stored in bottles with a special label "UGL OZ". At the same time, we find the same bottles also show up in the photos of "labsdirect" and "OZAlpha". According to the comments on the vendor profile, "OZAlpha" stated that he was stocking up the products of "UGL OZ". This indicates the relationships between the darknet vendors:"UGL OZ" is the producer of those bottles of steroid, and "labsdirect" and "OZAlpha" were purchasing the products and stocking them for reselling. With the help our tool, it is possible to further automate the analysis to infer the relationships between vendors and detect the stakeholders in the market (future work).

7.4 Photo Plagiarizing

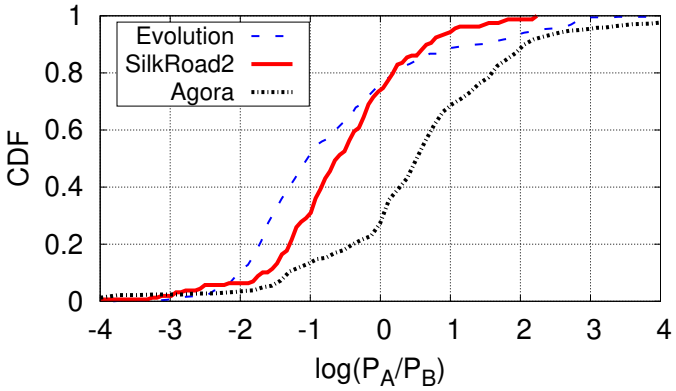
Photo plagiarizing is one of the reasons for the false positives. There are two main types. First, vendors may use the stock photos they find on the Internet. Second, vendors may "steal" photos from other vendors. The later case is more interesting to investigate further. For example, vendor "ICANTMTT2" (Agora) and "AussiesFinest" (Agora) share one identical photo of drugs. Based on the profile of "ICANTMTT2", this vendor is relatively new and his drugs were directly purchased from the drug maker. At the same time, "AussiesFinest" is a more established vendor and has many photos with the same background and layout. It looks like "AussiesFinest" is the original owner of the photo. There are several possible reasons for the photo plagiarizing. First, it is possible that "ICANTMTT2" purchased the drug from "AussiesFinest" for stocking and reselling, and thus it is reasonable to use the same photo. Second, it is also possible that "ICANTMTT2" stole the photo to make the product attractive to buyers (leveraging the established reputation of "AussiesFinest"'s drugs).



(a) Price Diff of Sybil Pairs (Different Markets)



(b) Price Diff of Sybil Pairs (Same Market)



(c) Sybil vs. Other Vendors (Same Market)

Figure 7.1: Price comparison for the same type of products around the same time (within 1 week). We compare the product prices for (a) the pairs of Sybil accounts from different markets; (b) the pairs of Sybil accounts (small account vs. big account) from the same markets; and (c), Sybil accounts vs. other vendors of the same markets.

Chapter 8

Discussion

8.1 Inter-market & Intra-market Sybils.

We identified hundreds of inter-market Sybil pairs, but only a handful of intra-market Sybils. There are two possible explanations: First, it is acceptable for a vendor to have accounts in different markets, but holding multiple accounts in the same market is usually prohibited. Due to the high anonymity of the darknet, the vendor reputation is a key factor to buyers' purchase decisions. Keeping one persistent account for each vendor helps the market administrator and buyers to assess the vendor's reputation. Second, after creating a vendor account, the vendor will need to pay several hundreds of US dollars as the "security deposit" in order to list products. The security deposit also makes it difficult for a vendor to create a large number of Sybil accounts in the same market.

8.2 Adversarial Countermoves

Our image-based fingerprinting method is not designed for adversarial settings. If a vendor wants to prevent her multiple accounts from being linked together, technically there are potential countermoves that the vendor can make. Before we discuss the adversarial countermoves, we want to stress that there are no real motivations for vendors to hide their multiple accounts in different markets. The only case where vendors may be motivated to hide their Sybil identifies is when they create Sybils in *the same market*. Intra-market Sybils are prohibited by the market administrators who actively seek to detect Sybil accounts.

To examine the impact of potential countermoves from vendors, we consider a number of image transformations. More specifically, to avoid detection, a vendor may slightly transform the photos (to hide personal styles) before posting them via the Sybil account. Here, we consider 3 simple transformations including *blurring the image*, *reducing the contrast*, and *adding random noises*. For simplicity, we apply Gaussian smoothing with $\sigma = 2$ for image blurring, we adjust the image contrasts to 50% of the original ones, and we add noise by randomly picking 5% of the pixels and changing them to black or white. Figure 8.1 shows an example.

Duplicated Images	Model	Original	Blur	Contrast	Noise
Yes	ResNet	0.979	0.960	0.969	**0.485**
	VGG	0.977	0.967	0.979	0.771
No	ResNet	0.883	0.715	0.803	**0.285**
	VGG	0.832	0.752	0.818	**0.394**

Table 8.1: Impact of adversarial image transformations to the classifier accuracy.

We run a quick test on the impact of the above adversarial countermoves using the Agora dataset with $T_r = 20$. We follow the same ground-truth evaluation workflow as §4.2, but apply image transformation to the testing images. The results are shown in Table 8.1. We

(a) Original (b) Blur (c) Contrast (d) Noise

Figure 8.1: Illustrating the image transformations.

observe that blurring and contrast adjustment only slightly decrease the matching accuracy. However, adding random noise points can greatly reduce the accuracy. With just 5% noise pixels, the products are still clearly recognizable in the images. Beyond adding random noises, vendors can also apply stronger *adversarial noises* that are optimized against the DNN based classifier [8, 8, 18, 31, 37, 41, 50, 52]. At the same time, defenders (in this case, the market administrators) can adopt defense techniques to "de-noise" the images to reduce the adversarial effect [3, 19, 36, 60] or enhance the robustness of the image classifier [18, 31, 44, 63]. Another way of defense is to set a smaller similarity threshold to include more candidate pairs for investigation.

In addition to adversarial image transformation, vendors can also sell different products using different accounts or change their photo style. Again, this type of adversarial behavior is only relevant to certain intra-market Sybils, but not the vast majority of the inter-market Sybil accounts. Our future work will measure the adversarial adaptations of vendors in the wild.

8.3 Limitations

Our study has a few limitations. First, our study only covers three darknet markets, and there are many other markets out there [5]. Our future work will explore to apply our tool to the more recent and a broader range of darknet markets. Second, although no evidence suggests that Sybil vendors are attempting to avoid detection by changing their photos, adversarial machine learning should be further explored to improve the robustness of the analysis. Third, during our manual inspection, we find additional features that can be used to match two accounts (*e.g.*, username, image trademarks, shipping information), which can be integrated to build a more powerful analysis tool.

Chapter 9

Related Work

9.1 Cybercrimes and Blackmarkets

Researchers have studied the darknet markets [11, 48] and underground forum [1, 23, 35] from various aspects. Some researchers use the underground forums to study specific cybercriminal activities such as pharmaceutical affiliate programs [35], large spam operations [49], trading stolen credit cards [20] and search engine optimization services [14]. Other researchers study the products sold on the blackmarkets [23], build automated tools to identify forum posts related transactions [42], and analyze the network footprints of underground vendors [51]. Recent works also have looked into the "social networks" and the communities among cybercriminals [15, 38, 61]. In this thesis, we develop a novel system to link Sybil identities through image analysis to support more efficient investigations of cybercrimes.

9.2 Stylometry Analysis

Stylometry analysis has been a useful tool to attribute authorship of anonymous online posts [13, 42]. The most related work to us is to use stylometry analysis to link Sybil accounts in underground forums [2, 6, 17, 22]. In this thesis, we show that stylometry analysis is less effective to model darknet market vendors due to the short and repetitive text. In comparison, our image-based approach achieved more promising results.

9.3 Image Analysis using Deep Neural Networks

Deep neural networks have contributed to the fast development of computer vision in recent years. Deep learning algorithms [30, 45] now reach the human-level accuracy in recognizing objects from images. Deep learning algorithms can take advantage of the massive training data to build highly accurate models. For many deep learning applications, transfer learning can be applied when the application-specific training dataset is insufficiently large [40, 46, 54].

A related body of work is photographer identification based on photos [9, 10, 26, 34, 43, 55, 56] or egocentric videos [24]. However, recent results show that lower-level features are not as effective as high-level features in photograph authorship attribution tasks [56]. Existing high-level feature based methods focus on several famous photographers who already have strong personal styles [56]. In contrast, we model a much larger population of darknet vendors who are typically not professional photographers.

Chapter 10

Conclusions

In this thesis, we demonstrate the feasibility of fingerprinting darknet vendors through their posted photographs. By evaluating the proposed system on real-world datasets, we demonstrate its advantage over existing stylometry methods in terms of both the accuracy and the coverage of fingerprintable vendors. In addition, we use the system to detect previously unknown Sybil account pairs in the wild, both within the same markets and across different markets. As a future work, we will continue to monitor the darknet markets and measure the potential adversarial evasions from vendors.

Bibliography

[1] Sadia Afroz, Vaibhav Garg, Damon McCoy, and Rachel Greenstadt. Honor among thieves: A common's analysis of cybercrime economies. In *Proc. of eCrime'13*, 2013.

[2] Sadia Afroz, Aylin Caliskan-Islam, Ariel Stolerman, Rachel Greenstadt, and Damon McCoy. Doppelganger finder: Taking stylometry to the underground. In *Proc. of IEEE SP'14*, 2014.

[3] Arjun Nitin Bhagoji, Daniel Cullina, and Prateek Mittal. Dimensionality reduction as a defense against evasion attacks on machine learning classifiers. *arXiv preprint arXiv:1704.02654*, 2017.

[4] Danny Bradbury. Silk road 2 loses over $2.6 million in bitcoins in alleged hack, 2014. URL https://www.coindesk.com/silk-road-2-loses-bitcoins-hack.

[5] Gwern Branwen, Nicolas Christin, David Dcary-Htu, Rasmus Munksgaard Andersen, StExo, El Presidente, Anonymous, Daryl Lau, Sohhlz, Delyan Kratunov, Vince Cakic, Van Buskirk, Whom, Michael McKenna, and Sigi Goode. Dark net market archives, 2011-2015, 2015. https://www.gwern.net/DNM-archives.

[6] Aylin Caliskan-Islam, Richard Harang, Andrew Liu, Arvind Narayanan, Clare Voss, Fabian Yamaguchi, and Rachel Greenstadt. De-anonymizing programmers via code stylometry. In *Proc. of USENIX Security'15*, 2015.

[7] Alfredo Canziani, Adam Paszke, and Eugenio Culurciello. An analysis of deep neural network models for practical applications. *arXiv preprint arXiv:1605.07678*, 2016.

[8] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *Proc. of IEEE SP'17*, 2017.

[9] M. Chen, J. Fridrich, M. Goljan, and J. Lukas. Determining image origin and integrity using sensor noise. *IEEE Transactions on Information Forensics and Security*, 3(1): 74–90, 2008.

[10] Kai San Choi, Edmund Y. Lam, and Kenneth K. Y. Wong. Source camera identification using footprints from lens aberration. In *Proc. of SPIE Digital Photography II*, 2006.

[11] Nicolas Christin. Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proc. of WWW'13*, 2013.

[12] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proc. of USENIX Security'04*, 2004.

[13] Greg Durrett, Jonathan K. Kummerfeld, Taylor Berg-Kirkpatrick, Rebecca S. Portnoff, Sadia Afroz, Damon McCoy, Kirill Levchenko, and Vern Paxson. Identifying products in online cybercrime marketplaces: A dataset for fine-grained domain adaptation. In *Proc. of EMNLP'17*, 2017.

[14] S. Farooqi, G. Jourjon, M. Ikram, M. A. Kaafar, E. De Cristofaro, Z. Shafiq, A. Friedman, and F. Zaffar. Characterizing key stakeholders in an online black-hat marketplace. In *Proc. of eCrime'17*, 2017.

[15] Vaibhav Garg, Sadia Afroz, Rebekah Overdorf, and Rachel Greenstadt. Computer-supported cooperative crime. In *Proc. of FC'15*, 2015.

[16] Stamatios Georgoulis, Konstantinos Rematas, Tobias Ritschel, Mario Fritz, Tinne Tuytelaars, and Luc Van Gool. What is around the camera? In *Proc. of ICCV'17*, 2017.

[17] Oana Goga, Howard Lei, Sree Hari Krishnan Parthasarathi, Gerald Friedland, Robin Sommer, and Renata Teixeira. Exploiting innocuous activity for correlating users across sites. In *Proc. of WWW'13*, 2013.

[18] Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *Proc. of ICLR'15*, 2015.

[19] Chuan Guo, Mayank Rana, Moustapha Cisse, and Laurens van der Maaten. Countering adversarial images using input transformations. *arXiv preprint arXiv:1711.00117*, 2017.

[20] Andreas Haslebacher, Jeremiah Onaolapo, and Gianluca Stringhini. All your cards are belong to us: Understanding online carding forums. In *Proc. of eCrime'17*, 2016.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. of CVPR'16*, 2016.

[22] Thanh Nghia Ho and Wee Keong Ng. Application of stylometry to darkweb forum user identification. In *Proc. of ICICS'16*. 2016.

[23] Thomas J Holt and Eric Lampke. Exploring stolen data markets online: products and market forces. *Criminal Justice Studies*, 23(1):33–50, 2010.

[24] Yedid Hoshen and Shmuel Peleg. An egocentric look at video photographer identity. In *Proc. of CVPR'16*, 2016.

[25] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*, 2016.

[26] C. R. Johnson, E. Hendriks, I. J. Berezhnoy, E. Brevdo, S. M. Hughes, I. Daubechies, J. Li, E. Postma, and J. Z. Wang. Image processing for artist identification. *IEEE Signal Processing Magazine*, 25(4):37–48, 2008.

[27] Rasoul Kaljahi, Jennifer Foster, Johann Roturier, Corentin Ribeyre, Teresa Lynn, and Joseph Le Roux. Foreebank: Syntactic analysis of customer support forums. In *Proc. of EMNLP'15*, 2015.

[28] Su Nam Kim, Li Wang, and Timothy Baldwin. Tagging and linking web forum posts. In *Proc. of CoNLL'10*, 2010.

[29] Frederik Kratzert. Finetune alexnet with tensorflow, 2016. URL https://github.com/kratzert/finetune_alexnet_with_tensorflow.

[30] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Proc. of NIPS'12*, 2012.

[31] Alexey Kurakin, Ian J. Goodfellow, and Samy Bengio. Adversarial examples in the physical world. In *Proc. of ICLR workshop*, 2017.

[32] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553): 436–444, 2015.

[33] Marco Lui and Timothy Baldwin. Classifying user forum participants: Separating the gurus from the hacks, and other tales of the internet. In *Australasian Language Technology Association Workshop 2010*, 2010.

[34] J. Lukas, J. Fridrich, and M. Goljan. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 1(2):205–214, 2006.

[35] Damon McCoy, Andreas Pitsillidis, Jordan Grant, Nicholas Weaver, Christian Kreibich, Brian Krebs, Geoffrey Voelker, Stefan Savage, and Kirill Levchenko. Pharmaleaks:

Understanding the business of online pharmaceutical affiliate programs. In *Proc. of UNENIX Security'12*, 2012.

[36] Dongyu Meng and Hao Chen. Magnet: a two-pronged defense against adversarial examples. In *Proc. of CCS'17*, 2017.

[37] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. Deepfool: A simple and accurate method to fool deep neural networks. In *Proc. of CVPR'16*, 2016.

[38] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M Voelker. An analysis of underground forums. In *Proc. of IMC'11*, 2011.

[39] NLTK. Natural language toolkit, 2017. http://www.nltk.org/.

[40] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *Proc. of CVPR'14*, 2014.

[41] Nicolas Papernot, Patrick D. McDaniel, Somesh Jha, Matt Fredrikson, Z. Berkay Celik, and Ananthram Swami. The limitations of deep learning in adversarial settings. In *Proc. of Euro SP'16*, 2016.

[42] Rebecca S Portnoff, Sadia Afroz, Greg Durrett, Jonathan K Kummerfeld, Taylor Berg-Kirkpatrick, Damon McCoy, Kirill Levchenko, and Vern Paxson. Tools for automated analysis of cybercriminal markets. In *Proc. of WWW'17*, 2017.

[43] Tong Qiao, Florent Retraint, Rmi Cogranne, and Thanh Hai Thai. Individual camera device identification from jpeg images. *Signal Processing: Image Communication*, 52: 74 – 86, 2017.

[44] Andras Rozsa, Ethan M. Rudd, and Terrance E. Boult. Adversarial diversity and hard positive generation. In *Proc. of CVPR Workshop*, 2016.

[45] 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. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[46] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016.

[47] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[48] Kyle Soska and Nicolas Christin. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In *Proc. of USENIX Security'15*, 2015.

[49] Brett Stone-Gross, Thorsten Holz, Gianluca Stringhini, and Giovanni Vigna. The underground economy of spam: A botmaster's perspective of coordinating large-scale spam campaigns. In *Proc. of LEET'11*, 2011.

[50] Jiawei Su, Danilo Vasconcellos Vargas, and Sakurai Kouichi. One pixel attack for fooling deep neural networks. *arXiv preprint arXiv:1710.08864*, 2017.

[51] Srikanth Sundaresan, Damon McCoy, Sadia Afroz, and Vern Paxson. Profiling underground merchants based on network behavior. In *Proc. of eCrime'16*, 2016.

[52] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In *Proc. of ICLR'14*, 2014.

[53] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proc. of AAAI'17*, 2017.

[54] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312, 2016.

[55] Thanh Hai Thai, Florent Retraint, and Rémi Cogranne. Camera model identification based on the generalized noise model in natural images. *Digit. Signal Process.*, 48(C): 285–297, 2016.

[56] Christopher Thomas and Adriana Kovashka. Seeing behind the camera: Identifying the authorship of a photograph. In *Proc. of CVPR'16*, 2016.

[57] Kristina Toutanova, Dan Klein, Christopher D Manning, and Yoram Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proc. of NAACL'03*, 2003.

[58] Xiangwen Wang, Peng Peng, Chun Wang, and Gang Wang. You are your photographs: Detecting multiple identities of vendors in the darknet marketplaces. In *Proc. of ASIACCS'18*, 2018. doi: 10.1145/3196494.3196529.

[59] Nicky Woolf. Bitcoin "exit scam": deep-web market operators disappear with $12m, 2015. URL https://www.theguardian.com/technology/2015/mar/18/bitcoin-deep-web-evolution-exit-scam-12-million-dollars.

[60] Weilin Xu, David Evans, and Yanjun Qi. Feature squeezing: Detecting adversarial examples in deep neural networks. *arXiv preprint arXiv:1704.01155*, 2017.

[61] M. Yip, N. Shadbolt, and C. Webber. Structural analysis of online criminal social networks. In *Proc. of ISI'12*, 2012.

[62] Felix Yu. Fine-tune convolutional neural network in keras with imagenet pretrained models, 2016. URL https://github.com/flyyufelix/cnn_finetune.

[63] Stephan Zheng, Yang Song, Thomas Leung, and Ian J. Goodfellow. Improving the robustness of deep neural networks via stability training. In *Proc. of CVPR'16*, 2016.