DARK VENDOR PROFILING

A Thesis

Presented to

the Faculty of the College of Graduate Studies

Tennessee Technological University

by

Susan Jeziorowski

In Partial Fulfillment

of the Requirements of the Degree

MASTER OF SCIENCE

Computer Science

May 2020

TABLE OF CONTENTS

Abstract vii

Certificate of Approval viii

Dedication ix

Acknowledgments x

List of Figures xi

List of Tables xiii

Chapter

1 Introduction 1

1.1 Thesis Motivation 2

Attribution Tasks 3

Contributions 3

Thesis Organization 4

2 Background 5

2.1 The Dark Web 5

2.2 The Onion Router 8

Hidden Services 11

2.3 Dark Marketplaces 13

2.4 Dark Web Investigations 15

2.5 Ethical Considerations 17

3 Related Work 19

3.1 User Attribution 20

3.2 Alias Attribution 22

3.3 Discussion of Limitations 26

4 The Dark Vendor Profiling Framework 33

4.1 System Overview 33
4.2 Key Contributions 35
Application of Image Hashing in Dark Marketplaces 36
Machine Learning Based Profiling 40
4.3 Dark Vendor Profiling Workflow 44
5 Data 47
5.1 Darknet Market Archives 47
5.2 DVP DB - The Dark Vendor Profiling Database 48
Information Collected 50
Data Cleansing 50
Versions 53
6 Methods and Design 57
6.1 Feature Engineering 57
Stylometry Based 58
Attribute Based 59
Image Based 61
Excluded Data 62
6.2 DVP Tasks 63
7 Dark Vendor Profiling Evaluation 65
7.1 Vendor Attribution 65
Random Forest Hyperparameter Tuning 68
Model Accuracy 69
Time Complexity and Memory Usage 77
Feature Subsets 80
7.2 Alias Attribution 82
Elementary Style Results 85

Enhanced Vendor Linkage Results 87

8 Future Work 93

9 Conclusion 99

References 101

Vita 107

ABSTRACT OF A THESIS

DARK VENDOR PROFILING

Susan Jeziorowski

Master of Science in Computer Science

Tor hidden services and anonymity tools alike provide an avenue for cyber criminals to conduct illegal activities online without fear of consequences. In particular, dark marketplaces are hidden services that enable the trade of paraphernalia such as drugs, weapons, malware, counterfeit identities, and pornography among other items of criminal nature. Several effective Dark Web analysis techniques have been proposed for Dark Web Forums and primarily focus on authorship analysis where the goal is one of two tasks: (a) user attribution, where a user is profiled and identified given an artifact they own, and (b) alias attribution, where pairs of users are identified to belong to the same individual. While these techniques may support dark web investigations and help to identify and locate perpetrators, existing automated techniques are predominately forum-based and stylometry-based, leaving non-textual artifacts, such as images, out of consideration due to the illicit nature of dark marketplace listings. Thus, new methodologies for adequate evidence collection and image handling in dark marketplaces are essential. In this thesis, stylometric, image, and attribute-based artifacts are collected from 25 dark marketplaces and machine learning based Dark Vendor Profiling methodologies are proposed to achieve dark vendor attribution and alias attribution across dark marketplaces, thereby supporting investigative efforts in deanonymizing cyber criminals acting on the anonymous web. Namely, we first propose the collection of image hashes in place of image content to reduce the storage demands of our proposed technique and reduce the risk of obtaining illicit digital material during data collection. Second, we design two unique feature sets for authorship analysis tasks that are extracted *per listing* and *per vendor*. Third, we propose a novel application of the Random Forest machine learning technique for the task of vendor attribution in dark marketplaces, achieving over 90% accuracy in distinguishing between over 2,500 unique dark vendors from various marketplaces. Lastly, we propose a novel application of the Record Linkage technique for the task of alias attribution and obtain imperative preliminary observations from Support Vector Machine and Logistic Regression based models that can assist in the design of future alias attribution models. Therefore, this thesis presents a detailed description of these contributions along with an evaluation of our proposed Dark Vendor Profiling system and several future research directions.

APPROVAL CERTIFICATE OF A THESIS

DARK VENDOR PROFILING

by
Susan Jeziorowski

Graduate Advisory Committee:

Ambareen Siraj, Chair	Date
Muhammad Ismail	Date
Stacy Prowell	Date

Approved for the Faculty:

Mark Stephens, Dean College of Graduate Studies	Date

DEDICATION

To my father, who has dedicated his life to enriching the lives of his daughters,
and my mother, who exemplified academic excellence and nurtured us.
Thank you for teaching me the value of hard work,
for leading me to discover my passion in computer science,
for inspiring me to set my aspirations high,
and for raising me to be the woman
I am today.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Ambareen Siraj, who took me under her wing early in my college career and inspired me to pursue my master's degree with a specialization in cybersecurity. Without her continuous encouragement, I could have never imagined completing my graduate studies, attending and presenting at several professional conferences, competing in national cyber competitions, leading clubs and organizations for Tennessee Tech students, and publishing academic research papers. She is an inspiration to me and to all women in engineering fields for her dedication to making a difference in our world and paying it forward. Dr. Siraj, thank you for the invaluable opportunities you have provided me. Someday, I hope I may pave the roads for future young women just like you had paved the roads for me. Secondly, I would like to thank my committee members along with the faculty of the Computer Science department and the Cybersecurity Education, Research & Outreach Center (CEROC) for their continuous support and guidance throughout my college career. I have an immense amount of gratitude and respect for the effort and dedication Tennessee Tech's faculty shows their students and could not feel more proud to have been a part of their program. Lastly, I would like to thank the CyberCorps Scholarship for Service federal program for granting me the means to complete my graduate degree. This program provides incredible opportunities to students across the nation, enabling us to enhance our talents in technology and discover our passions for public service. I hope this research may serve as an example of the benefits of federal scholarship and encourage continuous funding of similar programs.

LIST OF FIGURES

Figure **Page**

2.1 A popular graphic for visualizing the depth of the Surface, Deep, and dark web [50]. 6

2.2 Graphic representation of a typical Tor *circuit*, consisting of an entry guard, middle relay, and exit relay that facilitate the flow of data through the Tor cloud [21]. 10

2.3 The number of unique onion addresses running on the Tor network on any given day between February 2019 and February 2020 estimated by the Tor Project Metrics Portal [2]. 13

2.4 Screenshot of Quality King prescription pill listings in November 2019. 14

2.5 Screenshot of hacking service listings on the dark web from November 2019. 15

2.6 Screenshot of the most recent version of the Silk Road taken in November 2019. 16

3.1 A summarizing timeline of dark web authorship analysis research progression between 2014 and 2019. 20

3.2 A comparison of the proposed Dark Vendor Profiling technique to related work in order of most similar to least similar, including uIdentifier [60], Photo-Based Deep Neural Networks [58], Hidden Relationships in Social Network Analysis [35], and Stylometric Text-Based Studies [5, 16, 19, 24, 51]. 30

4.1 A comparison of several popular machine learning techniques based on desirable characteristics of dark marketplace classification models. 41

4.2 An illustration of the DVP system workflow. 45

7.1 Model Accuracy vs. Number of Listings per Vendor results for DVP DB V4 single marketplace experimentation. 71

7.2 Model Accuracy vs. Number of Listings per Vendor results for DVP DB V5 single marketplace experimentation. 72

7.3 Model Accuracy vs. Number of Vendor results for DVP DB V4 and V5 multi-market experimentation. 74

Figure **Page**

7.4 Model Accuracy vs. Number of Listings per Vendor results for DVP DB V4 and V5 multi-market experimentation. 75

7.5 Model Accuracy vs. Number of Listings per Vendor results for *full* DVP DB V4 and V5 multi-market experimentation. 76

7.6 Time vs. Model Complexity results for DVP DB V4 multi-market experimentation. 78

7.7 Average Memory Usage vs. Model Complexity results for DVP DB V4 multi-market experimentation. 79

8.1 System architecture of a fully automated DVP implementation including a newly developed Dark Web Crawler (DWC). 95

LIST OF TABLES

Table		Page
4.1	Dark marketplaces listed in order of significance to DVP by calculating an average rank over seven characteristics: number of listing entries, number of vendors, number of images, number of images with metadata, number of images with GPS coordinate data, proportion of images with metadata, and proportion of images with GPS coordinate data.	39
4.2	Hash analysis results for average, difference, perceptual, and wavelet hashing in order of weighted average SSIM values.	40
5.1	Basic DVP DB Schema with five tables.	49
5.2	Data availability per dark marketplace from the DMN Archives. Each X represents whether or not the information is regularly available in a marketplace.	51
5.3	Summary of deleted DVP DB entries due to missing integral data.	52
5.4	Summary of deleted DVP DB entries due to suggestion of sex as a service.	53
5.5	Summary DVP DB versions and their contents.	55
6.1	Stylometric features extracted for DVP. Note, vendor description features are considered only in alias attribution, whereas product name and product description are considered for both vendor and alias attribution.	60
6.2	Attribute-based features extracted for DVP.	61
6.3	Image-based features extracted for DVP.	62
7.1	Summary of experiments ran to evaluate DVP-based vendor attribution in individual marketplaces.	66
7.2	Summary of experiments ran to evaluate DVP-based vendor attribution across multiple marketplaces ranging from to 10 to 25 marketplaces.	67
7.3	Description of hyperparameters tested during single marketpace experimentation.	68
7.4	Results of hyperparameter tuning in single-market-based models.	68
7.5	Overall model accuracy results of single marketplace experimentation.	70

Table **Page**

7.6 Correlations between single market model *accuracy* and three variables: (a) number of vendors, (b) number of listings per vendor, and (c) model complexity. 70

7.7 Overall model accuracy results of multi-marketplace experimentation. 72

7.8 Correlations between multi-market model accuracy and three variables: (a) number of vendors, (b) number of listings per vendor, and (c) model complexity. 73

7.9 Summary of full DVP experiments ran to evaluate multi-market vendor attribution. 75

7.10 Correlations between multi-market feature extraction and model execution *time* and (a) number of vendors, (b) number of listings per vendor, and (c) model complexity. 77

7.11 Correlations between multi-market feature extraction and model execution *average memory usage* and (a)number of vendors, (b) number of listings per vendor, and (c) model complexity. 79

7.12 Full multi-market DVP results in terms of time complexity (H:M:S) and memory usage (GB). 80

7.13 Summary of feature subset experiments ran on models trained with datasets containing up to 20 vendors per marketplace with 25 listings each. 81

7.14 Summary of experiments ran to evaluate DVP-based alias attribution. **Incomplete experiments are awaiting results due to the time required to analyze large numbers of potential alias pairs.* 84

7.15 Example version 5 single market alias pairs with at least 90% similarity found using DVP-based alias attribution and *any* number of listings per vendor. 86

7.16 Example version 4 multi-market alias pairs with at least 90% similarity found using DVP-based alias attribution and 75 listings per vendor. 86

7.17 Example version 4 multi-market alias pairs with the same names but only 50-60% similarity found using DVP-based alias attribution and 75 listings per vendor. 87

Table **Page**

7.18 Results of SVM and LR training *with* Vendor Name as a feature in terms of True Positives (TP), False Postivies (FP), False Negatives (FN), True Negatives (TN), Precision, Recall, F1 Score, and Accuracy. 89

7.19 Results of SVM and LR training *without* Vendor Name as a feature in terms of True Positives (TP), False Postivies (FP), False Negatives (FN), True Negatives (TN), Precision, Recall, F1 Score, and Accuracy. 91

CHAPTER 1

INTRODUCTION

Anonymity tools, i.e., tools for concealing a web user's online identity and activities, have grown increasingly popular. While these tools promote our human right to privacy, they also provide an avenue for cyber criminals to conduct illegal activities online without fear of consequences. This illegal conduct is specifically enabled through the *Dark Web*, a subsection of the Internet only accessible through the use of an anonymous network such as Tor, the Onion Router [14]. While the intent of anonymous networks is pure, these Internet technologies have been misused by cybercriminals and now serve as obstacles for law enforcement and intelligence agencies attempting to apprehend these criminals.

Dark marketplaces are one example of how anonymity-preserving technologies have enabled the migration of organized crime to the Internet where national borders are eradicated and the consumer base is world wide. By using pseudo-names, decentralized cryptocurrencies, encrypted communications and special means to hide the host location of web services, these dark marketplaces are able to facilitate the trade of illegal goods and services. Thus, it has now become the task of investigators to circumvent these anonymity techniques so the vendors of these dark marketplaces may be identified, their businesses can be shut down, and they can be prosecuted.

An integral part to any criminal investigation is the collection of evidence since the success of prosecution heavily relies on the evidence presented in court. Similarly, the success of the deanonimization of an online identity relies on gathering as much data on the identity as possible. Although the dark web is based upon preserving anonymity, dark web users may unconsciously leave traces of evidence behind regarding their activities whether it be in the text they write, the media they post, or the others they interact with. Just as regular criminal investigations

involve collecting evidence to build cases, the task of deanonimizing a dark vendor can be thought of as collecting and analyzing data to build vendor *profiles* where each profile is unique and distinguishable from other dark vendor profiles and can aid in vendor identification. In this thesis, we propose *Dark Vendor Profiling*: a new method to automate the data collection and profiling of dark vendors which can support investigative efforts to deanonymize their identities.

1.1 Thesis Motivation

Dark Vendor Profiling is defined as the task of collecting evidence on a dark vendor using only the publicly available data they generate in the dark marketplaces while conducting business. Currently, this investigative effort is predominately executed manually which limits the number of vendors we are able to investigate, profile, and report. While a few works attempt to automate investigative work, it is reported that the amount of interest in dark marketplaces is disproportionate to the impact they have on illegal trade [33]. Moreover, most existing work exhibits several limitations in scalability and performance. These limitations will be further discussed in Chapter 3. Thus, the analysis of dark marketplaces is an important field of research to the investigative community. We believe the application of data mining and machine learning techniques may greatly enhance the process of profiling dark vendors and may lead to more effective analysis of dark marketplaces. Thus, we propose this thesis to aid investigative efforts against cybercriminals acting on the dark web.

Attribution Tasks

Our research intends to support investigators by achieving the primary task of *vendor attribution* and the secondary task of *alias attribution.* With vendor attribution, we first examine the effectiveness of collecting several data from vendor listings (such as product name, description, shipping information, and image) and utilizing this data to build a machine learning classification model that may identify a dark vendor given a product listing. By demonstrating the ability to distinguish between dark vendors and produce profiles from their listing information, we show how vendors leave identifiable data in anonymous environments and how this may be used by investigators to aid criminal investigations in the dark web. With alias attribution, we examine additional data from vendor profile pages and utilize it to further develop vendor profiles. Using these enhanced profiles, we propose a method for determining vendor aliases, i.e., finding pairs of vendor profiles that likely point to the same individual. By doing so, we hope to better understand the dark web cybercriminal ecosystem and automate the collection and analysis of dark marketplace data.

Contributions

The contributions of this thesis will be further detailed in Chapter 4. In short, this thesis presents the following:

1. A new method for collecting image-based data without the need for downloading dark web imagery.

2. Two unique feature sets derived from a combination of product listings and vendor profile pages to be used for vendor and alias attribution tasks in dark vendor profiling.

3. A novel application of the Random Forest algorithm, a classic machine learning classification model, for the task of vendor attribution.

4. A novel application of Record Linkage for the task of alias attribution.

5. A novel scheme for Dark Vendor Profiling.

Thesis Organization

The thesis is organized into nine chapters. In the first, we have established the motivation behind our work and briefly presented our research objectives. In the following chapter, we provide the necessary context needed to understand the technologies that enable the dark web and why dark marketplaces are challenging to investigate. In Chapter 3 we discuss several related works and determine their main limitations. Then, Chapter 4 introduces our proposed Dark Vendor Profiling system. In Chapter 5 we introduce our data sources followed by Chapter 6, in which we present the features we extract from our data sources to develop a classification model. Chapter 7 discusses our experimental design and provides an evaluation of our proposed system. Finally, the limitations of this work are presented with future research directions in Chapter 8, and the thesis concludes in Chapter 9.

CHAPTER 2
BACKGROUND

To better understand the challenges in investigating cybercriminals acting under the protection of anonymous networks, this chapter will introduce key concepts related to the dark web research domain. First, a formal definition of the dark web and distinction from the Internet will be made in Section 2.1, followed by an introduction to anonymous network technologies and the criminal online marketplaces they enable. Further, this chapter will describe typical dark web investigation strategies to demonstrate the importance of pursuing automated means of cybercriminal profiling and conclude with a discussion of the ethical concerns taken into consideration throughout the duration this study.

2.1 The Dark Web

The dark web is a complex, generally misunderstood domain of the Internet. Usually, the dark web is perceived negatively due to the influence of stereotypes presented in media and a general lack of public Internet education. Thus, to encourage a well-informed discussion, it is important to first define the dark web and acknowledge both its legitimate and illegitimate uses.

To better understand the distinction between the dark web and the Internet as we know it, we may compare the depths of the Internet to the nature of an iceberg as illustrated by Figure 2.1. Specifically, the Internet consists of three main components, the *surface, deep* and *dark* web, each of which are distinguished by their degree of visibility and accessibility to the public. The topmost part of the iceberg is the part of the Internet we regularly engage with. This web access level represents the *surface* web, which houses all of the regularly accessible sites and services that are indexed

by search engines and do not require any credentials, authorization tokens, or special network connections to access. It is estimated that surface level services make up a mere 4% of the Internet, leaving the rest of the Internet to be composed of the deep and dark levels.

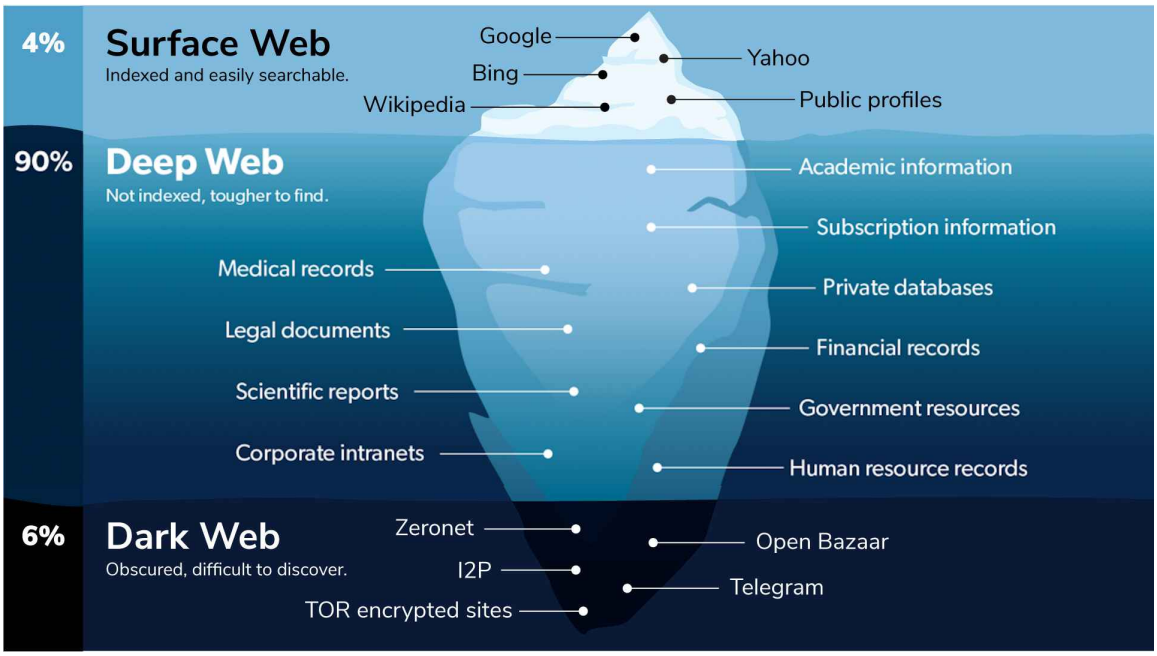


Figure 2.1: A popular graphic for visualizing the depth of the Surface, Deep, and dark web [50].

As we dive deeper into the water, we find services not indexed by search engines and instead protected behind firewalls, login credentials, CAPTCHAs, and other technologies alike. This level of the iceberg comprises of the deep web which is estimated to make up 90% of the Internet. Examples of web services residing at the deep web level include government resources, medical records, and private social media accounts among others. Generally, a user is unable to access a site at this

level without first bypassing some version of authentication since the web service is considered to lie "beneath the surface".

Finally, the *dark* web is the deepest part of the iceberg and the most difficult level of the Internet to access. This remaining 6% is home to hidden web services accessible only via special connections to anonymous networks such as The Onion Router, or **Tor** [14], which will be further discussed in Section 2.2. It is important to note that the dark web is a fascinating portion of the Internet for a wide variety of users, not just criminals and malicious users as the name may suggest. Since the Internet is a uniquely global environment, jurisdictional boundaries in anonymous networks are blurred. Thus, for the rest of this thesis, *legal* vs. *illegal* activities can be considered synonymous to the generally accepted definitions of right vs. wrong, ethical vs. unethical, or harmless vs. harmful.

Let us consider the sale and use of recreational cannabis as an example, which is generally unacceptable in the majority of the world. Therefore, using the dark web for cannabis trafficking can be considered illegal web activity, despite being jurisdictionally legal in several states and countries [59]. Other examples of illegal web services dwelling in the dark web include dark forums and marketplaces that facilitate the trafficking of paraphernalia like weapons, hacking-as-a-service, and counterfeit identities among others [9, 23]. Moreover, hidden services which exist to facilitate human trafficking and host child exploitation content are highly unethical and, therefore, considered illegal in this work despite being accepted by law in certain parts of the world [17, 18]. Therefore, this level of the Internet is known to be *dark* not only because of its limited degree of visibility but also because of the nature of the malicious activity it enables.

In contrast, the use of hidden services may also serve legitimate purposes. For example, a *legal* dark web hidden service may act as a gateway to the rest of the

world for Internet users physically located in oppressive regimes. Some governing bodies may restrict a person's freedom of expression, such as that of Venezuela where certain social media, political and social content is blocked and speaking up against one's government is punishable by law [40, 52]. Thus, utilizing anonymous network connections may indeed protect users from government surveillance and censorship by concealing web activity. Although the Venezuelan government may consider access to social media and news outlets illegal, the majority of the world generally accepts and supports the use of such web services, making them legitimate uses of anonymous networks in this research.

2.2 The Onion Router

As previously mentioned, dark web hosted services cannot be accessed without special connections to anonymous networks. In general, anonymous networks rely on peer-to-peer connections and other anonymity enhancing tools like VPNs and cryptocurrencies. Further, they are known as *overlay networks*, which means they use software solutions deployed on top of existing infrastructure, i.e. the Internet, to map virtual links between clients and services for the creation of new virtualized network infrastructures [32]. It is important to note that the connection between a user and an anonymity network is **not** hidden. However, a user's location and the content of the communications within an anonymous network remain obfuscated via multi-layered encryption and multi-hop proxies. Additionally, a user's traffic is delivered on shared bandwidth, making it even more difficult to distinguish between individual connections.

By far, the most prevalent anonymity network is The Onion Router, **Tor**, developed by The Tor Project, Inc. and initially released in 2002 [14]. When designing

Tor, the developers wanted to ensure that the anonymous network was founded on four main ideas: deployability, usability, flexibility, and simplicity. Since 2002, the Tor Project has developed into a massive nonprofit organization which supports a diverse group of developers, researchers, and privacy-seeking Internet users who make up the Tor community by offering training, outreach, and research opportunities. As advertised on the Tor Project website, these technologies advance human rights and promote online privacy by circumventing censorship, blocking trackers, defending from surveillance, resisting fingerprinting attacks, and enabling free web browsing. In fact, the Tor Project now has so much support that the network is currently comprised of over 6,500 volunteer servers that are able to support over two million directly connected users on a daily basis [2].

The basic concept of Tor involves the implementation of three main privacy preserving technologies - multi-layered encryption, volunteer proxies, and shared bandwidths - which are applied during all communications in the network. Typical Tor connections are based on *circuits* of three relay nodes: an entry node, a middle node, and an exit node as shown in Figure 2.2. When preparing a stream of data to be sent down a circuit, a user will encrypt their data three times, using each relay's public key once. As the data is passed from the entry node to the middle node and to the exit node, a layer of encryption is removed at each hop, similar to peeling layers off of an onion. Finally, once the data has reached the circuit's exit node, the data is fully decrypted and passed to the destination node. This scheme allows anonymity of the user not only by performing severely rounds of encryption but also by ensuring each node is only aware of its neighboring nodes in the circuit. In other words, no node is aware of the overall end-to-end communication and clients and services are **never** directly connected to each other. Additionally, when the Tor browser is used, little to no remnants of Internet activity can be forensically recovered

from the device. Forensic analysis may verify whether or not the Tor browser was installed on a client computer, but not if and when it was used, nor what it was used for [48].

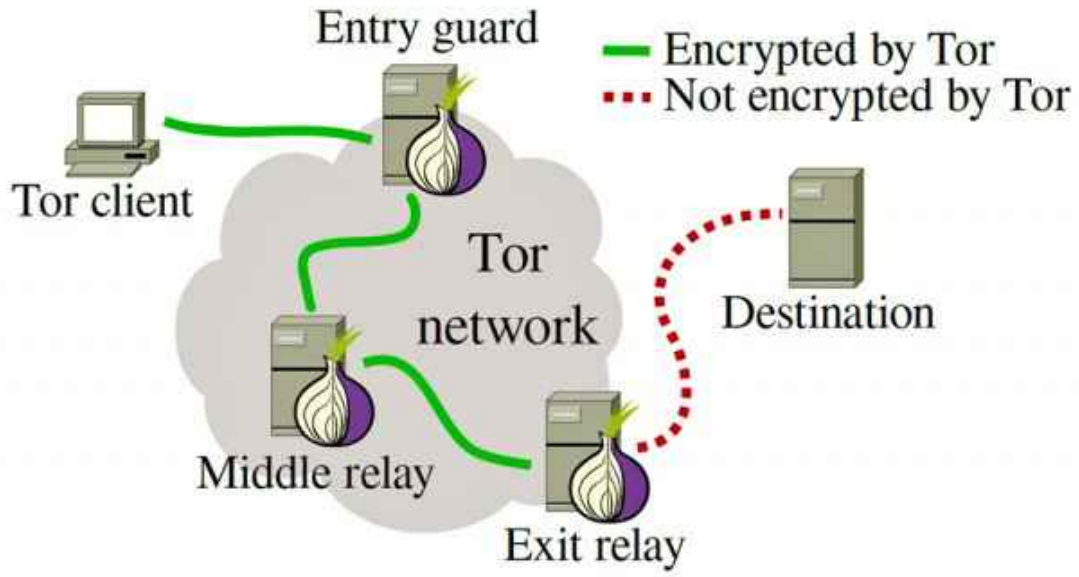


Figure 2.2: Graphic representation of a typical Tor *circuit*, consisting of an entry guard, middle relay, and exit relay that facilitate the flow of data through the Tor cloud [21].

Typically, users interact with the Tor network through local Onion Proxies, which can be downloaded in the Tor Browser Bundle that is now available for Windows, Apple, Linux, and Android devices [54]. The Tor browser is Firefox-based, so users are able to experience a familiar web user interface while enhancing their level of security. For uses who wish to access Tor services without using the browser bundle, the Tor community has developed a separate software project known as Tor2Web, which grants users access to Tor-based onion services, i.e. hidden services that are

normally only accessible via the Tor Browser Bundle on regular browsers. However, developers warn users that Tor2Web is designed to protect content providers, and not the users, thereby trading anonymity for convenience. For greater anonymity and protection, the developers highly recommend utilizing the browser bundle instead [1].

Another notable anonymous network is the Invisible Internet Project, I2P [25], which is an overlay network that incorporates encrypted, unidirectional circuits, formally known as *tunnels*, to connect its users. Thirdly, JonDonym, previously known as Java Anon Proxy [29], is an anonymity network that passes data between multiple nodes, formally known as *mixes*. Unlike Tor, the circuits, or *cascades*, incorporated in JonDonym are fixed. While both of these networks have a significant user base, we consider Tor as the sole anonymous network in this research because of its popularity with dark vendors.

Hidden Services

Tor's most distinctive feature is its ability to provide *hidden services*, each of which are hosted with *onion* addresses [14]. These services are comparable to I2P eepsites and enable users to host anonymous, theoretically untraceable websites by implementing additional security measures. Unlike typical Tor network connections, which involve one entry, one middle, and one exit node, connections to hidden services involve additional interactions with Introduction Points and Rendezvous Points. The process of connecting a user to a hidden service includes the following steps [32, 56]:

1. A hidden service server selects up to ten *introduction points* (IP) to advertise its service to. Like most connections, the hidden service and IP are separated by a three-node circuit.

2. The hidden service creates a *descriptor* containing information on how to request access via one of the services' IPs. Additionally, the service generates an onion address using a hash algorithm on its public key.

3. Both the descriptor and onion link are published to the *Distributed Hash Table* (DHT) distributed among the *Hidden Service Directories* (HSDir). Hidden Service Directories are simply trusted nodes that have been granted both the *HSDir* and *stable* flags.

4. When a user wants to connect to an onion address, they will request to download the address's associated descriptor from the DHT and establish a three-node connection to an arbitrary relay that will eventually become the *Rendezvous Point* (RP).

5. Using the descriptor data obtained from the DHT, the user will reach out to the hidden service's IP, sharing the chosen RP and a one-time secret.

6. The IP will pass the data along to the hidden service. If the hidden service approves of the connection, it will build a three-node circuit to the RP, offering the same one-time secret.

7. Once a connection is established, the RP will notify the user and serve as a relay between the two circuits. Therefore, the final connection between client and hidden service contains six relays in total.

With double-sided anonymity, both users and service providers are able to mask their identities by hiding behind their respective three-node circuits, thereby preventing either party to discover the other party's true location. For the anonymous community, this is a very attractive web hosting solution. In fact, it has been

estimated that 60,000-100,000 hidden services were running on the Tor network at any time during this past year, as shown in Figure 2.3. This statistic, along with many others, are estimated by the Tor Project [53] and accessible on their metrics portal [2].

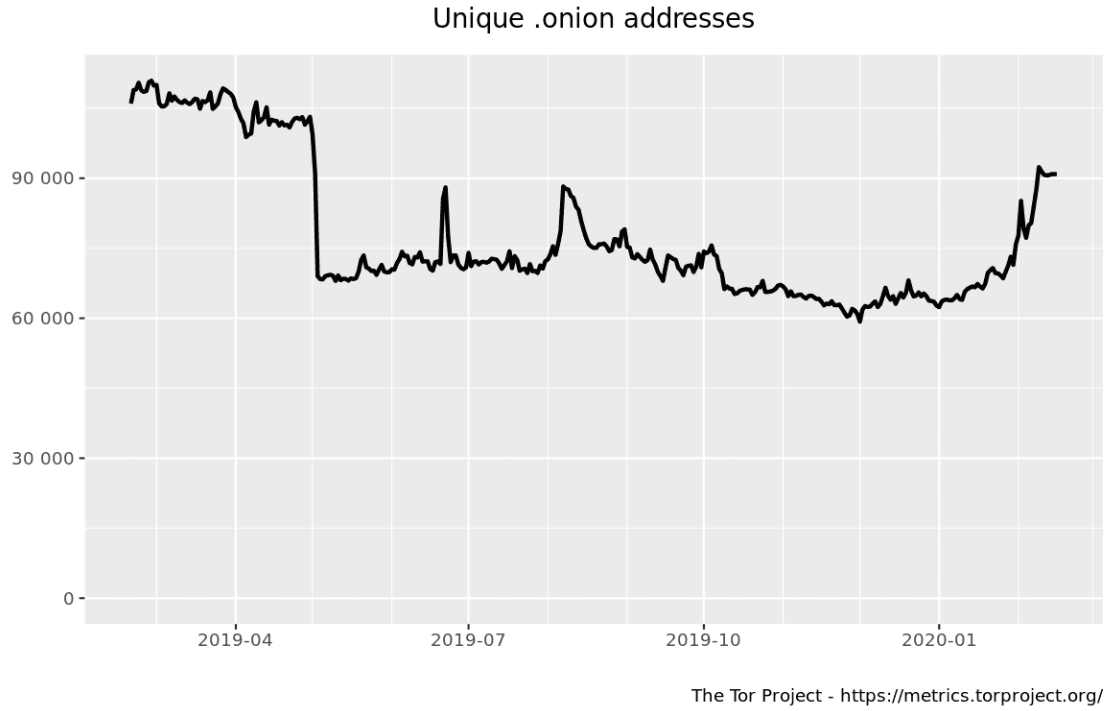


Figure 2.3: The number of unique onion addresses running on the Tor network on any given day between February 2019 and February 2020 estimated by the Tor Project Metrics Portal [2].

2.3 Dark Marketplaces

Although Tor and other popular anonymity networks promote our human right to privacy, they also provide an avenue for criminals to conduct illegal activities online without the fear of consequences. Specifically, Tor allows for the hosting of hidden

online marketplaces where *dark vendors* are able to anonymously engage in illegal trade. These hidden services are known as *dark marketplaces* and offer a wide variety of illegal goods including but not limited to marijuana, prescription pills, psychedelics, weapons, sex, hacking manuals, pirated digital content, and hacking services. For example, Figure 2.4 is an example of prescription pill listings being sold on the dark web by the dark vendor 'Quality King' in November 2019. Furthermore, Figure 2.5 illustrates how hacking services may be offered in these anonymous environments. Evidently, these listings are very criminal and therefore significant targets for law enforcement and intelligence agencies.

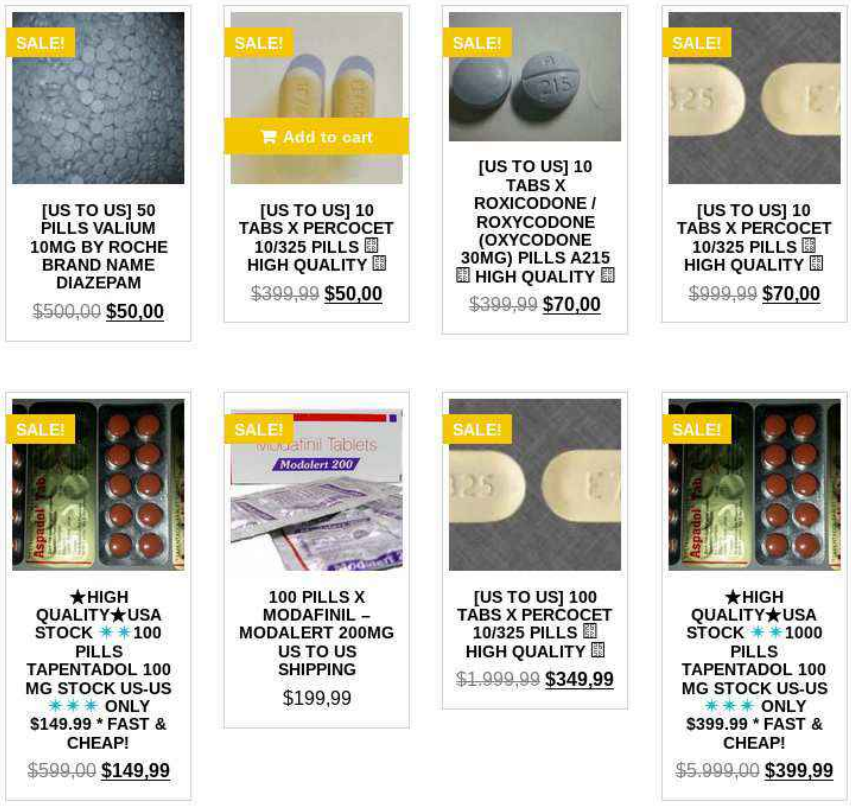
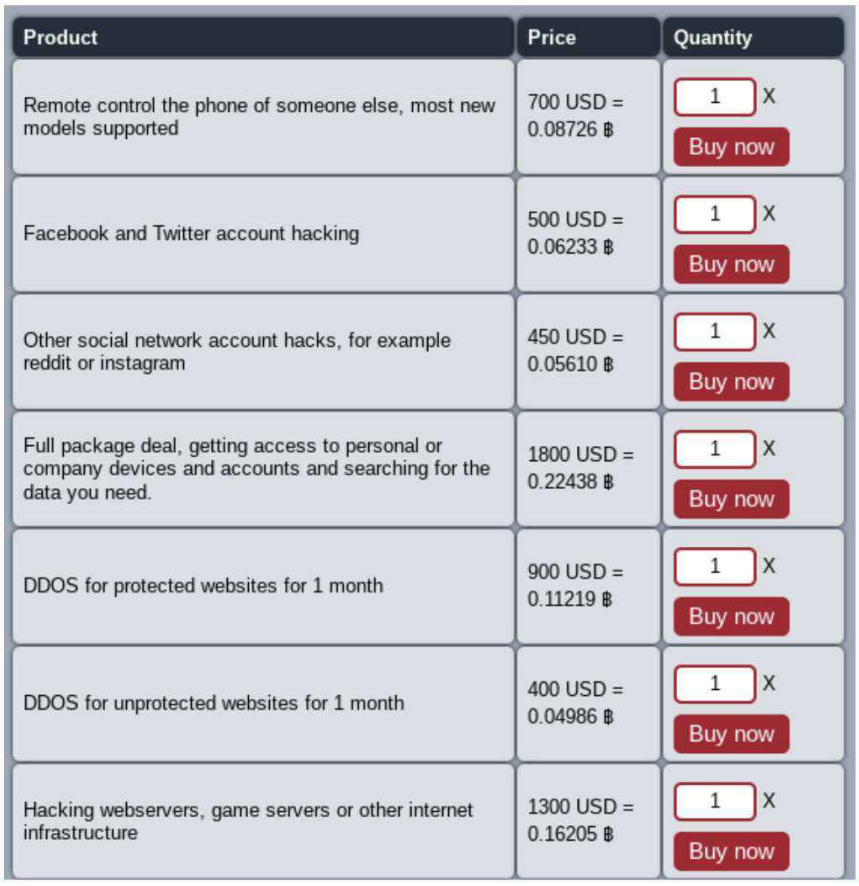


Figure 2.4: Screenshot of Quality King prescription pill listings in November 2019.



Product	Price	Quantity
Remote control the phone of someone else, most new models supported	700 USD = 0.08726 ฿	1 X Buy now
Facebook and Twitter account hacking	500 USD = 0.06233 ฿	1 X Buy now
Other social network account hacks, for example reddit or instagram	450 USD = 0.05610 ฿	1 X Buy now
Full package deal, getting access to personal or company devices and accounts and searching for the data you need.	1800 USD = 0.22438 ฿	1 X Buy now
DDOS for protected websites for 1 month	900 USD = 0.11219 ฿	1 X Buy now
DDOS for unprotected websites for 1 month	400 USD = 0.04986 ฿	1 X Buy now
Hacking webservers, game servers or other internet infrastructure	1300 USD = 0.16205 ฿	1 X Buy now

Figure 2.5: Screenshot of hacking service listings on the dark web from November 2019.

2.4 Dark Web Investigations

Due to Tor's hidden service infrastructure, owners, vendors, and users of these marketplaces are difficult to identify and locate. In fact, the location of a hidden service is *theoretically* untraceable. However, there are many cases in which law enforcement has been able to successfully locate a criminal hidden service server and prosecute the owner. One of the most famous law enforcement cases was that of the Silk Road anonymous marketplace takedown executed by the FBI and Europol in 2013

[22]. The Silk Road was a multi-million U.S dollar hidden marketplace specialized in narcotics and controlled substances. Consequently, it was one of the most popular international business avenues available through Tor. Ultimately, the takedown was enabled by *manual* investigative work accomplished by federal agents. However, despite the successful investigation, newer versions of the Silk Road became available through other hidden service operators (see Figure 2.6). In fact, now dozens dark marketplaces exist at any given time, and the ability for law enforcement to identify and locate these markets is mainly limited by the amount of manual analysis it requires. Therefore, as these marketplaces continue to grow in popularity and maliciousness, improved automated means for dark marketplace analysis will need to be established to support investigative efforts and aid in the takedown of these illegal services.

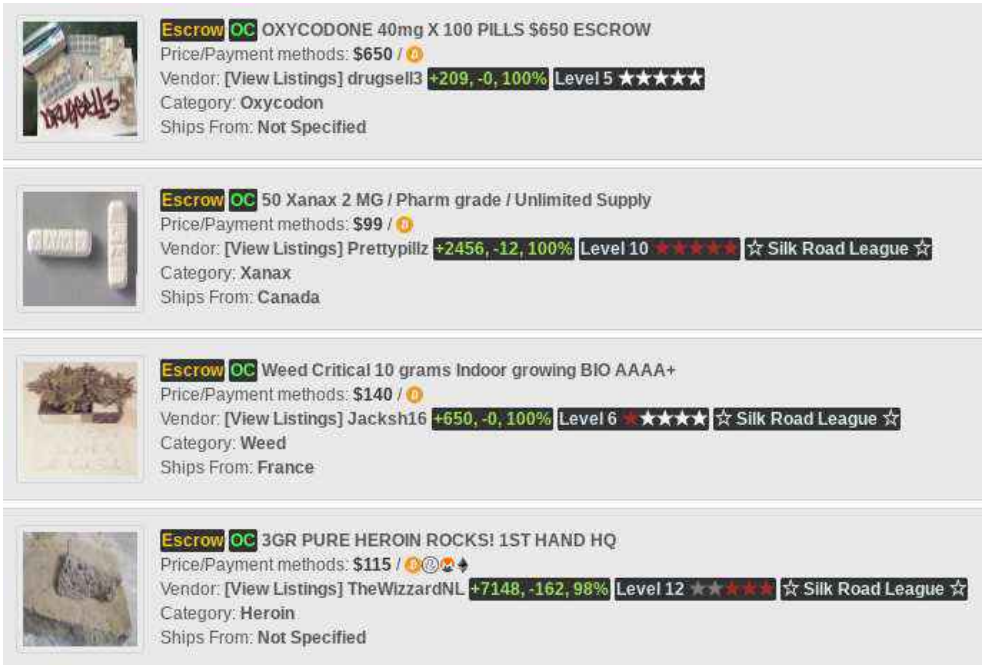


Figure 2.6: Screenshot of the most recent version of the Silk Road taken in November 2019.

2.5 Ethical Considerations

In the deanonymization research domain, data collection and analysis must be performed in an ethical matter, taking user privacy rights and safety into consideration. It is important to understand that research in cyber criminal networks may examine the behavior of people more than that of computers and networks. Hence, data gathered for research should undergo ethical review to consider the potential harms that could result from publishing the study.

Several published works offer a more in-depth discussion on ethics within this domain, especially when engaging with Tor. In particular, researchers have analyzed the ethics of conducting studies over dark web marketplaces [33] and established community standards for research in anonymous networks [49]. In [33], Martin & Christin discuss one of the main challenges of Internet based research in general since it is difficult to define jurisdictional boundaries within the scope of the global Internet. Different countries abide by different laws over paraphernalia and Internet use, thus the definition of 'illegal' business may become blurred. Especially in the dark web domain, it becomes challenging to govern dark marketplace research since the location of users and service hosting sites are hidden behind layers of encryption and proxy severs. With this in mind, Martin & Christin [33] attempt to define a universal set of Internet research standards in which it is suggested that data collection on dark marketplaces is acceptable as long as the data is public. Furthermore, they suggest that if a marketplace is of ill-intent (as determined by the eye of the beholder), any Terms of Service the site provides may be considered obsolete for both the research and intelligence communities. Similarly, authors of [49] conclude that community standards should ensure Tor based studies (a) are legal in the countries where they are performed, (b) engage in minimal user data collection and retention, (c) are

vetted by institutional review boards when available, and (d) use the Tor network to study Tor service usage exclusively, instead of using the anonymous environment as a convenient mean to study general Internet activity.

Along with considering the privacy and safety of anonymous users, it is important for researchers of this domain to practice safe browsing behavior. Evidently, the dark web can be intimidating and unpredictable. Therefore, researchers are encouraged to exercise common sense, protect their identities, avoid using personal accounts or credit cards, and avoid downloading or opening any files unless operating in special, disposable, virtual environments [45]. Furthermore, Martin & Christin [33] warn that the risk to researchers may increase after publication due to the anonymous web user base not wanting deanonimization efforts to be made against them. To combat these concerns, the Tor project organization offers research safety guidelines to protect the privacy of their user base and protect the researchers conducting studies on the user base. These guidelines are enforced via a Tor Research Safety Board which accepts research proposals, offers advice, and publishes any resulting discussions for public viewing [55].

Since this thesis considers publicly available data, does not rely on personally identifiable data provided by human subjects, performs similar experiments to those that have already been conducted and avoids pornographic content, we determine our research to meet Tor research ethical standards. Furthermore, all vendor names mentioned throughout this thesis are in pseudo name form and therefore are not considered to be personally identifiable information. Thus, the exposure of dark vendor identities used in this study is significantly limited and their anonymity is innately protected.

CHAPTER 3

RELATED WORK

This research began with an encompassing literature survey over open-ended research problems in the dark web. We conducted this survey to better understand the dark web research community, focusing on efforts to surpass the anonymity that Tor and anonymity tools alike provide. Specifically, our survey presents a new categorization of dark web deanonymization research, determining each work to be related to one of the following five domains: (1) Data Mining, (2) Classification, (3) Attribution, (4) Hidden Service Exploitation, or (5) Forensic Analysis. Further, several open-ended research opportunities are offered, thereby providing direction to other dark web researchers seeking to make contributions to this academic domain.

The focus of this thesis was drawn from one of the open-ended research problems presented in the survey we conducted under the *attribution* category. More precisely, this thesis considers *authorship analysis* in Tor-based dark marketplaces which can be defined by two objectives [51]. First, *user attribution* involves the identification of a given user based on their browsing behavior, traffic, semantic styles, or other such means. Second, *alias attribution* involves the correlation of distinct online profiles between communities. This task determines whether a user in one dark platform is that same user acting in another platform, forum, or marketplace. Several researches have been conducted for authorship analysis within the dark web domain, however, there are major distinctions between the goals of this thesis and that of the preceding works. For example, many current researches analyze hidden services and dark web users in general, whereas this thesis considers dark marketplaces and dark vendors specifically, which makes it a study of *vendor* and *vendor alias* attribution.

The following sections offer a literature review of research related to user attribution and alias attribution in the dark web domain. The intentions of the ensuing

discussion are to illustrate the relevance of this particular line of study, describe existing methodologies, introduce limitations in preceding works, and distinguish the contributions of this thesis from previous dark web authorship analysis research. A summarizing graphic of the progression of authorship analysis in the dark web is offered in Figure 3.1.

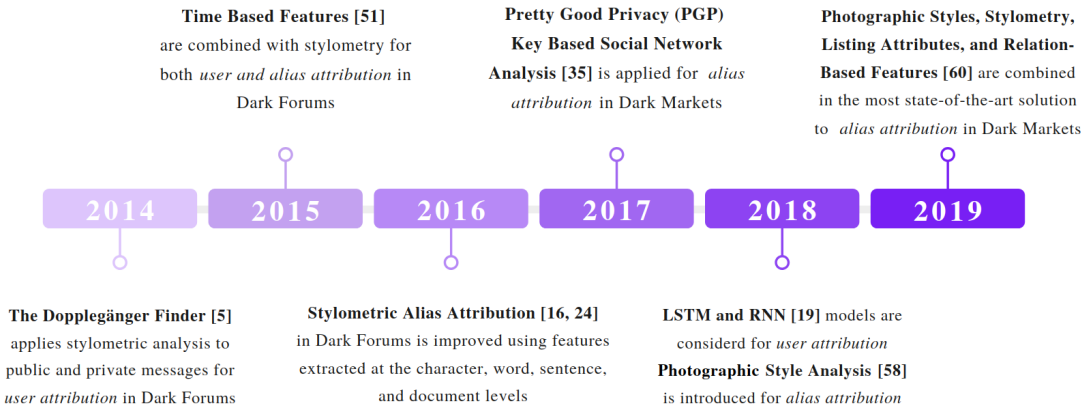


Figure 3.1: A summarizing timeline of dark web authorship analysis research progression between 2014 and 2019.

3.1 User Attribution

User attribution is innately a non-trivial task since dark web users generally try to mask their identities and become as indistinguishable as possible to maintain their anonymity. Due to the lack of personally identifiable data on hidden services, many attempts have been made to distinguish users based on *stylometric analysis*, i.e. the users' writing styles. For example, the method describe in [5] extracts features such as frequencies of punctuation, special characters, character n-grams, and function words among others from a vendor's public content (messages that are used to advertise

and request services) and private content (messages that are used to discuss more intricate details for an exchange). The features from these two types of messages are then used to train a Support Vector Machine (SVM) model to be able to distinguish an author given a document and a set of authors. Results indicated that the cleverly named model, The Dopplegänger Finder, peaked in accuracy once 4,500 words per author are used in training. Moreover, precision results are substantially better when using private messages as opposed to public messages, suggesting that text-based data is insufficient in training a model when considering public data alone.

User attribution is improved upon in [51], which introduces an SVM-based methodology where user profiles are analyzed based on topic-independent features, such as length of text and words, use of function words, interpunction and shallow syntactic patterns, along with *time-based features* and character n-grams. The classification task yields a ranking of the top-N candidates for author attribution of a given user post. Results show that model accuracy increases when combining time, stylometrics, and character n-gram features in comparison to each feature types used alone. Further, the model is able to achieve 88% accuracy in classifying the correct author to be the top-1 candidate and 97% accuracy in classifying the correct author to be in the top-5 ranked candidates. However, the author subsets used in the study are relevantly small in comparison to the user space that would need to be considered in practice. Further, cross platform user attribution is not considered in the scope of this work.

Subsequently, stylometric user attribution is approached in a study of applying LSTM and RNN neural networks to authorship predictions based on short user posts, but does not obtain results as successful as the preceding work [19]. Despite high accuracy on training set data, testing validation data sets resulted in insufficient accuracy, leaving room for improvement in stylometric authorship attribution. In

fact, user attribution research in the anonymous web is limited and most stylometry-based user attribution techniques are not suitable in practice because they under perform. Evidently, the user identification task one of the most challenging and most integral tasks for law enforcement, intelligence agencies, and the cybersecurity research community.

3.2 Alias Attribution

When individuals operate under a number of different accounts, they are considered to have *aliases*. Attributing these aliases is an important aspect in performing authorship analysis because it may assist in the identification of a user and potential prosecution of a cybercriminal. However, most alias attribution works in the anonymous web domain are challenged by a lack of ground truth availability.

In the following works, the alias attribution task is interpreted in many ways making the definition of alias attribution flexible. For example, some works consider alias matching a *link prediction* task in neural networks. Others choose to use a *distance heuristic* to measure the difference between user accounts and determine aliases by a specific distance threshold. Whatever the problem definition may be, a *classifiable*, i.e. distinguishable, user profile must first be created and then compared to other profiles. Since the task of alias attribution so closely relates to that of user attribution, many of the following studies are continuations of the aforementioned works from Section 3.1.

In an attempt to achieve alias attribution, the aforementioned Doppelgänger Finder [5] further considers the task of detecting multiple identities in both surface web level environments and underground forums. Using stylometric analysis to calculate pairwise probability scores, the model achieved significant precision and recall

rates in surface web environments. The results of experiments on dark forums were less conclusive, but manual analysis of the identified aliases overall supported the results of the Doppelgänger Finder classifier. However, the proposed method requires N classifiers for N authors, which may be too computationally expensive to apply in practice. Likewise, the work combining time, stylometry and character n-grams for user attribution extended their study to achieve alias attribution using the same features [51]. This new approach achieved sufficient precision for small sets of forums and users, but underachieved in terms of recall, resulting in 25% and 45% in pseudo user set sizes of 177 and 25 respectively. This shows that the solution is not scalable for practice.

Stylometry-based alias attribution in dark forums is also studied in [24] where a more promising methodology for dark web forums is proposed. Again, an SVM is implemented to determine if two accounts that share the same username across forums belong to a single individual. Specially, stylometric features are extracted at the character, word, sentence, and document levels, but not at the phrase, clause, and paragraph levels. Results were better compared to [51], however, required candidate aliases had to have at least 400 posts each and around 6,000 words to be viable for training and testing. Since it is challenging to find dark vendors with over 400 marketplace listings, this method is not very realistic to implement.

In the most recent work on stylometric alias attribution, tests are performed for both Twitter alias matching and dark web forum alias matching. In this study, the most successful models are able to achieve substantially high accuracy in Twitter alias attribution experiments (over 98%) and 90% accuracy in dark web forum experiments when users made at least 25 posts [16]. This is a huge improvement from [24] where it was suggested that a user had to have 400 posts in order to be classifiable. Notably, this work demonstrates how an unsupervised learning model can yield positive results

despite the absence of ground truth data which is a particularly important revelation in the dark web research domain since it is generally lacking in most cases.

Whereas the majority of alias attribution studies are stylometry-based, the alias attribution task may be considered with other techniques. For example, social network analysis can be applied to the dark web domain by analyzing vendor public PGP keys (i.e. encryption keys derived from the *Pretty Good Privacy* encryption scheme) and using PGP signatures to form graphical representations of *hidden relationships* between PGP key owners [35]. This analysis discovered many vendors who use different alias names on different markets while maintaining the same PGP key, thereby attributing blatant aliases across dark marketplaces. However, considering PGP key data alone is only sufficient for discovering a subset of vendor aliases. It should be used in conjunction with other methods, such as stylometric analysis, to ensure significant recall rates in alias attribution models. Interestingly, the same study demonstrated the ability to detect service *authorities* using PGP-key-based social network analysis, i.e. the owners of the marketplaces. Although, the focus of our research is dark marketplace vendor attribution rather than identification marketplace owners, the use of PGP key analysis is an intriguing concept for the Dark Vendor Profiling task and will be further discussed in Chapter 6.

In contrast to previous work mentioned, few works tackle alias attribution using photographic styles rather than writing styles or PGP key relationships. Namely, high-level image features, such as object, scene, background, camera angle, etc., can be extracted from images made available on product listings to profile dark vendors. In [58], Wang et al. use these exact high-level features to achieve alias attribution. Image metadata was at first considered with high-level image features, however, the metadata was determined to affect only a small portion of vendors, and therefore it was not included in the final classifier for this photo-based study. Their deep neural

network model is trained on high-level features alone to classify probable owners of an image, thereby identifying candidate vendor account pairs. The research demonstrates the effectiveness of image-based analysis over stylometric methods due to its accuracy and ability to fingerprint more vendors. However, it does not consider the potential ethical dilemmas when applying the suggested methods in more devious domains, such as in dark marketplaces involved in human and sex trafficking. Since this is a supervised learning model, training and testing samples are required to be labeled, so vendor aliases have to known. In [58], this is accomplished using a synthetic ground truth formulated by splitting a single vendors data into multiple pseudo-vendors. Furthermore, relying on photographic styles alone may dismiss important artifacts for the classification task such as text-based data present in market listings.

Building on top of the previous photo-based study, efforts to further improve solutions for the dark web alias attribution task combine text-based stylometery and photographic styles and apply their features to deep neural network models [60]. In this context, alias attribution is redefined as a link prediction task in a neural network. Using text and image content from four dark marketplaces, the *uIdentifier* classification model determines whether or not two vendors from the same marketplace are aliases and demonstrates how combining writing styles with photographic styles may improve model accuracies by an average of 10% compared to models using either writing or photographic styles alone [60]. In this work, writing styles are defined by several lexical, syntactic, and structural features, and photographic styles are defined by metadata and high level features. Further, drug and vendor features such as username, PGP key, shipping information and contact information are included in training as attributes. Finally, several relation-based features are considered, such as one that indicates whether or not a vendor sells a particular drug. The proposed

system, *uStyle-uID*, achieves up to 90.3% accuracy among the four tested marketplaces for determining whether or not a pair of vendors are the same individual. It is the most state-of-the-art methodology for this problem set. This research also briefly discusses cross-market vendors but focuses on alias attribution within a single platform. Again, a synthetic ground truth is used to test these techniques. Finally, this combination of techniques is considered for the task of alias attribution alone and not user attribution. Therefore, future research should consider (a) finding better ways to construct a ground truth for anonymous web analysis tasks, (b) establishing better methods for alias attribution across dark web platforms and across anonymous networks, and (c) applying similar profiling techniques to the task of user attribution.

3.3 Discussion of Limitations

Evidently, each of the aforementioned researches possess limitations which consequently serve as the motivation for this research. From our literature review, it was determined that the main challenges of attribution research was their limitations in representation of the dark web user universe and performance measured in accuracy, time, and memory usage. This section will discuss the limitations of current research as they relate to the general authorship analysis research domain rather than the sub-domains of user and alias attribution.

One of the earliest observations made during the literature survey process was that the majority of researches regarding authorship analysis relied strictly on text-based data and stylometric analysis where each dark web user was fingerprinted based on his/her unique writing style. Several problems arise from this strategy. For example, stylometric fingerprinting requires users to provide rich and diverse text samples which are most frequently available in dark forums and blogging services but

less guaranteed in dark marketplaces. Furthermore, the general goal for a user acting on the dark web is to remain hidden and indistinguishable. Therefore, as determined in [16] and [34], it is reasonable to suspect that criminals selling paraphernalia via dark marketplaces could use author obfuscation techniques to make their writing styles less apparent. For successful attribution in dark marketplaces, authorship analysis should consider a variety of data sources and user fingerprinting techniques so that accurate attribution is attainable even in the absence of rich and diverse text samples and the presence of purposefully obfuscated writing styles.

Several studies realize this limitation and consider other artifacts such as photographic styles in [58] and PGP keys in [35]. Similar to using writing styles to fingerprint users, photographic styles are proven to have success in distinguishing users. However, this method is limited to users who post photos rather than the entire dark marketplace user base since only photographic styles are considered. Further, analyzing photographic styles requires the methodology to include an image download process which may be a resource intensive task in itself and may heighten the risk of acquiring and possessing illegal content such as child pornography. In many cases, however, it is important for law enforcement and intelligence agencies to be able to profile sex and human traffickers active on the dark web. Thus, to achieve accurate attribution in dark marketplaces where pornographic content is shared, an authorship analysis technique must handle images legally and ethically.

Additionally, image analysis normally requires the use of neural networks, a popular machine learning technique used for mapping complex relationships between features. Unfortunately, the training of a neural network is not a trivial task and may be computationally expensive due to their complexity and memory/storage demand. Garrahan also considers neural networks in their stylometry-based work [19] but the

model does not perform well, demonstrating how neural networks may not be the optimal solution to authorship analysis in dark marketplaces.

The use of PGP keys is an interesting and unique take on authorship analysis, however, it is not sufficient to be used alone for several reasons. First, not all dark marketplaces are PGP-key-based. Thus, in marketplaces where keys are not used, the PGP method would have no success. Second, it is likely for a single user to use several keys for transactions, especially if they are purposefully trying to obfuscate their identity. In cases where single users utilize many keys, PGP analysis would again achieve limited success. However, Me et al. demonstrated in [35] that interesting connections could be made when considering PGP keys, making them a valid artifact for authorship analysis.

The analysis of writing styles, photographic styles, and PGP-based social networks each result in successful authorship analysis but are limited in scalability. In other words, stylometry is sufficient only in environments where rich and diverse text is available. Photographic style analysis is sufficient only in environments where images are prevalent. Finally, PGP key analysis is sufficient in only PGP-based marketplaces. Furthermore, each of these strategies may be hindered by author obfuscation when authors purposefully change their writing and photographic styles or use several PGP keys. Rather than using each technique alone, combining these artifacts may result in increased model performance by not only making users more distinguishable but also enabling fingerprinting users despite the absence of a subset of artifacts. This concept is realized in [60] with the *uIdentifier* classification model by analyzing a combination of photographic styles, writing styles, and other supporting attributes such as vendor and drug information. Thus, the uStyle-uID system presented by Zhang et al. is most similar to the profiling system proposed in this research.

Although the uStyle-uID makes up for many of the limitations of previous research, we are still able to identify some shortcomings that we attempt to improve upon. For example, since uIdentifer considers photographic styles, the uStyle-uID system relies on image downloading thereby limiting its scalability to marketplaces that do not possess child pornography or human trafficking content. Also, similar to other alias attribution studies, the uStyle-uID system imposes a pre-screening process such that vendor pairs are only considered if they can be related by the type of drug they sell. Other alias attribution methods impose similar limitations, taking vendor pairs into consideration only if their usernames are the same or at least similar. While these pre-screening processes may save system resources by not having to perform as many similarity computations, they may result in many alias accounts going undetected especially in cases where authors try to obfuscate their multiple identities by using diverse pseudo-usernames.

Finally, an overall limitation of all works examined is that the majority consider only one to four different dark marketplaces at a time. In practice, dozens of dark marketplaces may exist at any given time, so an effective user and alias attribution model should be robust enough to handle several marketplaces at a time. In other words, for a classification model to be truly effective and useful, it's performance should be independent of the number of marketplaces being analyzed.

Figure 3.2 summarizes and illustrates how the aforementioned techniques differ from the one proposed in this research. Overall, our proposed system attempts to alleviate the short-comings of previous dark web attribution work. In summary, the short-comings we improved upon include the following:

- Many models rely strictly on certain artifacts present in dark web services such as text, images, or user attributes alone. Due to this, many research

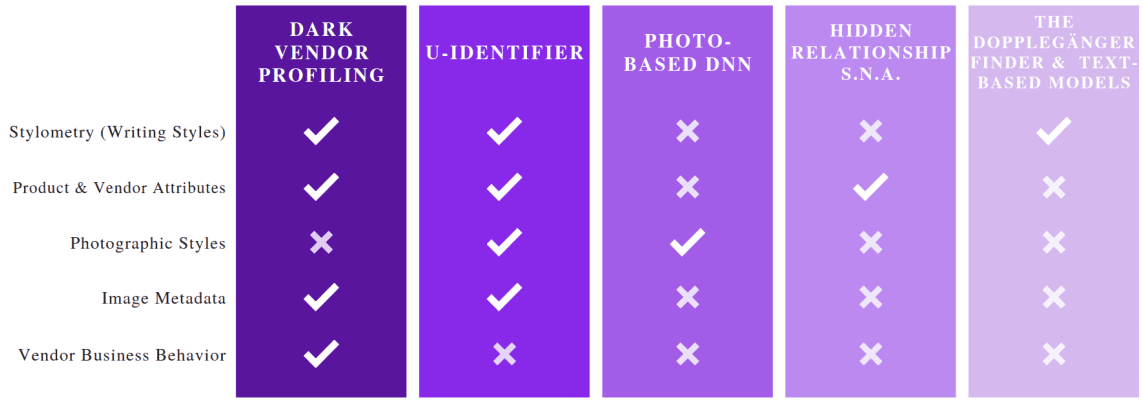


Figure 3.2: A comparison of the proposed Dark Vendor Profiling technique to related work in order of most similar to least similar, including uIdentifier [60], Photo-Based Deep Neural Networks [58], Hidden Relationships in Social Network Analysis [35], and Stylometric Text-Based Studies [5, 16, 19, 24, 51].

studies are missing usable evidence. By increasing the modality of attribution data sources, models may increase their performance and remain functional even when a subset of artifacts are not available in a particular service.

- To the best of our knowledge, random forests have not been applied in this domain. Thus, there is a research gap in understanding how this machine learning technique may perform in dark web marketplace investigations.

- Downloading illicit content and imagery provokes ethical dilemmas for researchers and investigators. To avoid such scenarios and decrease the risk of exposure to illicit image content, it would be advantageous to find ways to use image-based data without having to download and own the image itself.

- Users may purposefully attempt to mask their identities by using distinct pseudo-names and selling different paraphernalia under different accounts.

Thus, alias attribution models which rely on pre-screening to first identify likely vendor pairs may miss many important vendor connections. By removing prior username and other pre-screening assumptions, a model may be more successful in attribution even if the user hides behind pseudo-usernames and sells various paraphernalia using several vendor accounts.

- Most systems have only been tested on one to four marketplaces at a time, but, in practice, dozens of marketplaces may exist at any given time. By developing methods that are robust to data usage from many marketplaces at a time, machine-learning-based attribution can become more practical.

Next, a discussion of how our proposed system attempts to maintain and improve model performance while curtailing the short-comings of previous work is presented in the following chapter.

CHAPTER 4
THE DARK VENDOR PROFILING FRAMEWORK

First, it is important to define the principle terminology used in this thesis. From the discussion presented in Chapter 3, we can recognize that there exists a close relationship between user attribution and alias attribution works. In many cases, authors present a solution to both within a single work. In this thesis, both objectives are considered. First, we place an emphasis on user attribution since success in this domain also provides a solid foundation for success in alias attribution. However, since the methodology presented in this research relates to vendors active in dark marketplaces specifically, we redefine user attribution to *vendor* attribution and the overall task of authorship analysis to *Dark Vendor Profiling* (DVP). The goal of profiling may be considered very similar to that of fingerprinting. However, the distinction is that fingerprinting often refers to identifying a vendor based on a single type of artifact, such as writing style, photographic style, or attribute. In DVP, we propose *combining* artifacts such that a vendor can be identified by analyzing a variety of sources and artifacts at once, similar to how a social media profile identifies a user by their writing, images, interests and other attributes.

4.1 System Overview

The overall goal of DVP is to lessen the manual workload investigators handle when pursuing dark web investigations and aid in the deanonimization of dark vendors. To automate this process, DVP is presented as a framework for taking HyperText Markup Language (HTML) scrapes of dark web pages and processing this data to facilitate vendor and alias attribution. Specifically, the goal of vendor attribution is to be able to identify the owner of a product listing given data on

previous listings and their vendors. Since vendor usernames are always available to investigators examining dark marketplaces, vendor attribution may not seem like an advantageous capability in practice. However, we argue that vendor attribution demonstrates the effectiveness of profiling vendors using our proposed technique and is therefore an integral piece of the overall Dark Vendor Profiling task. If our proposed technique can achieve accurate vendor attribution, applying the same artifacts and methods to the task of alias attribution should result in effective and accurate profiling as well.

Therefore, the first focus of this thesis is to evaluate our vendor attribution technique and examine its robustness against varying numbers of vendors, listings, and marketplaces considered. Further, we present how these artifacts can be translated to the task of alias attribution and evaluate the accuracy, precision, and recall of the proposed methodology based on a synthetic ground truth representation of the dark vendor universe.

An important clarification about our work is that the DVP system does *not* include an automated process for obtaining dark web scrapes. Instead, it assumes scrapes have already been systematically collected. Again, this is an extension that may be considered for future work. The task of dark web crawling and scraping has been considered by many researchers and we categorize it as research in the *Data Mining* domain. Generally, data mining tasks in the dark web suffer from several limitations such as attaining limited completeness and under-utilizing available data sources due to challenges presented by the nature of the dark web. Specifically, most dark marketplaces consider data mining and web crawling to be undesired traffic by web services. According to a study on dark web data mining, the most common techniques for obfuscating web crawling include Turing tests, CAPTCHAs, user-agent identification, throttling of HTTPD requests, data tainting, injecting markers, and

network traffic analysis [6]. Therefore, automating the collection of HTML pages is a research task in itself and was not in the scope of this thesis.

4.2 Key Contributions

As determined in Chapter 3, several studies have been conducted for this task but each encompass limitations that prohibit the reliability of authorship analysis in practice. The DVP methodology attempts to improve on these limitations so that attribution tasks are more achievable in practice. In particular, DVP offers the following contributions to authorship analysis in dark marketplaces:

1. A new methodology for collecting image-based evidence that, if implemented during the preceding web crawling phase, would eliminate the need to download and own any dark web images.

2. A new feature set that is uniquely developed from a variety of artifact sources. Namely, stylometric-based, image-based, and attribute-based features serve as the basis for DVP.

3. A Random-Forest-based methodology for the task of vendor attribution in dark marketplaces. To the best of our knowledge, decision-tree-based tactics have never been applied to this domain.

4. A Record Linkage methodology for crafting vendor profiles and utilizing them to achieve alias attribution across dark marketplaces. To the best of our knowledge, record-linkage-based tactics have never been applied to this domain.

5. An evaluation of performance and resource usage of the proposed DVP system with regards to its robustness against varying amounts of vendors, listings, and marketplaces.

Application of Image Hashing in Dark Marketplaces

As listed in our Key Contributions, the DVP methodology proposes a unique way to collect image-based evidence such that the image *content* is not required for the classification model but image data is still considered in making decisions. This is accomplished by collecting any image *metadata* that may be embedded within the image file via EXIF (exchangeable image file format) and by recording a *hash* of the image which serves as a text-based unique representation of it's contents. This subsection will explain how image metadata and hashes are utilized for the DVP task.

First, whenever an image is taken or created, the image file embeds data about the image along with the data on the image content. For example, typical metadata may offer information with regards to when, where, how and by whom the image was taken. Obviously, any location-based information available to an investigator can be extremely advantageous, especially when seeking to deanonimize an anonymous web user. Unfortunately, cybercriminals are just as aware of this fact as investigators are, and often times, dark vendors and dark marketplace authorities will clear an image's auxiliary metadata before posting it in a hidden service. However, in cases where automatic metadata deletion does not occur, images could assist in the development of incriminating evidence if they contain metadata concerning the image's author, date and time of creation, location, camera make and model, and more. Thus, it is useful for the DVP classifier to take any present image metadata into consideration during classification.

Second, we previously determined several problems with using image content for authorship analysis tasks in the dark web. Namely, image analysis may not only be computationally expensive but also places a researcher at risk of exposure to illicit image content such as child pornography. However, image content may provide investigators with important incriminating evidence. According to Black Widow [46], a cyber intelligence gathering framework for dark web applications, there is substantial overlap between actors across dark forums, even if the forums are not based in the same language. Therefore, it is reasonable to suspect a similar overlap exists between dark marketplaces as well. Presumably, images have the potential to result in the identification of dark vendor aliases since it is likely a vendor would use the same images to sell their products if they were participating on several dark marketplaces and/or utilizing several pseudo-usernames. Therefore, we implement *image hashing* in DVP to capture a text-based representation of the image without having to own its content.

With image hashing, any image of arbitrary size is translated into a fixed sized string of characters; in our case, a 16 character string. Typical hashing algorithms are designed such that any change in the original content produces a drastic change in the hash. In contrast, image hashing algorithms are designed to be tolerant to changes in the original photo such as cropping, rotating, and filtering. Several image hashing algorithms exist, however, we selected four popular image hashing algorithms for our experiment to determine which algorithm would be the most appropriate for matching images within the dark web domain. This research is detailed in [28].

In short, in this research, we collected the images of product listings from 47 dark marketplaces scraped throughout 2013-2015. The data source of these images is the same source used in the DVP research and will be further detailed in Chapter 5. In the first part of the study, we ranked the dark marketplaces based on the amount

of data they offered and analyzed the presence of metadata in each image. Specifically, each of the 47 marketplaces in our dataset were ranked against each other using the following measures: number of listing entries, number of vendors, number of images, number of images with metadata, number of images with GPS coordinate data, proportion of images with metadata, and proportion of images with GPS coordinate data. Then, an average rank was calculated for each of the marketplaces using the seven aforementioned metrics, which denoted the overall significance of each marketplace in comparison with the others. The resulting ranks are listed in Table 4.1 in order of their significance as determined by this analysis. This table was an integral reference in determining the top 25 most effective dark marketplaces present in our dataset which were later selected for further analysis and DVP experimentation. Notably, many marketplaces did not contain any images with metadata. This may be due to automated metadata deletion techniques as discussed above.

Next, for every image present in our dataset, we calculated four image hashes using the following algorithms: Perceptual (PHASH), Difference (DHASH), Average (AHASH), and Wavelet (WHASH) hashing [42]. For each unique hash developed by the four algorithms, we then formed groups of images such that every image belonging to the group resulted in the same perceptual, difference, average, or wavelet hash. Then, within each group of unique hashes and hash types, we calculated an average Structural Similarity Index Metric (SSIM) [26] which allowed us to determine the percentage of similarity between image contents within the group. After calculating average SSIM values for each group, an overall *weighted* average SSIM was calculated for each of the four hash types. The weighted average SSIM provides each groups' SSIM value a weight determined by the number of image pairs used to calculate it. This way, groups with a large number of images more heavily influenced the overall

Marketplace	# Listings	# Vendors	# Images	# W/ Meta	# W/ GPS	% W/ Meta	% W/ GPS
Agora	116,858	3,154	64,535	2,292	214	3.55%	0.33%
Blackbank Market	12,852	905	10,565	2,086	470	19.74%	4.45%
Evolution	89,208	3,922	69,492	749	62	0.88%	0.07%
Alphabay	88,722	1,446	79,060	0	0	0%	0%
Pandora	15,223	516	15,066	11	0	0.07%	0%
Tor Escrow	958	185	866	241	43	27.83%	4.97%
Abraxas	16,641	432	11,979	0	0	0%	0%
Tor Market	1,502	200	817	230	14	28.15%	1.71%
Cloudnine	10,952	1,088	10,070	0	0	0%	0%
Dream Market	7,251	398	6,385	0	0	0%	0%
Cryptomarket	4,422	411	3,941	0	0	0%	0%
Middle Earth	6,650	359	6,167	0	0	0%	0%
Andromeda	3,054	237	2,947	0	0	0%	0%
Bluesky	2,400	213	2,089	0	0	0%	0%
Oxygen	2,212	257	2,012	0	0	0%	0%
Freebay	507	175	417	97	6	23.26%	1.44%
Hydra	2,282	166	2,240	0	0	0%	0%
Cannabis Road 2	1,537	155	1,442	0	0	0%	0%
Area51	489	74	479	89	6	18.58%	1.25%
East India Company	1,429	143	1,232	0	0	0%	0%
The Real Deal	981	82	873	0	0	0%	0%
Black Services	639	167	621	0	0	0%	0%
The Marketplace	823	124	584	0	0	0%	0%
Amazon Dark	199	41	190	57	5	30%	2.63%
Haven	741	74	704	0	0	0%	0%
Darkbay	538	124	533	0	0	0%	0%
Cannabis Road 3	318	95	258	3	0	1.16%	0%
Panacea	461	21	459	0	0	0%	0%
Silkstreet	35	14	33	11	7	33.33%	21.21%
Freemarket	169	6	167	52	1	31.14%	0.6%
Poseidon	427	17	427	0	0	0%	0%
Torbazaar	383	27	332	0	0	0%	0%
Tochka	197	29	192	0	0	0%	0%
1776	171	37	170	0	0	0%	0%
Darknet Heroes	207	28	153	0	0	0%	0%
Deepzon	56	6	55	19	0	34.55%	0%
Dogeroad	112	28	100	0	0	0%	0%
Underground Market	143	20	112	0	0	0%	0%
The Majestic Garden	88	16	63	0	0	0%	0%
Horizon	44	11	44	2	0	4.55%	0%
Cantina	21	9	19	5	0	26.32%	0%
Topix2	34	24	24	0	0	0%	0%
Sheep	8048	370	0	0	0	0%	0%
Bloomsfield	8	3	8	0	0	0%	0%
White Rabbit	313	62	0	0	0	0%	0%
Kiss	415	10	0	0	0	0%	0%
Greyroad	21	9	0	0	0	0%	0%

Table 4.1: Dark marketplaces listed in order of significance to DVP by calculating an average rank over seven characteristics: number of listing entries, number of vendors, number of images, number of images with metadata, number of images with GPS coordinate data, proportion of images with metadata, and proportion of images with GPS coordinate data.

Image Hash Type	Weighted Avg SSIM	Avg SSIM
PHASH	0.991	0.986
DHASH	0.987	0.989
AHASH	0.881	0.976
WHASH	0.660	0.975

Table 4.2: Hash analysis results for average, difference, perceptual, and wavelet hashing in order of weighted average SSIM values.

weighted averaged SSIM compared to groups with a small number of images. Table 4.2 lists the results of this analysis.

Evidently, the PHASH algorithm resulted in the grouping of the most similar images as depicted by its high average SSIM value of 91.1%. Therefore, the PHASHing technique was selected to represent images in the DVP system.

Machine Learning Based Profiling

DVP accomplishes vendor attribution by implementing a supervised machine learning classification model. Whereas other studies implement Neural Networks and Support Vector Machines, DVP implements a Random Forest Classifier [31]. The Random Forest algorithm is a popular machine learning model that uses ensemble learning to make classifications. These forests consist of N Decision Trees (also known as estimators) each of which build themselves based on the provided labeled training data. The goal of using many Decision Trees instead of one is to reduce over-fitting and improve model accuracy [13]. When classifying a single input, each Decision Tree in the forest will make a prediction, and the forest will output the final classification based on a majority vote.

In choosing an appropriate model for the vendor attribution task, we considered several desirable characteristics which would be important to have in order to

apply the model in the dark marketplace domain. For example, it was important for our model to be able to handle missing values since often times vendors choose to omit certain details about their products and profiles. Further, we needed a model that was able to handle categorical data since many of our features are not numerical, such as the shipping information or category of a product listing. With these considerations in mind, we compared the appropriateness of several common machine learning models based on their general model behavior. Our comparison is illustrated in Figure 4.1. Several other characteristics and models were considered during the comparison, however, we ultimately determined the following five characteristics to be most integral to the success of DVP and the following five models to be the best candidates for DVP.

	Decision Trees	Naive Bayes	Bayesian Networks	Artificial Neural Networks	Support Vector Machines
Applicable to Categorical Data	✓	✓	✓	✓	✓
Can Accomplish Multi-Class Classification (without Binarization)	✓	✓	✓	✓	×
Can Handle Missing Values	✓	✓	✓	×	×
Can Handle Irrelevant Data	✓	✓	✓	✓	✓
Can Handle Redundant Data and Correlation Variables	✓	×	✓	✓	✓

Figure 4.1: A comparison of several popular machine learning techniques based on desirable characteristics of dark marketplace classification models.

As it can be seen, many of the considered classic machine learning models were similar in characteristics. Although the Bayesian Network model was the only other classification technique to have each of the five necessary characteristics, it requires a manual configuration of how each feature is mapped to each other and to

an output, whereas Decision Trees automate this process by including a weighting and configuration mechanism within its algorithm. Therefore, for the sake of simplicity, we determined that Decision Trees are better candidates for DVP in comparison to Bayesian Networks.

Furthermore, Naive Bayes, Artificial Neural Networks (ANN), and Support Vector Machines (SVM) each had most of the necessary characteristics but were missing some key capabilities. For example, neither ANN's nor SVM's handle missing values well. Considering the hidden and obfuscated nature of dark web marketplaces, it is likely that listings will often contain missing values, so it is desirable to choose a model that can handle such scenarios. Additionally, the Naive Bayes model cannot handle redundant data, i.e., variables that are strongly correlated with one another, due to its unique *conditional independence* assumption. In other words, the Naive Bayes model requires its variables to be independent from one another - an assumption that is not reasonable to make in the dark marketplace domain since several variables are usually correlated, such as product category, subcategory, and name.

Since the Decision Tree method showed the most promise in meeting all of the demands of dark marketplace analysis, it was chosen to serve as the basic classification model for DVP. Further, since ensemble methods generally perform better than stand-alone methods, we decided to apply the ensemble-based Random Forest model in place of it's stand-alone counterpart, the Decision Tree.

The second component of DVP is alias attribution. In this research, we experiment with two methods for attributing aliases. The first method, which we designate as *elementary style* attribution, works by computing the average cosine similarities between the features that make up vendor profiles. Cosine similarity is a measure of similarity frequently used in determining the proximity of entities belonging to a single cluster as determined by an unsupervised learning algorithm [41]. Thus, it is a

natural candidate for determining the degree of similarity between two vendors. We propose that pairs which result in high cosine similarity values indicate likely vendor aliases. In short, the elementary style model works by first calculating the cosine similarities of each individual feature between every possible pair of vendors and then using the individual feature similarities to calculate the overall average similarity of each vendor pair. It is important to note all features considered in this version of alias attribution have equal weight in the final cosine similarity metric.

The elementary style method has many flaws that make it impractical for investigative purposes. These will be further discussed in Chapter 7. However, the results this method yielded provided us with important insight for designing an enhanced DVP alias attribution method. In the *enhanced* methodology, we interpret the task of alias attribution as a *record linkage* problem. Typically, record linkage is the task of linking database entries that are believed to belong to the same entity [43]. For example, proper record linkage methods would be able to identify that the database record for Jane Doe living on West Street represents the same individual as the database record for J. Doe living on West St. Therefore, record linkage can also be considered as a data matching or duplicate attribution technique.

Considering these definitions, we explored record linkage in DVP by considering *alias vendors* to be synonymous with *duplicate vendors* within a dataset. Namely, we used Python's record linkage library to train and test both Support Vector Machine (SVM) classifiers [10] and Logistic Regression (LR) classifiers [11] for alias attribution in DVP [43]. These classifiers were chosen for our experimentation because they are two of the three supervised machine learning models implemented by the record linkage package and are able to handle categorical, numeric, and string based features.

In short, the *enhanced* alias attribution model works by first creating candidate pairs instead of linking every vendor with one another like in elementary style attribution. The details of candidate pair formulation will be further discussed in Chapter 7. Next, the model calculates *comparison vectors* for each of the candidate pairs where all categorical features are compared using the levenshtein distance metric [7], all numerical features are compared using the gaussian distribution function, and all frequency-based features are compared using the cosine similarity. The particular features used in training will be further discussed in Chapter 6. However, we note that by incorporating several similarity metrics instead of using cosine similarity alone, we are better able to tailor to the data we have at hand, thereby resulting in more accurate models. Further, by using SVM and LR models, we are able to calculate *weighted* similarities where each feature is not necessarily considered equally important as other features. Finally, after candidate pairs are formulated and their comparison vectors are computed, we train our models to be able to predict alias pairs based on comparison vector values.

4.3 Dark Vendor Profiling Workflow

Based on our findings described above, we designed the DVP system as illustrated in Figure 4.2 beginning with the collection of dark web marketplace HTML pages. Since the actual scraping mechanism was out of scope for our research, it is grayed out in the figure; however, it remains an integral part of the DVP system architecture. Following the collection of HTML pages, we parse and store listing and vendor information into a database which we call *DVP DB*. The specific information collected will be further detailed in Chapter 5. From the parsed data, we extract three types of features - stylometry-based, attribute-based and image-based - and

feed each feature into our DVP machine learning models. Each of the steps in this workflow were implemented using Python 2.7 and higher and MySQL 8.0. The implementations of the feature extraction and model development will be further detailed in Chapter 6.

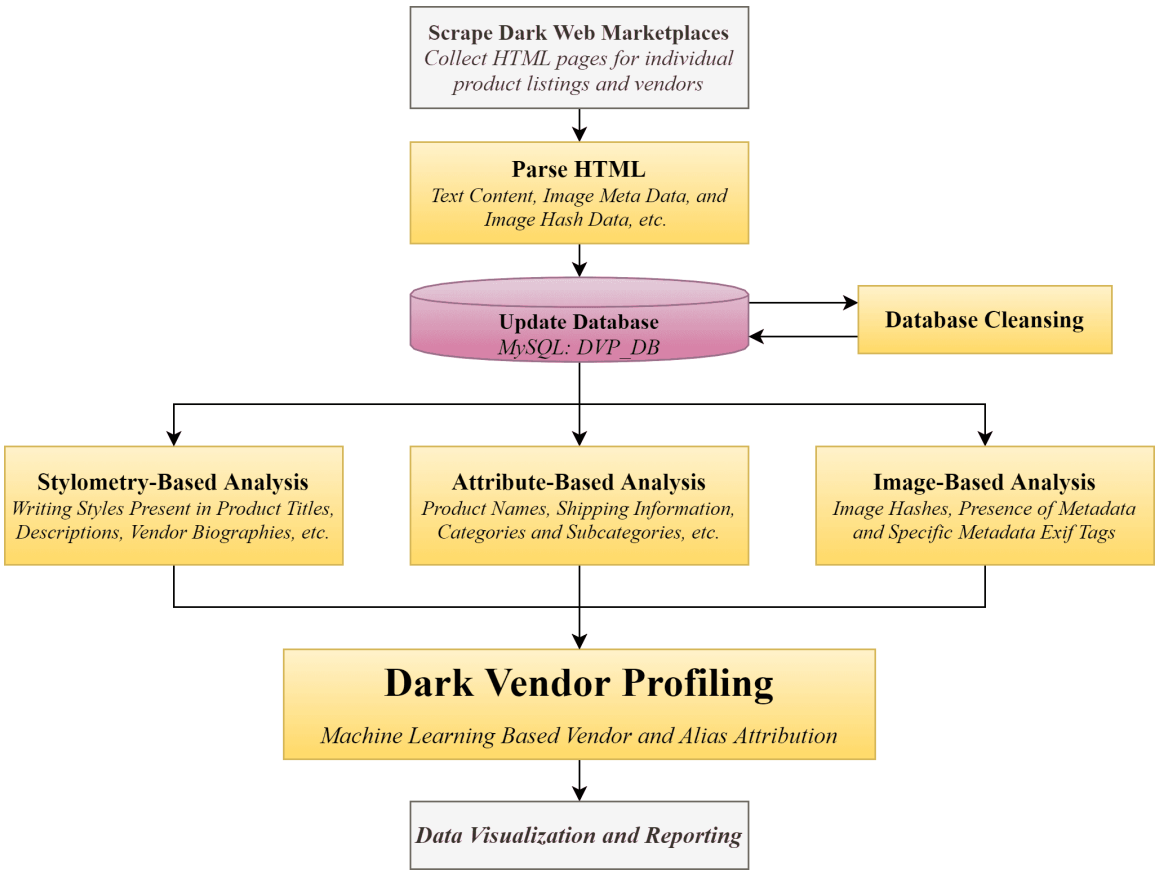


Figure 4.2: An illustration of the DVP system workflow.

CHAPTER 5

DATA

One of the greatest challenges in experimenting with machine learning techniques in the dark web domain is the lack of standard and complete datasets. Due to the hidden nature of dark web services and the absence of site indexing, it is imposisble to determine the size of the dark marketplace universe. Consequently, it becomes challenging not only to find relevant dark marketplaces to collect data from but also to determine what proportion of all dark marketplaces we are considering. Further, given a valid dark marketplace onion address, it may be challenging to access and scrape the site if it requires a membership and surpassing of obfuscation techniques like CAPTCHAs. Overcoming these challenges is a huge undertaking in itself. Hence, it was not in the scope of this research. Instead, we reuse publicly available HTML scrapes taken between 2013 and 2015. This chapter will discuss our data source and how it was used in the development of the MySQL-based DVP database (DVP DB).

5.1 Darknet Market Archives

In an effort to provide dark web researchers with data on dark marketplaces and forums, Branwen et al. hosts an archive of approximately 89 dark web forums and markets scraped from 2011-2015 [8]. The Darknet Market Archives (DNM Archives), which consists of approximately 1,500 gigabytes of data, has been viewed over 30,000 times and referenced in a number of researches, covering topics on dark web user behavior [12, 37], dark web research ethics [33], dark market impacts [27], automated analysis of anonymous environments [4, 15, 20, 39], and intelligence availability in anonymous web settings [19, 24, 30, 36, 57, 60], among many others. Branwen et al. explicitly states child pornography is not a concern in the darknet market archives

since each marketplace and forum in the archive ban this type of content. Further, Branwen et al. warns that the scrapes are large, complicated, redundant, and highly noisy due to several factors. For example, listings may appear and disappear in the market during a crawl, vendors may be banned at any time resulting in the deletion of all of their listings and profile, Tor connections may fail, vendors may purposefully post misleading listings and labels, or the marketplaces themselves may go unavailable at any time. Therefore, it is important to consider the incompleteness and high likelihood of discrepancies present in any dark web scrape dataset. Despite these warnings, we determined the darknet market archives to be the best option available under the circumstances and most appropriate for our DVP research. Thus, we extracted a subset of this archive which we transformed into our database, DVP DB.

5.2 DVP DB - The Dark Vendor Profiling Database

We selected a relational database to store our DVP data to avoid over-storing redundant information. For example, for each listing, we collected data relative to the listing itself and to the vendor who owned the listing. If, for each listing, we saved both listing and vendor data in a single entry, we would end up with highly redundant vendor data since vendor profiles generally remain constant over time. Instead, using a relational database, we can create separate listing and vendor tables so that several listing entries may be related to a single entry in the vendor table, thereby eliminating overly redundant vendor information. Thus, we organized the parsed data into five tables which are described in Table 5.1.

Table Name	Stored Data
Listings	Listing ID, Scrape Date, Marketplace, File Path, Product Name, Vendor Name
Images	Listing ID, Image Path, PHASH, Exif Data
Details	Listing ID, Category, Subcategories, Description, Price, Ships From, Ships To
Vendor Details	Scrape Date, Marketplace, Vendor Name, Date Registered, Biography, PGP Key, Biography Hash, PGP Key Hash
Vendor Activity	Scrape Date, Marketplace, Vendor Name, Rating, Last Seen Date, Trade Count

Table 5.1: Basic DVP DB Schema with five tables.

The dataset we originally pulled from the DNM Archive consisted of 34.4 GB of zipped directories containing HTML, JavaScript, and styling files along with images where each directory was formed from a different marketplace. After extracting the zipped directories, 47 of the 74 downloaded directories (totaling 667.6 GB) were determined to be useful for the image hash analysis database introduced in Chapter 4. The remaining 27 directories (115 GB) were not processed into the DVP DB due to a number of factors including lack of image data or inconsistent HTML formatting which would have made organizing web scraped data into the DVP DB time consuming and futile. Further, after ranking the 47 marketplaces based on their size and significance, we limited our scope to the top 25 marketplaces, which consisted of approximately 123 GB of HTML scrapes from the DNM Archive. This may be considered a limitation to this work since the DNM Archive used for this study was not considered in its entirety.

All data was parsed from the HTML scrapes and written to DVP DB using Python 2.7 scripts. Due to the uniqueness of each marketplace's HTML format, a separate parsing script had to be implemented for every marketplace. In the following subsections, we will describe in detail what information is stored in the DVP DB, how the data was cleansed and pre-processed in preparation for feature extraction, and why we developed our database in stages resulting in five DVP DB versions.

Information Collected

As mentioned before, each of the 25 marketplace directories consisted of HTML pages with unique formatting, thereby requiring us to implement 25 separate parsing scripts. Despite inconsistent formatting, the majority of the marketplaces offered similar information. Through manual analysis, we determined the most frequently available data on dark marketplaces and aimed to collect each of the 15 pieces of information from each marketplace listed in Table 5.2. According to this analysis, the only data that was consistently available across all marketplaces was product name, description, price, image, and vendor name.

Data Cleansing

Inevitably, the parsing process resulted in some error in DVP DB entries. For example, any scrapes from the marketplace archive created by error, i.e. scraping login pages rather than listing and vendor pages, resulted in erroneous entry creation in DVP DB. In order words, pre-existing errors in the DNM archives propagated to the DVP DB, as author Branwen et al. cautioned when presenting his datasets. Therefore, after parsing, we ran several MySQL and Python scripts as part of the *database cleansing* phase.

First, by directly manipulating the DVP DB entries using MySQL scripts, we deleted any entries that where missing critical information needed for analysis. These erroneous entries were caused when attribute values were not found in the HTML as expected. For example, if our parser attempted to find the `<h1>` HTML tag which typically stored the product name of a listing and was provided the HTML file of the login page instead of the listing page, then our parser may have accidentally stored `Home` or `Login` as a product name. Obviously, such data is not useful to the task of

	Product Name	Product Description	Product Category	Product Subcategory	Ships From	Ships To	Product Price	Product Image	Vendor Name	Vendor Biography	Vendor PGP Key	Registered Since Date	Last Seen Date	Vendor Rating	Vendor Trade Count
Abraxas	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Agora	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Alphabay	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Andromeda	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Area 51	X	X			X	X	X	X	X	X	X	X	X	X	
Black Bank Market	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Bluesky	X	X			X		X	X	X	X	X	X	X	X	X
Cannabis Road 2 X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Cannabis Road 3	X	X			X	X	X	X	X	X	X			X	X
Cloudnine	X	X	X	X	X	X	X	X	X	X	X		X		X
Cryptomarket	X	X	X		X	X	X	X	X	X	X		X	X	
Darkbay	X	X			X	X	X	X	X	X	X	X	X	X	
East India Company	X	X	X	X	X	X	X	X	X					X	
Evolution	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Freebay	X	X			X	X	X	X	X	X	X			X	
Freemarket	X	X			X		X	X	X	X	X	X	X		
Haven	X	X	X	X			X	X	X	X	X		X	X	
Hydra	X	X			X	X	X	X	X	X	X	X	X	X	X
Middle Earth	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Oxygen	X	X	X	X	X	X	X	X	X	X	X	X	X		X
Panacea	X	X	X	X	X	X	X	X	X	X	X		X		X
Pandora	X	X			X	X	X	X	X	X	X	X	X	X	X
The Marketplace	X	X	X	X	X		X	X	X	X	X	X	X	X	
Tor Escrow	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Tor Market	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Table 5.2: Data availability per dark marketplace from the DMN Archives. Each X represents whether or not the information is regularly available in a marketplace.

Entry Type	Missing Values	Number of Entries Affected
Listings	Product and Vendor Name	19
Listings	Vendor Name Only	2
Images	PHASH	53
Vendor Details	Date Registered, Description Hash and PGP Hash	968
Vendor Activity	Rating, Last Seen and Trade Count	431
Total		*1,473*

Table 5.3: Summary of deleted DVP DB entries due to missing integral data.

DVP, so we had to remove it from our database before we could perform any further meaningful analysis. It is important to note that the *listings* table can be considered a root table, as any entry deletions in this table were cascaded to both the *images* and *details* tables as well.

Specifically, we removed any *listings* table entries that were missing (a) both product and vendor names or (b) just vendor names. This resulted in the deletion of only 21 entries. We also removed any *images* table entries that were missing PHASHes since an empty PHASH indicated an error in opening and reading image files. This resulted in 53 deletions. Finally, we removed any *vendor detail* table entries that were missing values for date registered, description, and PGP key, and any *vendor activity* table entries that were missing values for rating, last seen and trade count. Any vendor-related entry in the database with empty values for these attributes were clearly erroneous and incapable of producing any analytical value. This resulted in the deletion of 1,399 entries. A summary of deletions due to missing integral data is presented in Table 5.3.

Additionally, we tried to delete any listings that had suggested sex as a service. This attempts to remove listings selling pornography and access to pornographic accounts but does not remove drugs and medications that are related to pornography or sex. While the DVP system could be used for human and sex trafficking cases, we

Entry Type	Criteria	Number of Entries Affected
Listings	Product Name contains 'Porn' or 'Erotica'	2,236
Details	Category or Subcategory contains 'Porn' or 'Erotica'	973
Details	Description contains 'Teen Porn'	5
Details	Description contains 'Porn' *and* 'Account'	946
Total		*4,160*

Table 5.4: Summary of deleted DVP DB entries due to suggestion of sex as a service.

chose to omit related listings in this study to avoid ethical and legal dilemmas since including that was not necessary to evaluate DVP's performance. Unfortunately, this may be considered a limitation to our work if the cleansing process resulted in the deletion of non-pornographic content. Table 5.4 describes the criteria for deleting entries based on their relatedness to offering sex-as-a-service.

Finally, using a python script, we attempted to clear any duplicate entries in DVP DB. We suspect that the duplicate entries were a result of a parsing error where listing entries were considered 'new' based on the file name they were saved under in the DNM Archives instead of being based on the files' content. Since some files in the archive were found to be saved under different file names but contained the same listings, duplicate removal was a necessary step in preparing our data for our DVP classification model. Thus, if entries shared matching Marketplace, Vendor Name, and Product Name entries in the listings table, all but one of the entries were deleted. This resulted in 98,027 entry deletions in the listings table.

Versions

In the development of DVP DB, we chose to save 'snapshots' of database progression to ensure large amounts of data would not be lost on accident and to

ensure we would not need to re-parse all 123 GB of HTML data. These snapshots are saved as separate MySQL databases, resulting in five versions of the DVP DB.

Version 1 was the original database used for our first DVP-related study on image hashing. This database was specifically used for the hash analysis, thus it only contained data on images and their hashes.

Version 2 was created after completing the hash analysis and determining the top 25 best marketplaces in the DNM Archive for the DVP task. This version reflects the schema from Table 5.1 and attempts to collect all of the data listed in Table 5.2. Lastly, DVP DB V2 is the version that underwent the *data cleansing* phase, thereby removing erroneous entries, entries that suggested sex as a service, and duplicates.

Version 3 underwent the *pruning* phase during which we eliminated data columns that were unused in feature selection. A more detailed discussion on the omission of these columns will be presented in Chapter 6; however, in short, it was determined that we would not be able to extract valuable features from a few of the attributes we initially collected including price, date registered, trade count, rating and last seen. Thus, DVP DB V3 prunes the *vendor activity* table in it entirety as it contained no valuable information for our feature extraction procedure. However, since the data was maintained in DVP DB V2, future work may consider producing new mechanisms for extracting meaningful features from the omitted data.

Version 4 underwent a *merging* process such that the original schema was simplified into two tables - *listings* and *vendors*. This was accomplished to simplify the feature extraction process later on. Specifically, this process merged the *images* table and *details* table entries into the *listings* table such that no image or detail entries were present in DVP DB V4 without having a corresponding listing entry. However, it is possible for a listing entry to have no corresponding image or detail entries. In these cases, the empty columns are filled with NULL or empty strings. It

Version	Tables	Marketplaces	Scrape Dates	Listing Table Entries	Unique Vendors in Listings Table	Unique Vendors in Vendors Table
1	3	47	391	400,741	15,890	N/A
2	5	25	375	303,359	15,129	62,215
3	4	25	375	303,359	15,129	62,215
4	2	25	375	303,359	15,129	12,739
5	2	14	331	147,031	10,508	7,424

Table 5.5: Summary DVP DB versions and their contents.

is important to note DVP DB V4 contains both complete and incomplete listing and vendor entries, meaning that empty valued attributes are permissible.

Version 5 contains the subset of DVP DB V4 entries that have a value for all columns in the *listings* and *vendors* tables. In other words, DVP DB V5 contains complete listing and vendor entries only, which allows us to examine the effect of missing data on the DVP system. With DVP DB V4 and V5, we aim to demonstrate how the DVP classifier is resilient to missing values by comparing the performance of classifiers constructed from the two versions.

The contents of each of the database versions are listed in Table 5.5. One notable observation is that versions two and three have many more unique vendors present in their vendor tables than their listings tables. This is indicative of one of two things: (a) many vendors did not have any listings posted over the duration of the DNM Archive scrapes or (b) the DNM Archive scrapes missed the listings of many vendors and were therefore omitted from the DVP DB. This may be a limitation to the completeness of our DVP data but may also suggest future work in strictly considering vendor profile pages alone for DVP rather than analyzing both their listings and profile pages.

CHAPTER 6
METHODS AND DESIGN

In the majority of works related to authorship analysis in the dark web, authors build their classification schemes based on a single type of feature. In this research, our goal is to examine the effectiveness of building classification models on several types of features. Namely, we consider stylometry, attribute, and image-based features for our DVP model. In this chapter, we discuss which features we consider from each feature type and explain how they were applied to vendor and alias attribution tasks. All code for feature extraction and model development was written in Python 3.

6.1 Feature Engineering

Our feature engineering process was heavily based on the literature review we performed in the making of the DVP system. Based on existing research, we came up with several ideas for stylometric, attribute, and image-based features that could influence the performance of our classifier. Then, we implemented a feature extraction process using a python script which would translate an input CSV (comma-separated values) file into a new CSV file that was ready for interpretation by our DVP model.

In this study, we wanted to examine the effect of training and testing classification models with missing data since missing and mislabeled data is often present in dark marketplace platforms. Therefore, for our experimentation, we extract features from both version 4 and version 5 of the DVP DB. While all extracted features are considered in the development of our vendor attribution model, the Random Forest classifier algorithm incorporates an automated feature selection mechanism by assigning weights to features based on their importance for classification.

It is also noteworthy that the tasks of vendor and alias attribution require slightly modified feature sets due to the fact that the vendor attribution model is trained on individual listings, whereas the alias attribution model is trained on individual vendor profiles. Specifically, the alias attribution feature set includes all features used in the vendor attribution feature set in addition to a few features that may describe a vendor's overall behavior rather than the behavior they exhibit on an single product listing. These features will be detailed in the following subsections.

Stylometry Based

From the literature review, is it evident that the use of stylometric features is a popular technique for authorship analysis. Therefore, we use several similar features to those that are suggested in related works [3, 38, 44]. These features are listed in Table 6.1. Note, the vendor attribution model extracts the stylometric features of an individual listing's Product Name and Product Description. In contrast, the alias attribution model takes a holistic view of a vendor and extracts *average* features from a collection of a vendor's listings based on the listings' Product Names and Product Descriptions, and the Vendor Description.

One of the most unique stylometric features we propose in DVP is the frequency of *top keywords*. In this manner, we may incorporate a term frequency analysis based on the TF-IDF technique, a popular text mining methodology for finding important words in a collection of documents, while also avoiding the development of an overfit model [47]. Specifically, given a dataset from which we wish to extract features, we calculate the top most important term out of all terms found in product names and the top 10 most important terms out of all terms found in the product and vendor descriptions. Then, for each listing considered for vendor attribution or

each vendor considered for alias attribution, we calculate the frequency (or average frequency) of each 'top keyword' used.

In other studies, word level and character level N-grams were considered for stylometry-based analysis [5, 51]. However, we chose to exclude these features from DVP to simplify the feature extraction process and lessen the computational resources required to both extract features and train models. Additionally, we could have considered paragraph level features, but chose to exclude them due to our inability to guarantee that the scraping and parsing processes preserved the paragraph level structure of the bodies of text taken from dark marketplaces. In fact, to further alleviate the effect of scraping and parsing errors, we removed all new lines and tabs from Product Name, Product Description, and Vendor Description before extracting stylometry-based features from their values.

Attribute Based

We also extract attribute-based features based on *Base* Product Name, Category, Subcategory, Ships From, and Ships To values stored in DVP DB. For alias attribution, we additionally consider Vendor Name and PGP Key data. Similar to how top keywords were calculated for stylometric feature extraction, *base product names* were determined by identifying the top keyword of each product name. Additionally, shipping and category information was pre-processed such that if a listing's shipping, category, or subcategory values were not present in a pre-computed list of allowed values, they were marked 'unknown'. The lists of allowed values were formulated from an analysis of all shipping and category information present in DVP DB V4 and DVP DB V5 separately. This way, attribute-based features extracted from version 5 data were based on DVP DB V5 lists and features extracted from version 4

Source	Type	Feature
Product Name	Count	Length in Characters Words ASCII Letter Characters Digits Special Characters / Punctuation Spaces Non-printable Unicode Characters
	Frequency	Top Important Keywords Uppercase and Lowercase Letters A-Z Digits 0-9 Special Characters / Punctuation Non-printable Unicode Characters
Product/Vendor Description	Count	Length in Characters Words Sentences Average Characters per Sentence Average Words per Sentence Average Word Length Misspelled Words ASCII Letter Characters Digits Special Characters / Punctuation Spaces Non-printable Unicode Characters
	Frequency	Top Important Keywords Uppercase and Lowercase Letters A-Z Digits 0-9 Special Characters / Punctuation Non-printable Unicode Characters

Table 6.1: Stylometric features extracted for DVP. Note, vendor description features are considered only in alias attribution, whereas product name and product description are considered for both vendor and alias attribution.

data were based on DVP DB V4 lists. These measures were taken in an attempt to standardize the categorical data shared across all marketplaces, minimize feature dimensionality and reduce the risk of developing overfit models. Table 6.2 summarizes the attribute-based features considered for DVP.

Task	Source	Type	Feature
Vendor Attribution	Product Name Category Subcategory Ships From Ships To	Categorical	Base Product Name Category Subcategory Ships From Ships To
Alias Attribution	N/A Product Name Category Subcategory Ships From Ships To PGP Key	Numeric	Total Product Listings Unique Base Product Names Unique Categories Unique Subcategories Unique Ships From Locations Unique Ships To Locations Unique PGP Keys
	Product Name Category Subcategory Ships From Ships To PGP Key	Frequency	Unique Base Product Names Categories Subcategories Ships From Locations Ships To Locations PGP Keys
	Vendor Name Product Name Category Subcategory Ships From Ships To PGP Key	Categorical	Vendor Name Top Base Product Name Top Category Top Subcategory Top Ships From Location Top Ships To Location Top PGP Key

Table 6.2: Attribute-based features extracted for DVP.

Image Based

Finally, we extract a few image-based features in an attempt to boost DVP model performance. These select features are listed in Table 6.3. Notably, we consider the EXIF tags that are present within an image's set of metadata, but not the actual values associated with the tags. For example, if an image contains GPS related data, we extract the presence of this tag as a feature but not the actual GPS data that the image stores. This is due to a limitation of our selected model type, the Random Forest classifier, which requires all categorical data to be encoded. Encoding the values of exif tags would have resulted in a highly dimensional feature set which would

Task	Source	Type	Feature
Vendor Attribution	PHASH	Categorical	PHASH
	EXIF Tags	Boolean Categorical	Metadata Availability EXIF Tags
Alias Attribution	PHASH	Numeric Frequency Categorical	Unique PHASHes PHASHes Top PHASH
	EXIF Tags	Numeric Frequency Categorical	Images with Metadata EXIF Tags Top EXIF Tag

Table 6.3: Image-based features extracted for DVP.

increase the risk of facing the curse of dimenstionality and overfitting. Therefore, the inclusion of exif tag values remains a direction for future work.

Excluded Data

In the previous chapter, we determined that we would not be able to extract valuable features from price, date registered, trade count, rating and last seen data, so we pruned these data columns from DVP DB versions 3 through 5. However, each of these data may be considered for future DVP development. Our reasoning for excluding these attributes during this study are the following:

- **Price** is difficult to standardize since it is so reliant on the product being sold. For example, a hacking book listing would likely be valued differently from a prescription pill listing, so if we wanted to compare a vendor who sells books to a vendor who sells drugs, price may not be a good indicator of similarity. Perhaps, if we were to include a monetary-based feature in DVP, it could be translated into a categorical feature that describes whether the product is under, average, or over priced in comparison to similar items. However, this type of feature translation would not only require domain

expert knowledge but also require all items to be listed with their unit prices. This data analysis was not in the scope of this study.

- **Date Registered and Last Seen** were the two least precise attributes available due to a lack standardization between dark marketplaces. For example, some marketplaces would list exact dates when a vendor registered and was last seen while other marketplaces would simply denote that they had been registered 'for a few months' or that they were seen 'a while ago'. Our parser attempted to translate these data into estimated dates in DVP DB V2, however, we decided that estimations would not be useful in practice and did not consider these attributes for DVP.

- **Trade Count and Rating** both relied on external factors more than on the vendors themselves. In other words, the number of trades a vendor made on a particular marketplace is likely dependent on the popularity of the marketplace. Thus, in order to incorporate trade count data as a feature, our model would have to learn it's relative trade count (high, average, low) in relation to the popularity of the marketplace the vendor is selling on. Likewise, a vendor rating may be more dependent on a vendor's customer base than on the vendor themselves. Through manual analysis, we determined most vendors were rated with the highest possible value. Therefore, even if ratings were considered in DVP, they would not add much value in distinguishing vendor accounts.

6.2 DVP Tasks

Again, we define vendor attribution as the task of determining the owner of a product listing given data on previous listings and their vendors. The features we

consider for vendor attribution are extracted *per listing* and fed into our Random Forest classifier. So, the only features available for vendor attribution are those that reflect individual listings. It is important to note that the Random Forest algorithm requires all data to be in numeric form. Thus, each of our categorical features are either one-hot encoded or hash encoded such that they may be interpreted by the Random Forest algorithm without assigning weights or ranks to the data.

We define alias attribution as the task of determining the degree of similarity between pairs of vendors. The features we consider for alias attribution are extracted *per vendor* and used to compute the similarity metric for each vendor pair in a dataset. For this reason, we are able to extend the feature set used in vendor attribution to gain a holistic view of a vendor and reflect several listings in a single vendor profile.

As discussed in Chapter 4, the first focus of this thesis is to demonstrate the effectiveness of using our proposed feature set and a Random Forest classifier for vendor attribution in dark marketplaces. We argue that success in vendor attribution is a strong indicator of success in alias attribution using the same feature set because it demonstrates the ability for a vendor to be distinguished by their product listings alone. It is also reasonable to suspect that alias attribution can be further improved by including additional features that are extracted in the same manner as in vendor attribution but from the unique data sources that are present only in vendor profile pages such as vendor descriptions and PGP keys. Thus, we propose an extended feature set derived from the vendor attribution model which can be used to determine the degree of similarity between vendors. Also, we evaluate the effectiveness of using our extended feature set in training record-linking SVM and LR models for vendor alias attribution.

CHAPTER 7
DARK VENDOR PROFILING EVALUATION

To evaluate the effectiveness of our approach for vendor and alias attribution in dark marketplaces, we designed a series of experiments using a variety of data subsets derived from DVP DB V4 and V5. Namely, we crafted experiments to evaluate the performance, time complexity, and memory usage of the DVP models based on the number of vendors, marketplaces, and listings per vendor used in training and testing phases. Furthermore, we examined the effect of using complete and incomplete data as well as various feature subsets. In this chapter, we present the results of these experiments.

7.1 Vendor Attribution

Since the majority of related work focuses on single marketplace experimentation, we chose to evaluate our vendor attribution model on single marketplaces as well using our top three most informative marketplaces, Abraxas, Agora and Evolution. We consider these marketplaces to be most informative since they have more listings and vendors than all other marketplaces in both DVP DB V4 and V5. For each of these three marketplaces, we crafted datasets from both DVP DB V4 and V5 made up of their top 5, 10, 15, and 20 vendors using 5, 10, 25, 50, 75, and 100 listings per each vendor. Furthermore, we created three datasets comprised of all vendors in the individual marketplaces who had 100 listings available. This experimental design resulted in 150 total experiments for single marketplace vendor attribution evaluation. Notably, the Abraxas marketplace did not have enough vendors with 100 listings to develop a model for 20 vendors. Therefore, the model was built with 17 vendors instead. We followed similar strategy for other such similar cases, thus our evaluation

DVP DB	Marketplace	Vendor Count	Listings per Vendor
Version 4	Abraxas	5	5, 10, 25, 50, 75, 100
		10	5, 10, 25, 50, 75, 100
		15	5, 10, 25, 50, 75, 100
		20	5, 10, 25, 50, 75, 100
		All (57)	100
	Agora	5	5, 10, 25, 50, 75, 100
		10	5, 10, 25, 50, 75, 100
		15	5, 10, 25, 50, 75, 100
		20	5, 10, 25, 50, 75, 100
		All (205)	100
	Evolution	5	5, 10, 25, 50, 75, 100
		10	5, 10, 25, 50, 75, 100
		15	5, 10, 25, 50, 75, 100
		20	5, 10, 25, 50, 75, 100
		All (103)	100
Version 5	Abraxas	5	5, 10, 25, 50, 75, 100
		10	5, 10, 25, 50, 75, 100
		15	5, 10, 25, 50, 75, 100
		20	5, 10, 25, 50, 75, 100
		All (17)	100
	Agora	5	5, 10, 25, 50, 75, 100
		10	5, 10, 25, 50, 75, 100
		15	5, 10, 25, 50, 75, 100
		20	5, 10, 25, 50, 75, 100
		All (39)	100
	Evolution	5	5, 10, 25, 50, 75, 100
		10	5, 10, 25, 50, 75, 100
		15	5, 10, 25, 50, 75, 100
		20	5, 10, 25, 50, 75, 100
		All (61)	100

Table 7.1: Summary of experiments ran to evaluate DVP-based vendor attribution in individual marketplaces.

was not impacted. Table 7.1 summarizes the experiments used to evaluate DVP in single marketplace settings.

Since our primary goal was to attain vendor attribution across many marketplaces to show how DVP can be used in practice, we additionally attempted to craft datasets comprised of 1, 5, 10, and 20 vendors with 25, 50, 75 and 100 listings from *every* marketplace available in DVP DB V4 and V5. However, in many cases, our marketplaces did not have enough vendors with enough listings to create the datasets

we had considered in our experimental design. This meant that (a) not every marketplace in the DVP DB could be considered in every multi-market experiment and (b) not all marketplaces could contribute exactly 1, 5, 10, and 20 vendors to the datasets. This variation of marketplace representation was expected, however, since the size and popularity of dark marketplaces varies greatly in practice. To compensate, our code extracts *up to* 1, 5, 10, and 20 vendors per marketplace and results in datasets with data from anywhere between 10 to 25 marketplaces. Additionally, since the performance of our DVP model relies more on the number of vendors used in training than it relies on the number of marketplaces the vendor data is coming from, we refocus our experiments to evaluate the effect of the number of vendors our model considers regardless of how many marketplaces their data is extracted from. Since our experiment is still able to compare the results of single marketplace analysis versus multi-marketplace analysis, we do not consider this a limitation. This experimental design resulted in 32 experiments for multi-market vendor attribution evaluation which is summarized in Table 7.2.

DVP DB	**Listings per Vendor**	**Number of Vendors**
Version 4	25	24, 102, 180, 297
	50	20, 80, 132, 216
	75	17, 63, 107, 163
	100	15, 53, 83, 132
Version 5	25	14, 67, 118, 196
	50	12, 48, 83, 144
	75	11, 42, 71, 102
	100	10, 35, 53, 80

Table 7.2: Summary of experiments ran to evaluate DVP-based vendor attribution across multiple marketplaces ranging from to 10 to 25 marketplaces.

Random Forest Hyperparameter Tuning

We began our evaluation of DVP vendor attribution with the single marketplace datasets. Since these datasets were significantly smaller than multi-market datasets, we chose to utilize them for hyperparameter tuning tests without exhausting memory and time resources. In addition to building each single-market-based random forest model with default parameters, we performed a random grid search to build a model with optimal hyperparameters. For each single market model, the random grid search selected 100 random combinations of the parameters detailed in Table 7.3 and determined the combination of hyperparameters resulted in the highest model accuracy. Then, out of the 150 models we developed, we determined which parameters most often resulted in the highest accuracies. Interestingly, the 'best' parameters varied between DVP DB versions. The results are shown in Table 7.4.

Parameter	**Description**	**Values Tested**
Number of Estimators	The amount of decision trees in the forest	25, 81, 137, 193, 250
Minimum Sample Split	The minimum number of samples required to split a leaf node	2, 5, 10
Minimum Samples Leaf	The minimum number of samples required to be at a leaf node	1, 2, 4
Maximum Features	The number of features to consider when looking for the best split	Auto, Sqrt
Maximum Depth	The maximum depth allowed for each decision tree	25, 81, 137, 193, 250, None
Boostrap	Whether or not the entied dataset is used to build each tree	True, False

Table 7.3: Description of hyperparameters tested during single marketpace experimentation.

Model Version	**Estimators**	**Min. Sample Split**	**Min. Samples Leaf**	**Max. Features**	**Max. Depth**	**Boostrap**
Default	100	2	1	Auto	None	True
Version 4	160	3	1	Auto	105	False
Version 5	152	4	1	Auto	131	False

Table 7.4: Results of hyperparameter tuning in single-market-based models.

Since the random grid search can be resource demanding, we did not rerun hyperparameter tests for multi-market-based models. Instead, the results of the single market hyperparameter tuning tests were utilized for multi-market experimentation by having each dataset train a *base* model with default parameters and a *best* model with the optimal parameters as determined by the single market tests.

Model Accuracy

Our models were primarily evaluated for their accuracy in distinguishing dark vendors and determining their listings. For each dataset, our models were trained on 75% of the given data and were tested on the remaining 25%. First, we examined the results of single marketplace experimentation.

We took the maximum accuracy of each model, whether it was using the base parameters or the tuned parameters, and measured the average, maximum, and minimum accuracies for Abraxas, Agora, and Evolution models. These results are presented in Table 7.5 and show that models trained on version 5 data generally attained higher accuracies than models trained on version 4 data. We relate this behavior to the completeness of listings in DVP DB V5. However, we also argue that while incomplete listings resulted in lower model accuracy, models trained on this data still attained an acceptable level of performance illustrating the robustness of DVP in dealing with missing data unlike other related work models.

Furthermore, we calculated the statistical correlation between model accuracy and (a) number of vendors, (b) number of listings per vendor, and (c) model complexity, where model complexity is defined as the *number of rows X number of columns* in the dataset used to train the model. This way, we were able to determine which variable had the greatest impact on model accuracy. As shown in Table 7.6, despite

DVP DB	Marketplace	Average	Maximum	Minimum
Version 4	Abraxas	89.69%	100%	63.16%
	Agora	94.77%	100%	69.23%
	Evolution	89.61%	97.99%	63.16%
Version 5	Abraxas	94.48%	100%	57.14%
	Agora	97.21%	100%	83.33%
	Evolution	91.91%	100%	64.00%

Table 7.5: Overall model accuracy results of single marketplace experimentation.

having a relatively weak positive correlation, the number of listings per vendor had the greatest impact on increasing the accuracy of both DVP DB V4 and V5 models.

	Version 4 Model Accuracy	Version 5 Model Accuracy
Number of Vendors	0.0530	0.0840
Number of Listings per Vendor	0.3831	0.5413
Model Complexity	0.1280	0.1110

Table 7.6: Correlations between single market model *accuracy* and three variables: (a) number of vendors, (b) number of listings per vendor, and (c) model complexity.

These results suggest that the DVP model accuracy does not rely on the number of vendors or model complexity. This is an exceptionally significant finding because it indicates that DVP is scalable as long as there is sufficient listing data for each vendor. Next, we needed to determine how many listings per vendor would be needed to have sufficient training data. Figures 7.1 and 7.2 investigate this task for both database versions, finding that most models achieved high accuracy as long as they were trained with over 25 listings per vendor. Again, we note that version 5 models generally performed better than version 4 models. However, these results reflect single marketplace experimentation only, therefore our natural next step was to evaluate the results of our models trained on multi-marketplace data in the same manner.

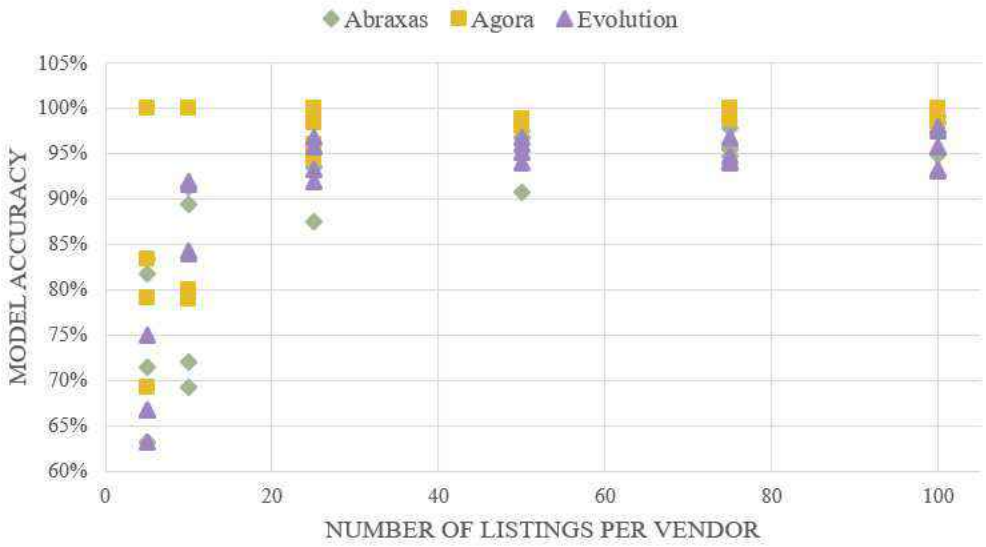


Figure 7.1: Model Accuracy vs. Number of Listings per Vendor results for DVP DB V4 single marketplace experimentation.

Similar to before, we trained our multi-market models on 75% of the given data and tested their accuracies on the remaining 25%. Further, we demonstrate how version 5 models are generally more accurate than version 4 models but conclude that all models achieve remarkable accuracy. Table 7.7 lists the overall accuracy results of multi-market experimentation. Interestingly, all multi-market models achieve higher accuracy than found in single market experimentation. However, we relate this to the experimentation of 5 and 10 listings per vendor models with single markets. In other words, since single market models were tested with datasets containing fewer listings per vendor and multi model market models were not, the average results of single market models are expected to be lower.

In the same fashion as before, we calculated the statistical correlation between multi-market model accuracies and related variables. The results of this analysis are listed in Table 7.8. Unlike single market models, we see a strong negative correlation

Figure 7.2: Model Accuracy vs. Number of Listings per Vendor results for DVP DB V5 single marketplace experimentation.

DVP DB	Average	Maximum	Minimum
Version 4	95.74%	100%	91.06%
Version 5	98.13%	100%	94.69%

Table 7.7: Overall model accuracy results of multi-marketplace experimentation.

between the model accuracy and the number of vendors used in training and testing as well as a weak negative correlation between model accuracy and model complexity. This trend is illustrated linearly by Figure 7.3 and illustrates that model accuracy generally decreases as more distinct vendors are used in training and testing. We presume model complexity correlation mimics number of vendor correlation since model complexity is dependent on the number of vendors present in our training datasets. We also presume that strong correlation is exhibited in multi-market experimentation and not single market experimentation because multi-market models were trained on

a balanced range of up to 297 vendors, whereas single market models were trained mostly on 5 to 20 vendors with few exceptions (one Agora-based model was trained on 39 vendors with 100 listings each and one Evolution-based model was trained on 61 vendors with 100 listings each). Thus, if the same amount of vendors were considered in single market experimentation as in multi-market experimentation, we hypothesize that a similar strong negative correlation would exist between model accuracy and number of vendors considered.

	Version 4 Model Accuracy	**Version 5 Model Accuracy**
Number of Vendors	-0.8960	-0.9376
Number of Listings per Vendor	0.3242	0.4987
Model Complexity	-0.6498	-0.4607

Table 7.8: Correlations between multi-market model accuracy and three variables: (a) number of vendors, (b) number of listings per vendor, and (c) model complexity.

Table 7.4 shows that the weak positive correlation between single market model accuracy and the number of listings per vendor used is similarly present for multi-market models. As depicted by Figure 7.4, there is limited variation in accuracy based on the number of listings per vendor available regardless of how many vendors are considered in the model. In fact, for both version 4 and version 5 based models, there is only about a 2% difference in average accuracy between models that were trained with 25 listings per vendor and 100 listings per vendor.

Surely, if we were to consider more listings per vendor, we would be able to represent each vendor better. While this leads to more accurate models, requiring vendors to have 100 listings in order to be considered for DVP analysis is somewhat restrictive since the majority of dark vendors do not have 100 listings available. In fact, in DVP DB V4, vendors have an average of 20 listings, and in DVP DB V5, vendors have an average of 14. Therefore, for DVP to be used in practice, the trained

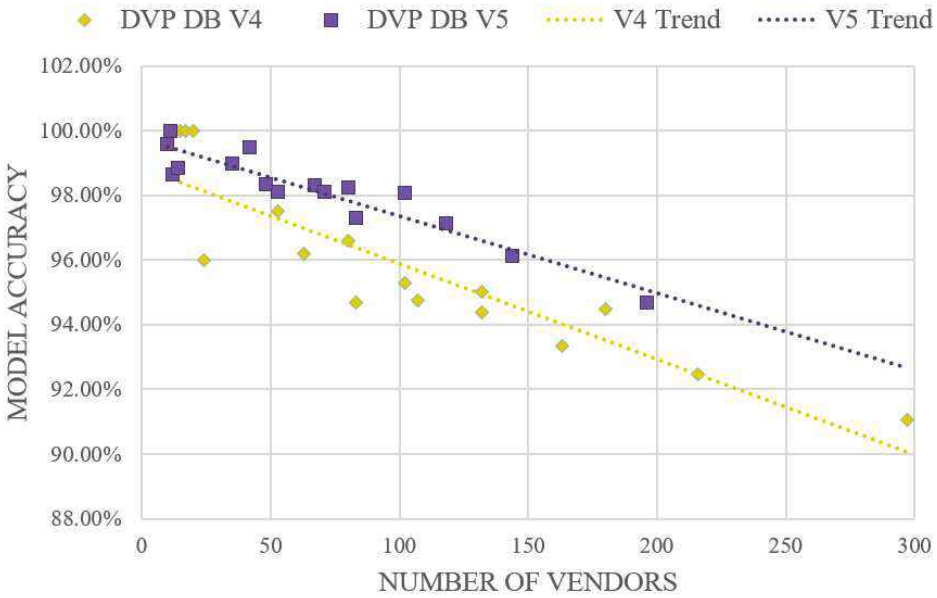


Figure 7.3: Model Accuracy vs. Number of Vendor results for DVP DB V4 and V5 multi-market experimentation.

model would need to strike a balance between listing count requirements and model performance.

After running each of the aforementioned experiments, we decided to test a more practical version of multi-market vendor attribution. Thus, we designed 8 more multi-market experiments with *all* vendors having 25, 50, 75, or 100 listings in DVP DB versions 4 and 5 were used in training the models as opposed to only considering up to 1, 5, 10, and 20 from each marketplace. These experiments can be considered *full* DVP experiments since they consider all possible vendors. These additional experiments are summarized in Table 7.9. With these experiments, we can clearly observe how increasing the number of listings required in order to be considered for DVP is restrictive and results in smaller sets of vendors.

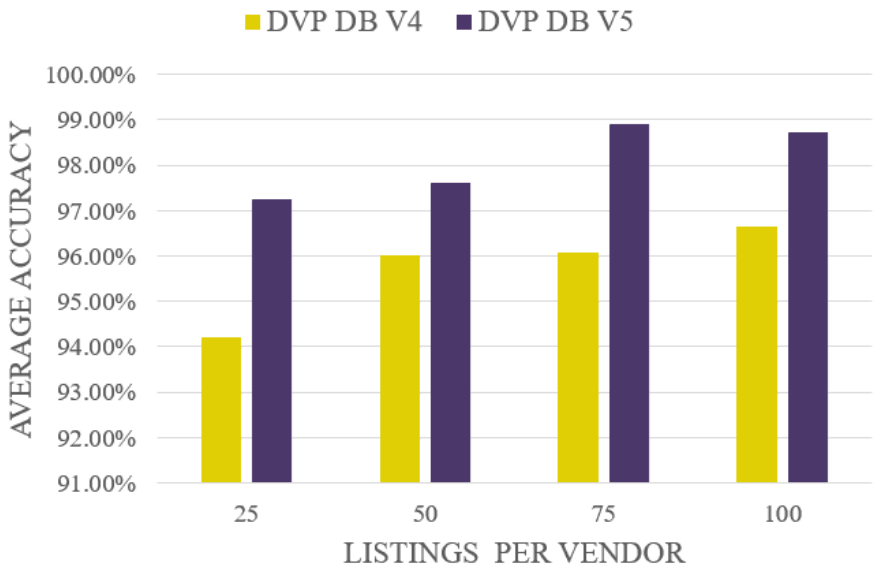


Figure 7.4: Model Accuracy vs. Number of Listings per Vendor results for DVP DB V4 and V5 multi-market experimentation.

DVP DB	Listings per Vendor	Number of Vendors
Version 4	25	3240
	50	1401
	75	745
	100	446
Version 5	25	1501
	50	525
	75	250
	100	140

Table 7.9: Summary of full DVP experiments ran to evaluate multi-market vendor attribution.

Once again, we can see that the model accuracies have a negative relationship with the number of vendors considered during training. However, we exhibit the robustness of DVP vendor attribution by demonstrating how our models maintain above 88% accuracy as long at least 25 listings per vendor are used in training. Figure

7.5 illustrates these results and depicts their linear trends. Interestingly, this figure suggests that DVP DB V4 models are more robust to increasing numbers of vendors than DVP DB V5 models as depicted by the slope of the V4 trend line being less steep than the slope of the V5 trend line. However, since we (a) cannot confirm that the most reliable trend line is a *linear* trend (as opposed to polynomial, logarithmic, or exponential) and (b) do not have a sufficient amount of data points to extrapolate our data with high confidence (there are only four data points for each line), we do not conclude that these trends accurately predict the accuracies of models trained with more vendors. Thus, future work may consider further experimentation with larger sets of vendors.

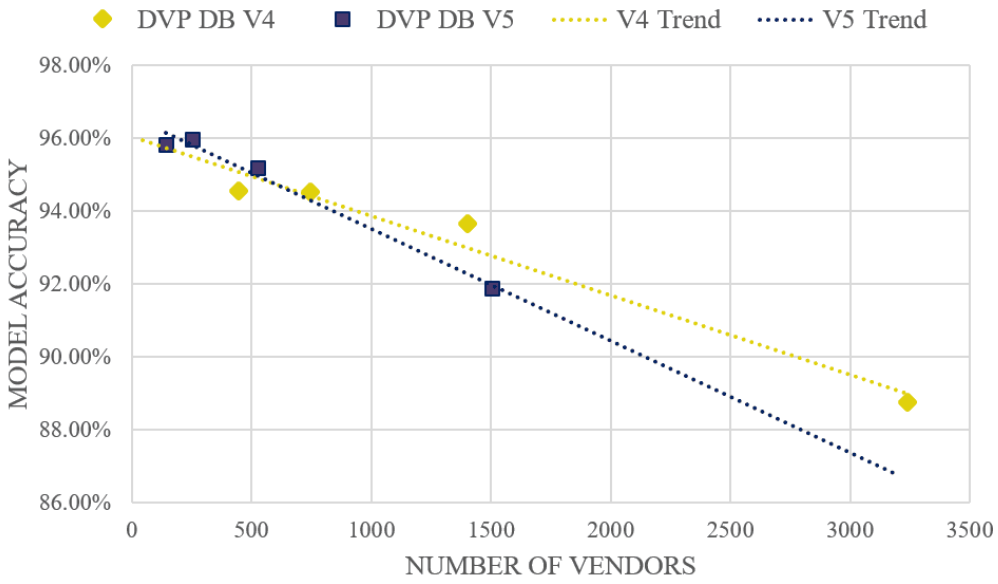


Figure 7.5: Model Accuracy vs. Number of Listings per Vendor results for *full* DVP DB V4 and V5 multi-market experimentation.

Time Complexity and Memory Usage

Next, we wanted to evaluate how the time to extract features and run the vendor attribution model was effected by the same variables that we examined for model accuracy. Thus, we ran the same statistical correlation test on the original 32 multi-market experiments to evaluate the relationships between feature extraction and model execution times and the number of vendors, number of listings per vendor, and model complexity. The results in Table 7.10 show that there is a strong positive correlation between feature extraction time and model complexity as well as a weaker positive correlation between model execution time and model complexity. These trends are illustrated in Figure 7.6 using DVP DB V4 results. We chose to demonstrate these trends using the version 4 results since they considered more complex models. However, we do note that the version 5 tests resulted in nearly identical trends but with smaller datasets. From this graphic, we can conclude that the Random Forest algorithm is robust to increasing dataset sizes in terms of time complexity, but that the time to complete the feature extraction process may grow exponentially as we consider more vendors and listing data. This is an expected limitation of the current DVP implementation but can likely be improved with code optimization techniques, parallelization, or improved hardware, all of which may be considered for future DVP work.

	Version 4		**Version 5**	
	Feature Extraction	Model Execution	Feature Extraction	Model Execution
Number of Vendors	0.4966	0.7798	0.5341	0.5267
Number of Listings per Vendor	0.3330	0.0926	0.3203	0.0205
Model Complexity	0.9906	0.9496	0.9943	0.4783

Table 7.10: Correlations between multi-market feature extraction and model execution *time* and (a) number of vendors, (b) number of listings per vendor, and (c) model complexity.

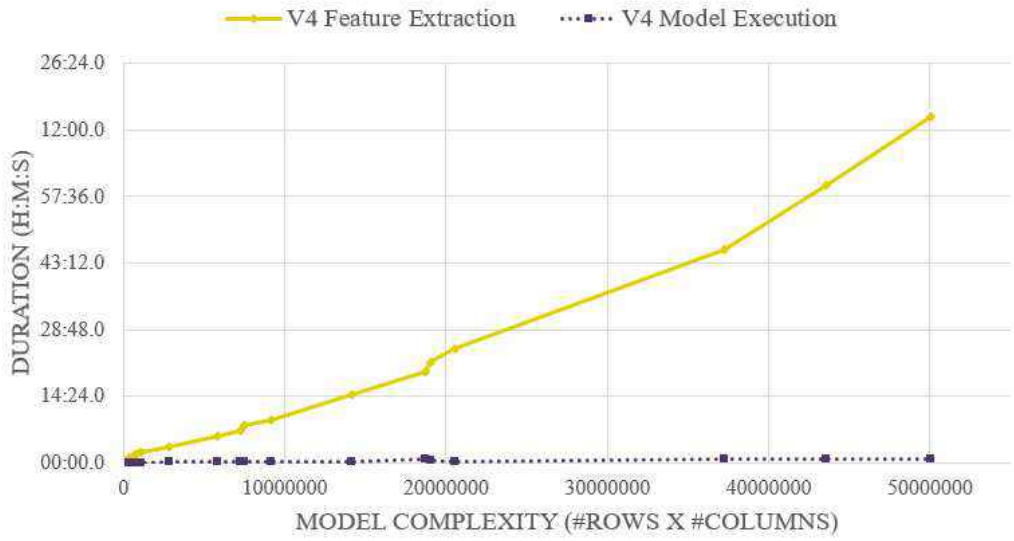


Figure 7.6: Time vs. Model Complexity results for DVP DB V4 multi-market experimentation.

Likewise, we conducted the same evaluation for the average memory use during feature extraction and model execution. We believe this is an important evaluation for this particular domain because in order for DVP to be applied in practice, it should not rely on extensive computing resources. Again, we use data from the original 32 multi-market experiments to calculate correlations between the three variables and average memory usage. The results listed in Table 7.11 again suggest a strong positive correlation between memory usage and model complexity. This trend is illustrated by Figure 7.7. Notably, this graphic shows how the feature extraction process is more robust in terms of memory usage than model execution. This is the opposite of what we saw in the time complexity graphic in Figure 7.6. Therefore, we can conclude that the Random Forest algorithm is *not* robust to increasing dataset sizes in terms of memory usage, but that the memory required to complete the feature

extraction process is. This is a limitation to using the Random Forest algorithm for large numbers of vendors.

	Version 4		**Version 5**	
	Feature Extraction	Model Execution	Feature Extraction	Model Execution
Number of Vendors	0.5606	0.8035	0.6538	0.7961
Number of Listings per Vendor	0.0635	0.0926	0.2710	0.0640
Model Complexity	0.9983	0.9363	0.9790	0.9492

Table 7.11: Correlations between multi-market feature extraction and model execution *average memory usage* and (a)number of vendors, (b) number of listings per vendor, and (c) model complexity.

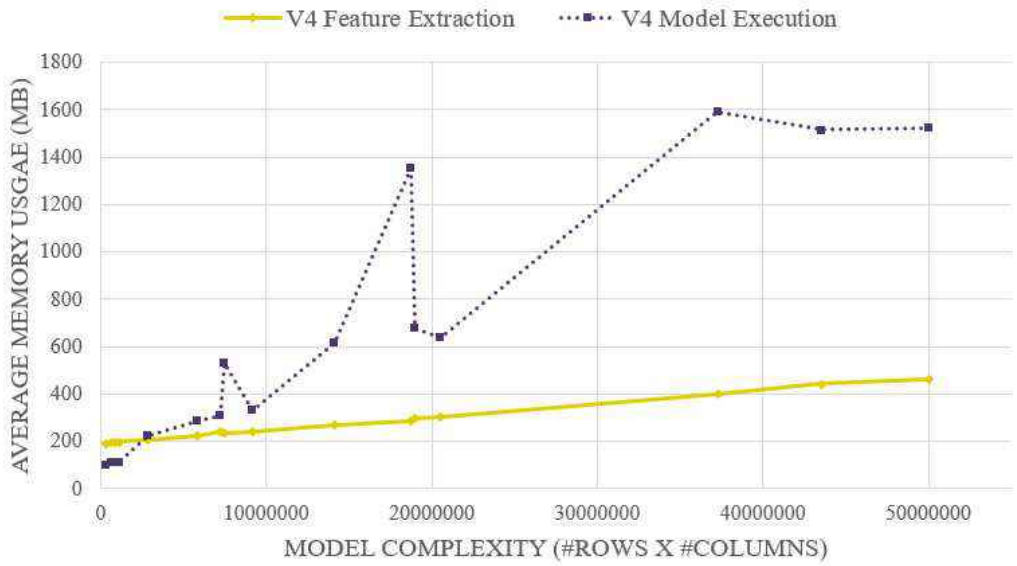


Figure 7.7: Average Memory Usage vs. Model Complexity results for DVP DB V4 multi-market experimentation.

In the original 32 multi-market experiments, we considered up to 297 vendors at a time, but in the 8 *full* DVP experiments, we considered up to 3,240. Thus, we expected the full DVP experiments to be far more expensive in terms of time and

memory. With the assumption that full DVP would be more practical to be used by law enforcement and intelligence agencies, we continued our evaluation of time and memory usage on full DVP models. The results are provided in Table 7.12. Notably, our largest model was able to attribute 3,240 distinct vendors with 88.76% accuracy. However, these results were achieved after using approximately 4.5 days worth of time and up to nearly 114 GB of memory at one point in time. These findings are very significant since they demonstrate how DVP in practice achieves high accuracy at the cost of expensive time and memory resources. Once more, these results demonstrate how the DVP task requires an effective balance between the number of vendors being analyzed and the time and memory resources we are willing to exhaust.

DVP DB	**Number of Vendors**	**Average Feature Extraction Memory**	**Maximum Feature Extraction Memory**	**Feature Extraction Time**	**Average Model Execution Memory**	**Maximum Model Execution Memory**	**Model Execution Time**
Version 4	446	3.03	9.80	25:48:26	11.06	15.17	00:12:32
	745	4.09	13.38	44:50:55	16.40	20.94	00:22:02
	1,401	5.47	18.86	73:58:26	29.65	37.71	00:37:59
	3,240	7.06	23.01	108:50:02	79.97	113.35	00:54:23
Version 5	140	0.59	1.35	1:17:38	1.51	1.98	00:01:07
	250	0.86	2.11	02:53:44	2.92	3.97	00:02:38
	525	1.39	4.74	07:03:08	6.72	9.31	00:07:16
	1,501	2.43	8.13	17:07:13	19.03	26.54	00:19:04

Table 7.12: Full multi-market DVP results in terms of time complexity (H:M:S) and memory usage (GB).

Feature Subsets

Finally, we evaluated the effect of using subsets of the features proposed in this thesis to train models. This way, we were able to verify if using a combination of stylometry, attribute, and image based features provides the most effective results for

DVP. In an effort to strike a balance between computational resources and model complexity, we experimented using different feature subsets to train multi-market models with datasets derived from up to 20 vendors per marketplace with 25 listings each. The results of this analysis are presented in Table 7.13. Evidently, the combination of all three feature sets resulted in the most accurate models. However, we note that the models trained on both stylometry-based and attribute-based features achieve similar accuracy. This suggests that the image-based features are the least important of the three feature sets. Therefore, in future work, we may consider modifying image-based features to provide more value to the DVP classification scheme. Additionally, future work may consider more thorough feature subset experimentation on datasets made up of varying amounts of vendors and listings per vendor.

DVP DB	**Feature Set**	**Maximum Accuracy**
Version 4	Image	58.83%
	Attribute	78.78%
	Stylometric	79.59%
	Image, Attribute	83.04%
	Image, Stylometric	80.45%
	Attribute, Stylometric	90.31%
	All	**91.06%**
Version 5	Image	59.72%
	Attribute	84.97%
	Stylometric	87.75%
	Image, Attribute	84.80%
	Image, Stylometric	88.07%
	Attribute, Stylometric	94.36%
	All	**94.69%**

Table 7.13: Summary of feature subset experiments ran on models trained with datasets containing up to 20 vendors per marketplace with 25 listings each.

We also note that the image-based features had more of an impact in version 4 experiments than they did in version 5 experiments. For example, for version 4 experiments, accuracy increased by approximately 4.5% when image features were used

in conjunction with attribute features instead of attribute features alone. In version 5 experiments, the accuracy was not improved at all for this scenario. Likewise, version 4 accuracy increased approximately 1% when all features were considered rather than just attribute and stylometric features alone. In version 5, this increase was only 0.33%. This supports our initial hypothesis that combining feature types may be advantageous in applications where data is often incomplete or missing. Thus, we conclude that image-based features are still valuable additions.

7.2 Alias Attribution

In this research, we define alias attribution as the task of determining the degree of similarity between two vendors based on their listings and vendor page data and experiment with two attribution models: elementary style and enhanced. Using each of the features considered in vendor attribution as well as a few additional features derived from a vendor's overall behavior, we consolidated each unique vendor's data into a single vector. Then, in elementary style alias attribution, we used the cosine similarity metric to compute a similarity measure for every pair of vendor vectors in a dataset. Specifically, for every pair of vendors, we calculated a cosine similarity measure between each of their individual features and used these values to determine an overall *average* similarity.

We acknowledge three problems that arise from using such a simplistic approach. First, using an average calculation is very elementary in that it considers each feature to have equal importance in determining the overall similarity. Thus, in our enhanced model, we determine overall vendor similarity using a *weighted* average instead so that less important features may have less of an impact on the final similarity measure than more important features. Second, calculating *every* pair of

vendors becomes an immense task when considering a large number of vendors. For example, to calculate the similarities between all pairs of 500 vendors, we would need to calculate overall cosine similarities for 124,750 vendor pairs. If our dataset grew to 5,000 vendors, we would need to calculate similarities for 12,497,500 vendor pairs. Therefore, to reduce the amount of calculations needed to accomplish alias attribution, the enhanced alias attribution model implements a vendor pairing scheme where pairs are only constructed for vendors who share a common attribute instead of pairing all vendors with each other. Third, we have no ground truth available to evaluate the performance this approach. Thus, we cannot confirm the reliability of this approach and instead only use it to make preliminary observations of our data. This is perhaps the most impeding limitation which we hope to improve upon in the enhance model.

With these limitations in mind, we still examined how our proposed feature set fared in elementary-style alias attribution. We set up three series of experiments to test our system. In the first set of experiments, we took every vendor from the Abraxas, Agora, and Evolution marketplaces who had 75 listings and tested single market alias attribution in which we sought to find aliases within a single marketplace. Additionally, we took every vendor from *any* marketplace who had 75 listings and tested multi-market alias attribution in which we sought to find aliases within and between marketplaces. In the second set of experiments, we attempted the same approach as above but used 25 listings instead of 75. Finally, in the third set of experiments, we eliminated the requirement of having a certain amount of listings and attempted *full* alias attribution by considering all vendors having *any* number of listings. The final set of experiments resulted in much larger datasets than the other two and therefore were much more time consuming to run.

Upon completion, we had cosine similarity values for every pair of vendors from each dataset. To analyze our preliminary results and examine which vendors would be considered aliases using our approach, we selected a threshold value of .90 such that two vendors would have to be at least 90% similar (as determined by their cosine metric) in order to be considered aliases. Each of these experiments used both DVP DB V4 and V5 data and are summarized by Table 7.14 along with the resulting number of alias pairs they discovered.

DVP DB	Marketplace	Listings per Vendor	Unique Vendors	Vendor Pairs	Alias Pairs
Version 4	Abraxas	75	89	3,916	0
		25	329	53,956	1
		Any	1,138	646,953	13
	Agora	75	338	56,953	3
		25	1,232	758,296	16
		Any	3,149	4,956,526	133
	Evolution	75	187	17,391	0
		25	862	371,091	1
		Any	3,921	7,685,160	117
	All	75	745	277,140	36
		25	3,240	5,247,180	130
		Any	15,160	114,905,220	*incomplete*
Version 5	Abraxas	75	30	435	0
		25	162	13,041	1
		Any	850	360,825	5
	Agora	75	73	2,628	0
		25	399	79,401	4
		Any	2,047	2,094,081	35
	Evolution	75	105	5,460	0
		25	580	167,910	0
		Any	3,310	5,476,395	38
	All	75	250	31,125	8
		25	1,501	1,125,750	40
		Any	10,520	55,329,940	*incomplete*

Table 7.14: Summary of experiments ran to evaluate DVP-based alias attribution. **Incomplete experiments are awaiting results due to the time required to analyze large numbers of potential alias pairs.*

Elementary Style Results

As illustrated in Table 7.14, version 4 datasets contained more vendors and resulted in more alias pairs. Since the number of vendors greatly varies between datasets and we do not have precision and recall rates for any of the datasets, we could not make any conclusions on the effect of missing data or varying numbers of listings per vendor on our approach. Therefore, implementing a *weighted* similarity metric and conducting a thorough evaluation of our approach was our immediate next research direction. However, from our preliminary results, we could conclude that (a) more inclusive datasets resulted in larger numbers of alias pairs, and (b) the majority of alias pairs found using our approach shared very similar pseudonames of the vendors which led us to suspect the identified pairs were in fact true positives. For example, Table 7.15 lists several version 5 single market vendor aliases with at least 90% similarity found using *any* number of listings per vendor. From these results, we observe highly similar pseudonames in the words they contain, the way they capitalize letters, and their use of punctuation.

Further, Table 7.16 lists several version 4 multi-market alias pairs found using 75 listings per vendor. This time, we observed exact pseudoname matches across various marketplaces as well as several similar pseudoname matches. The overall results of these experiments are significant for several reasons. First, the majority of multi-market alias pairs had exact vendor name matches and the majority of single market pairs had very closely related pseudonames. This suggests that vendor names are frequently reused among dark marketplaces. If we assume vendor names are reused only by the same individual, simply searching for similar pseudonames could be sufficient for alias attribution rather than crafting vendor profiles with extensive feature sets. However, it is possible for different vendors to use the same username, which

Marketplace	Vendor Name A	Vendor Name B	Similarity
Abraxas	fakeasanything	flawlessfakeids	0.9586
	fakeasanything	fake	0.9659
	flawlessfakeids	fake	0.9590
	Mountain	GreenMountainMan	0.9039
	indiabenzos_ib	indiabenzos	0.9322
Agora	Colorado	Colorado_Fantasy	0.9832
	UKPharma	UKPharmaceuticals	0.9620
	only	theonlysource	0.9296
	GotTheProduct	TheProduct	0.9628
	kingofcokeman	cokeman	0.9657
Evolution	TungstenGold	Goldenman	0.9730
	TheBitCoinGuru	Coin	0.9322
	monkey	howlingmonkey	0.9954
	crookscastle710	rook	0.9421
	Baron	Baron-JOY	0.9179

Table 7.15: Example version 5 single market alias pairs with at least 90% similarity found using DVP-based alias attribution and *any* number of listings per vendor.

Marketplace A	Vendor Name A	Marketplace B	Vendor Name B	Similarity
Agora	Peaceful	Abraxas	Peaceful	0.9629
	RepAAA		RepAAA	0.9470
	COFFEESHOP24		COFFEESHOP24	0.9306
	Righteous		Righteous	0.9547
	TheLeanZebra		The LeanZebra	0.9234
Evolution	StealthBomber	Agora	StealthBomber	0.9179
	SaltnPepper		SaltnPepper	0.9235
	Toyota		Toyota	0.9221
	ThreeKings		ThreeKings	0.9211
	Shiny-Flakes		Shiny-Flakes	0.9195
Abraxas	Drugs4you	Evolution	Drugs4you	0.9054
	GoingPostal		GoingPostal	0.9107
	InsideTheWhale		InsideTheWhale	0.9168
Pandora	RedBull	Hydra	RedBull	0.9110

Table 7.16: Example version 4 multi-market alias pairs with at least 90% similarity found using DVP-based alias attribution and 75 listings per vendor.

would make simple pseudoname searches insufficient for detecting aliases accurately. For example, as shown in Figure 7.17, the elementary approach found that many vendors with the same name were found to be only 50-60% similar which suggests that these users are *not* true aliases. Further, it is also possible that our results are

Marketplace A	Vendor Name A	Marketplace B	Vendor Name B	Similarity
Alphabay	Fake	MiddleEarth	Fake	0.5839
		Oxygen		0.5776
		Agora		0.5753
		Cryptomarket		0.5916
		Abraxas		0.5764
Abraxas	indianpilldaddy	Pandora	indianpilldaddy	0.5687
Evolution	etimbuk	Cryptomarket	etimbuk	0.5892
Black Bank Market	COLOR	Alphabay	COLOR	0.5942

Table 7.17: Example version 4 multi-market alias pairs with the same names but only 50-60% similarity found using DVP-based alias attribution and 75 listings per vendor.

misrepresenting the frequency of pseudoname reuse and that our elementary-style approach is not effective enough to identify vendors who deliberately obfuscate their alias identities. Second, despite our vendor attribution experimentation indicating that vendors should ideally be profiled from 25-75 listings, such high listing counts are too restrictive in practice and result in practical alias attribution in large marketplaces only. This is evident in our results since the majority of aliased vendors belonged to one of our three largest marketplaces: Abraxas, Agora, or Evolution. Third, calculating vendor similarities becomes an extremely time consuming process when every possible vendor combination is considered as potential aliases. While this is a comprehensive approach, it is not practical due to time demands. In fact, with this elementary approach, calculating a similarity metrics took about 0.1 seconds per vendor pair, so with high numbers of candidate pairs, some of our experiments had to run for days before presenting results.

Enhanced Vendor Linkage Results

In the enhanced alias attribution model, we address the short comings of the elementary style approach. Namely, we *do not* restrict vendors by requiring them to

have 25 listings but *do* restrict the number of potential alias pairs our models consider by only pairing vendors who share a common attribute. In particular, we train and test our models with data from all vendors who have at least 10 product listings and one vendor profile page in the DVP DB. While our vendor attribution results indicated vendors should have at least 25 listings to be well represented, we believe it is reasonably acceptable to trade off minor performance loss for higher scalability. Further, candidate pairs are formulated in the enhanced model by matching vendors who either share their (a) top ships from *and* ships to locations, (b) top categories *and* subcategories, (c) exact same (case-sensitive) vendor name, (d) top base product name, (e) top PGP key, or (f) top PHASH. Additionally, we compute vendor comparison vectors in parallel such that ten processes are calculating results at once. By incorporating parallelization and improving our vendor comparison technique, we are able to reduce model training time to approximately 0.013 seconds per candidate vendor pair as opposed to the 0.1 seconds it took in elementary style attribution. We test our enhancements with a DVP DB V5 dataset, however, plan to experiment with a DVP DB V4 dataset in future work in order to evaluate our models' robustness against incomplete data.

As mentioned in Chapter 3, ground truth availability is generally lacking in the dark web research domain. This makes it difficult to train a model to identify vendor aliases in the wild because there exists no publicly available dataset of known aliases. To combat this challenge, we develop a synthetic ground truth by taking all vendors with at least 10 listings and splitting their listing data into two pseudo vendors who share the same pseudo vendor identification number. Thus, we label *true vendor alias pairs* as vendors who share the same pseudo vendor ID. Additionally, because of the results of elementary style alias attribution experimentation, we decided to label

additional alias pairs if the vendors shared the same case-sensitive vendor name. Using the candidate pair formulation technique and the labeled pseudo-vendor datasets, we trained our models to predict alias vendor pairs in the wild. Our results are a promising indication that our methodology can achieve a very high level of performance in comparison to the most recent state-of-the-art system for alias attribution in dark web marketplaces, *uStyle-uID* [60] especially since our alias attribution model considers more dark marketplaces and dark vendors at a time than in [60].

To experiment our methodology, we used a DVP DB V5 dataset comprised of 4,357 pseudo vendors, we defined *true pairs* as any pair that shared a pseudo ID or a case sensitive vendor name, we implemented candidate pair formulation, and we split our dataset to use 75% for training and 25% for testing the SVM and LR models. Our true pair definition resulted in 5,269 true vendor pairs and 1,784,223 candidate pairs which were later split into 3,991 true pairs and 1,338,167 candidate pairs for training, and 1,278 true pairs and 446,056 candidate pairs for testing. Our initial results are shown in Table 7.18.

Model	**TP**	**FP**	**FN**	**TN**	**Precision**	**Recall**	**F1 Score**	**Accuracy**
SVM	1,278	0	0	444,778	1.0	1.0	1.0	1.0
LR	1,278	0	46	444,778	0.9653	1.0	0.9823	0.9999

Table 7.18: Results of SVM and LR training *with* Vendor Name as a feature in terms of True Positives (TP), False Postivies (FP), False Negatives (FN), True Negatives (TN), Precision, Recall, F1 Score, and Accuracy.

While the results indicate incredibly high performance, we suspected they were misleading of how the models would fare in real world applications. Notably, we noticed how the models never predicted any false positives. This meant that the models were not attributing aliases whose vendor names were not exact matches.

But, as we demonstrated with the elementary style alias attribution, many of the originally predicted pairs consisted of vendors who had *similar* but not exact same vendor names. We suspect that since our training data consisted only of true pairs who shared the exact same vendor name, the models might have learned this pattern and over-weighted the importance of the *Vendor Name* feature. Nonetheless, we saved the models such that they could be reloaded in other programs and utilized to predict pairs from other datasets. Then, we simulated a real world application of these models by having them predict alias pairs from a DVP DB V5 dataset that did *not* consist of any pseudo vendors. Of the 384,029 candidate pairs considered, the SVM model predicted 782 alias pairs while the LR model predicted 832. Unfortunately, after qualitative analysis of the models' predictions, we noticed that both models rarely predicted vendor pairs who did not share equivalent vendor names. While the logistic regression model was able to pick up on a few more vendor pairs with non-equivalent names, neither of our models were performing well enough to be used in practice.

To verify our suspicion that our training data was causing the models to over-weight the importance of the vendor name feature, we trained another set of models with the same DVP DB V5 pseudo vendor dataset. This time, however, we removed vendor names from our feature set to encourage the models to rely on the other features instead. The results of this experiment are illustrated in Table 7.19. Evidently, vendor names feature played an important role in model performance as we note how the precision, recall, F1 score, and accuracy of both models degraded in the second experiment. Therefore, vendor names should be considered in the enhanced model feature set. However, to avoid the inaccuracy of our initial models while still incorporating the vendor name feature, we plan to improve our training dataset in future work such that it consists of exact vendor name matches, similar name matches, and completely unrelated name matches. While the process of labeling true vendor pairs

via manual analysis may be tedious and time consuming, we believe that it may significantly improve model performance.

Model	TP	FP	FN	TN	Precision	Recall	F1 Score	Accuracy
SVM	987	372	134	44,563	0.8805	0.7263	0.7960	0.9989
LR	1,024	335	132	444,565	0.8858	0.7535	0.8143	0.9990

Table 7.19: Results of SVM and LR training *without* Vendor Name as a feature in terms of True Positives (TP), False Postivies (FP), False Negatives (FN), True Negatives (TN), Precision, Recall, F1 Score, and Accuracy.

Again, we saved the second set of trained models and evaluate their predicted pairs in a real world simulation. Interestingly, while the original SVM and LR models predicted 782 and 832 pairs in the wild, the second set of models predicted 891 and 873 pairs respectively and predicted many more alias pairs who did not share exact vendor names. These results are significant for a number of reasons. First, the spike in false positives tells us that when the models are not trained to match vendors with equivalent names, the models are able to pick up on potential aliases they would have otherwise overlooked. Second, the rise in false negatives tells us that vendor names are an important feature to include in training and testing since without them, our models do not pick up on as many true pairs as before. However, this observation also poses a question regarding the credibility of our *true alias pair* definition: are we mislabeling true vendor pairs by matching vendors with equivalent vendor names when, in fact, we shouldn't? Perhaps, our mislabeled data is training our models insufficiently. Lastly, the increase of predicted pairs from the real world simulation tells us the second set of models was still able to predict vendor pairs despite missing vendor name information.

With these observations in mind, we conclude that our proposed enhanced alias attribution method yielded promising results. However, we emphasize the importance of training future models with more accurately labeled vendor data. This is our immediate next research step, as well as evaluating our methodology for time complexity, memory usage, and robustness against profiling vendors with fewer than 10 listings. Additionally, we plan to evaluate this methodology with version 4 data to determine our methods robustness against missing data.

CHAPTER 8
FUTURE WORK

Several limitations of DVP have been mentioned throughout this thesis. In this chapter, we summarize these limitations and suggest future research directions. To begin, we reiterate that ground truth availability remains a challenge in dark marketplace applications. In our case, the elementary style alias attribution did not establish a ground truth; therefore, we were not able to evaluate the performance of using the simplistic approach for our task. In our enhanced model, we chose to rely on a synthetic ground truth which was not the most accurate representation of the dark vendor universe and caused our models to over-rely or *overfit* on the Vendor Name feature. Therefore, along with further investigating the cause of our model's over-reliance on vendor names, improved ground truth establishment is imperative future work and our immediate next research direction.

Evidently, it is very challenging to formulate a ground truth in dark web studies since (a) the size and contents of the dark marketplace universe remains unknown, and (b) it is impossible to verify *actual* aliases in the wild due to the anonymous nature of the dark web. Therefore, some suggestions for future ground truth establishment include manually examining datasets to better label aliases based on qualitative analysis. Also, it has been suggested that the training of the DVP models be conducted using surface web level data, such as data from Etsy and other less structured marketplaces. This way, the researcher could label surface web level aliases in the wild, train a model using this data, and then apply the trained models to unseen dark web level data and achieve profiling of dark vendors.

Also, our enhanced methodology is presented without time complexity and memory usage analysis. Since these variables have a large impact on the practicality of applying our methods in real world law enforcement investigations, it is integral

that future DVP work evaluates our alias attribution methodology with respect to time and memory usage. Further, our experiments consider vendors who have at least 10 product listings and one vendor page in DVP DB. Perhaps, future versions could consider any number of product listings and vendor pages or even the lack thereof.

Next, we clarify that our DVP framework includes an automated process for obtaining dark web scrapes. However, implementing an automated dark web crawler was not in the scope of this thesis since it is so challenging to develop comprehensive datasets from dark markets. Additionally, despite the DNM Archives having been used for many recognized academic studies, it may be outdated for current marketplace standards. Therefore, future work may include implementing a dark web crawler and scraper so the DVP system may be tested with more current data. Further, it would be advantageous if a new crawler was able to improve on some of the shortcomings of the DNM Archives. Thus far, we have considered the task of developing a new dark marketplace crawler with the intention of further supporting dark web investigations by combining the DVP classifier and a newly developed dark web crawler. Figure 8.1 illustrates the proposed system architecture of a fully automated DVP implementation. Overall, the more automated the workflow can be, the more aid DVP can offer to law enforcement and intelligence agencies.

Further, we note several limitations to the data used for DVP experimentation. For example, while constructing DVP DB V4 and V5, we pruned several dark marketplace artifacts since we did not include their data in the feature extraction process. Namely, we observed that data on product price, vendor registration date, trade count, last seen date, and vendor rating was regularly available on dark marketplaces, but we did not include this data for marketplace analysis. Likewise, while we considered the exif tags present in listing photos, our current implementation does not consider the values associated with the exif tags in order to avoid overly

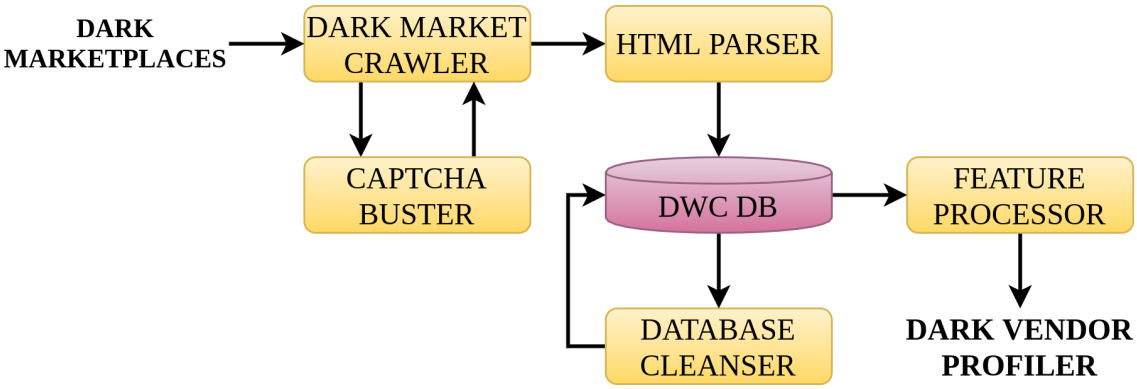


Figure 8.1: System architecture of a fully automated DVP implementation including a newly developed Dark Web Crawler (DWC).

complex feature dimensions. Also, while we considered several popular stylometry-based features, we excluded word-level and character-level N-grams for the sake of model simplicity. However, by excluding regularly available data, we may be missing valuable marketplace artifacts. Perhaps, if the exif tag values were considered for marketplace analysis, image-based artifacts would provide more value to the DVP system than observed in Section 7.1. Thus, future DVP work may revisit these artifacts and attempt to construct meaningful features from their values such that DVP performance may improve without exponentially increasing feature dimensionality.

Another limitation we observe from our data is that early versions of the DVP DB contained thousands of vendors who had vendor page entries but no listing entries. In this thesis, we only considered vendors who had listing data; thus, we excluded a significant amount of vendor page data derived from the DNM Archives. While it is still unclear whether (a) the DNM Archives accidentally missed listing information for several thousand vendors or (b) the vendors simply did not have any listings, we suggest future DVP work revisits the excluded vendor pages. Perhaps, we may eliminate the requirement of having listing data for future versions of DVP and

instead rely on vendor page data alone to achieve attribution. Further, we only used a fraction of dark marketplace data available in the DNM Archives. Since the DNM Archives were not used in their entirety for this research, future work may consider the excluded marketplaces in addition to the 25 marketplaces included in our study.

Like most works, we chose to omit any listing data related to human and sex trafficking to avoid ethical and legal dilemmas. While we argue that our methodology is more fit to be applied to these types of cases than previous work, we did not deem this kind of data necessary to evaluate DVP's performance and therefore did not include it during our evaluation. However, an evaluation of DVP for sex-oriented marketplaces is an important research direction before DVP can be applied in practice. Therefore, we suggest this future work to be completed by appropriate parties who have the proper jurisdiction to analyze this niche of dark marketplaces.

Finally, we highlight a few limitations of the current implementation of DVP and the evaluation of DVP presented in Chapter 7. First, we determined that our current feature extraction implementation is not robust in terms of time to increasing model complexity. Thus, in future work, we suggest improving feature extraction by implementing code optimization techniques and parallelization. We also determined that the Random Forest algorithm is not robust in terms of memory usage to increasing model complexity. Thus, future work may consider other machine learning algorithms that are more robust to increasing model complexity and can achieve similar accuracy. For example, although *Neural Networks* have been studied relatively thoroughly for the task of dark market analysis, revisiting this technique with new data and new perspectives could further improve DVP performance. Indeed, since the process of feature engineering is automated in neural network algorithms, a neural network implementation of DVP could alleviate the short comings of manually defining distinctive vendor features and result in even stronger feature sets for the

DVP tasks. Therefore, investigating various machine learning techniques could be an interesting future research direction.

Second, we note that our vendor attribution evaluation considered cases where listing data was evenly distributed among vendors resulting in *balanced* datasets. To further evaluate the performance of DVP, we may consider the problem of *class imbalance* where our classifier is trained to distinguish vendors with varying amounts of listings. This is an especially important research direction, for it may further determine how our proposal would fare in real-world dark web environments where vendor participation, marketplace popularity and listing availability are extremely variable.

CHAPTER 9
CONCLUSION

Due to the hidden nature of dark web technologies, cybercriminals are able to hide behind anonymous online identities and conduct illegal business. Consequently, it has become a challenge for investigators to identify these anonymous cybercriminals, shut down their businesses, and prosecute them. To facilitate deanonymization, investigators must collect as much data on these identities as possible; however, this is a non-trivial task considering the size of the dark web universe.

In this research, we proposed *Dark Vendor Profiling*: a novel way to automate the data collection and profiling of dark vendors which can support investigative efforts to deanonymize cybercriminal identities. Using a novel feature set derived from product listing pages and vendor profile pages collected from the DNM Archives, we demonstrated how vendor and alias attribution can be achieved using common data mining and machine learning techniques, thereby automating much of the *manual* labor typically required to carry out dark web investigations.

Dark marketplace analysis techniques have been researched and reported in several works. While these studies have achieved success for certain problem sets within this domain, they are generally limited by their scalability, model complexity, comprehensiveness and performance. Therefore, in this work, we improved upon the short-comings of related researches. In summary, we presented a new method for collecting image-based data, two unique feature sets to be used for vendor and alias attribution, a novel Random Forest based technique for the task of vendor attribution, a novel application of record linkage for vendor profile formulation and alias attribution, and an in-depth overall evaluation of our proposed Dark Vendor Profiling scheme.

Most importantly, we demonstrated how our vendor attribution model can distinguish between 1,500 unique vendors across several dark marketplaces with 92% accuracy provided that their product listings do not contain any missing values. Likewise, we demonstrated how this model may distinguish between over 3,200 unique vendors across several dark marketplaces with over 88% accuracy, even in the presence of missing values. These are significant achievements as no other study has attempted vendor attribution with such accuracy using such a large variety of dark marketplaces and such an extensive set of unique vendors.

Furthermore, we demonstrated how our vendor attribution techniques may be translated to achieve alias attribution using a record linkage based methodology. With our enhanced alias attribution model, we demonstrated how our method outperforms other state-of-the-art techniques and showed how the models trained with our labeled data may be applied to unlabeled data to identify vendor alias pairs in the wild. Finally, along with presenting our promising results, we have suggested several important future research directions for Dark Vendor Profiling in hopes that we may further support investigative efforts to deanonymize cybercriminal identities and reduce the impact illegal online markets have on modern day society.

REFERENCES

[1] Tor2web: Browse the tor onion services. `https://www.tor2web.org/`. (Accessed on 01/26/2020).

[2] Welcome to tor metrics. `https://metrics.torproject.org/`. (Accessed on 01/26/2020).

[3] Abbasi, A. & Chen, H.-c. (2008). Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems*, 26, 1–29.

[4] Adamsson, H. (2017). *CLASSIFICATION OF ILLEGAL ADVERTISEMENT WORKING WITH IMBALANCED CLASS DISTRIBUTIONS USING MACHINE LEARNING*. Technical report, Uppsala University.

[5] Afroz, S., Caliskan-Islam, A., Stolerman, A., Greenstadt, R., & McCoy, D. (2014). Doppelgänger finder: Taking stylometry to the underground. *Proceedings - IEEE Symposium on Security and Privacy*, (pp. 212–226).

[6] Baravalle, A., Lopez, M. S., & Lee, S. W. (2016). Mining the dark web: Drugs and fake ids. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)* (pp. 350–356).

[7] Bard, G. V. (2007). Spelling-error tolerant, order-independent pass-phrases via the damerau-levenshtein string-edit distance metric. In *Proceedings of the Fifth Australasian Symposium on ACSW Frontiers - Volume 68*, ACSW '07 (pp. 117–124). AUS: Australian Computer Society, Inc.

[8] Branwen, G., Christin, N., Décary-Hétu, D., Andersen, R. M., StExo, Presidente, E., Anonymous, Lau, D., Sohhlz, D. K., Cakic, V., Buskirk, V., Whom, McKenna, M., & Goode, S. (2015). Dark net market archives, 2011-2015. `https://www.gwern.net/DNM-archives`.

[9] Christin, N. (2013). Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13 (pp. 213–224). New York, NY, USA: ACM.

[10] Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Mach. Learn.*, 20(3), 273–297.

[11] Cramer, J. (2002). The origins of logistic regression. *Tinbergen Institute, Tinbergen Institute Discussion Papers*.

[12] Demant, J., Munksgaard, R., & Houborg, E. (2018). Personal use, social supply or redistribution? cryptomarket demand on silk road 2 and agora. *Trends in Organized Crime*, 21, 42–61.

[13] Deng, H. (2018). Why random forests outperform decision trees - towards data science. https://towardsdatascience.com/why-random-forests-outperform-decision-trees-1b0f175a0b5. (Accessed on 03/28/2020).

[14] Dingledine, R., Mathewson, N., & Syverson, P. (2004). Tor: The second-generation onion router. In *IN PROCEEDINGS OF THE 13 TH USENIX SECURITY SYMPOSIUM*.

[15] Du, P., Ebrahimi, M., Zhang, N., Chen, H., Brown, R. A., & Samtani, S. (2019). Identifying high-impact opioid products and key sellers in dark net marketplaces: An interpretable text analytics approach. In *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)* (pp. 110–115).

[16] Ekambaranathan, A. (2018). Using stylometry to track cybercriminals in darknet forums. Master's thesis, University of Twente.

[17] for Missing & Exploited Children, I. C. (2017a). Despite increase in global child protection laws many countries still do not consider child pornography a crime. https://www.icmec.org/press/despite-increase-in-global-child-protection-laws-many-countries-sti/ll-do-not/-consider-child-pornography-a-crime/. (Accessed on 02/18/2020).

[18] for Missing & Exploited Children, I. C. (2017b). New study reveals child pornography not a crime in most countries. https://www.icmec.org/press/new-study-reveals-child-pornography-not-a-crime-in-most-countries/. (Accessed on 02/18/2020).

[19] Garrahan, J. (2018). Authorship detection in dark web marketplaces using lstm and rnn neural networks.

[20] Ghosh, S., Porras, P. A., Yegneswaran, V., Nitz, K., & Das, A. (2017). Atol: A framework for automated analysis and categorization of the darkweb ecosystem. In *AAAI Workshops*.

[21] Goodin, D. (2014). Tor structure.

[22] Greenberg, A. (2013). End of the silk road: Fbi says it's busted the web's biggest anonymous drug black market. https://www.forbes.com/sites/andygreenberg/2013/10/02/. (Accessed on 01/27/2020).

[23] Gulati, S., Sharma, S., & Agarwal, G. (2018). *The Hidden Truth Anonymity in Cyberspace: Deep Web*, (pp. 719–730). Intelligent Computing and Information and Communication.

[24] Ho, T. & Ng, W. K. (2016). Application of stylometry to darkweb forum user identification. In *Information and Communications Security: 18th International Conference, ICICS 2016*, volume 9977 (pp. 173–183).

[25] I2P (2020). I2p anonymous network. `https://geti2p.net/en/`. (Accessed on 01/26/2020).

[26] Imatest (2019). Ssim: Structural similarity index. `http://www.imatest.com/docs/ssim/`. (Accessed on 12/11/2019).

[27] Janze, C. (2017). Are cryptocurrencies criminals best friends? examining the co-evolution of bitcoin and darknet markets. In *23rd Americas Conference on Information Systems (AMCIS 2017)*.

[28] Jeziorowski, S., Ismail, M., & Siraj, A. (2020). Towards image-based dark vendor profiling: An analysis of image metadata and image hashing in dark web marketplaces. In *Proceedings of the Sixth International Workshop on Security and Privacy Analytics*, IWSPA '20 (pp. 15–22). New York, NY, USA: Association for Computing Machinery.

[29] JonDonym (2020). Jondonym - the anonymisation service. `https://anonymous-proxy-servers.net/`. (Accessed on 01/26/2020).

[30] Lawrence, H., Hughes, A., Tonic, R., & Zou, C. (2017). D-miner: A framework for mining, searching, visualizing, and alerting on darknet events. In *2017 IEEE Conference on Communications and Network Security (CNS)* (pp. 1–9).

[31] Learn, S. (2007-2019). sklearn.ensemble.randomforestclassifier — documentation. `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html`. (Accessed on 03/28/2020).

[32] Marques, J. (2018). *Tor: Hidden Service Intelligence Extraction*. Technical report, Universiteit Van Amsterdam.

[33] Martin, J. & Christin, N. (2016). Ethics in cryptomarket research. *International Journal of Drug Policy*, 35.

[34] McDonald, A. W. E., Afroz, S., Caliskan, A., Stolerman, A., & Greenstadt, R. (2012). Use fewer instances of the letter "i": Toward writing style anonymization. In S. Fischer-Hübner & M. Wright (Eds.), *Privacy Enhancing Technologies* (pp. 299–318). Berlin, Heidelberg: Springer Berlin Heidelberg.

[35] Me, G., Pesticcio, L., & Spagnoletti, P. (2017). Discovering hidden relations between tor marketplaces users. In *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)* (pp. 494–501).

[36] Mittal, S., Joshi, A., & Finin, T. (2019). *Cyber-All-Intel: An AI for Security related Threat Intelligence*. Technical report, University of Maryland, Baltimore County.

[37] Moeller, K., Munksgaard, R., & Demant, J. (2017). Flow my fe the vendor said: Exploring violent and fraudulent resource exchanges on cryptomarkets for illicit drugs. *American Behavioral Scientist*, 61(11), 1427–1450.

[38] Narayanan, A., Paskov, H., Gong, N. Z., Bethencourt, J., Stefanov, E., Shin, E. C. R., & Song, D. (2012). On the feasibility of internet-scale author identification. In *2012 IEEE Symposium on Security and Privacy* (pp. 300–314).

[39] Pastrana, S., Thomas, D. R., Hutchings, A., & Clayton, R. (2018). Crimebb: Enabling cybercrime research on underground forums at scale. In *WWW*.

[40] Phelan, J. (2014). This is how these 12 countries will punish you for insulting their heads of state. `https://www.pri.org/stories/2014-03-12/how-these-12-countries-will-punish-you-insulting-their-heads-state`. (Accessed on 02/18/2020).

[41] Prabhakaran, S. (2018). Cosine similarity - understanding the math and how it works? (with python). `https://www.machinelearningplus.com/nlp/cosine-similarity/`. (Accessed on 03/28/2020).

[42] Python (2017). Imagehash · pypi. `https://pypi.org/project/ImageHash/`. (Accessed on 03/28/2020).

[43] Python (2019). Python record linkage toolkit 0.14 documentation. `https://recordlinkage.readthedocs.io/en/latest/about.html`. (Accessed on 04/03/2020).

[44] Qian, T. & Liu, B. (2013). Identifying multiple userids of the same author. In *EMNLP*.

[45] Rafiuddin, M. F. B., Minhas, H., & Dhubb, P. S. (2017). A dark web story in-depth research and study conducted on the dark web based on forensic computing and security in malaysia. In *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)* (pp. 3049–3055).

[46] Schäfer, M., Fuchs, M., Strohmeier, M., Engel, M., Liechti, M., & Lenders, V. (2019). Blackwidow: Monitoring the dark web for cyber security information. In *2019 11th International Conference on Cyber Conflict (CyCon)*, volume 900 (pp. 1–21).

[47] Scott, W. (2019). Tf-idf from scratch in python on real world dataset. `https://towardsdatascience.com/tf-idf-for-document-ranking-from-scratch-/in-python-on-real-world-dataset-796d339a4089`. (Accessed on 03/28/2020).

[48] Shavers, B. & Bair, J. (2016). *Hiding Behind the Keyboard*, chapter 2. Syngress.

[49] Soghoian, C. (2011). Enforced community standards for research on users of the tor anonymity network. In *Proceedings of the 2011 International Conference on Financial Cryptography and Data Security*, FC'11 (pp. 146–153). Berlin, Heidelberg: Springer-Verlag.

[50] Sonar, M. (2018). Threat intelligence & investigation platform | media sonar technologies. `https://mediasonar.com/`. (Accessed on 02/18/2020).

[51] Spitters, M., Klaver, F., Koot, G., & v. Staalduinen, M. (2015). Authorship analysis on dark marketplace forums. In *2015 European Intelligence and Security Informatics Conference* (pp. 1–8).

[52] Time-Magazine (2019). Inside the battle to get news to venezuelans | time. `https://time.com/5571504/venezuela-internet-press-freedom/`. (Accessed on 03/23/2020).

[53] Tor-Project. Anonymity online. `https://www.torproject.org/`. (Accessed on 01/26/2020).

[54] Tor-Project. Download. `https://www.torproject.org/download/`. (Accessed on 01/26/2020).

[55] Tor-Project (2017). Research safety board. `https://research.torproject.org/safetyboard/`. (Accessed on 02/01/2020).

[56] Tor-Project (2019). Tor: Onion service protocol. `https://2019.www.torproject.org/docs/onion-services.html.en`. (Accessed on 03/23/2020).

[57] Wang, X. (2018). Photo-based Vendor Re-identification on Darknet Marketplaces using Deep Neural Networks. Master's thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

[58] Wang, X., Peng, P., Wang, C., & Wang, G. (2018). You are your photographs: Detecting multiple identities of vendors in the darknet marketplaces. In *ASIACCS '18*.

[59] Wikipedia (2020). Legality of cannabis. `https://en.wikipedia.org/wiki/Legality_of_cannabis`. (Accessed on 02/18/2020).

[60] Zhang, Y., Fan, Y., Song, W., Hou, S., Ye, Y., Li, X., Zhao, L., Shi, C., Wang, J., & Xiong, Q. (2019). Your style your identity: Leveraging writing and photography styles for drug trafficker identification in darknet markets over attributed heterogeneous information network. In *The World Wide Web Conference*, WWW '19 (pp. 3448–3454). New York, NY, USA: Association for Computing Machinery.

VITA

Susan Jeziorowski was born in Chicago, Illinois on May 1, 1997. She graduated third in her high school class in 2015 and accepted a full athletic scholarship to play volleyball for Tennessee Technological University starting the following August. Upon graduating with a Bachelor of Science degree in Computer Science, she was awarded Tech athletics' highest honor, Woman of the Year, and the university's W.A. Howard award for completing her undergraduate studies with a 4.0 GPA. She began her graduate studies in January 2019 and received her Master of Science degree in Computer Science from Tennessee Technological University in May 2020.