Doppelgängers on the Dark Web: A Large-scale Assessment on Phishing Hidden Web Services

Changhoon Yoon
S2W LAB Inc.

Kwanwoo Kim
KAIST

Yongdae Kim
KAIST

Seungwon Shin
KAIST, S2W LAB Inc.

Sooel Son*
KAIST

ABSTRACT

Anonymous network services on the World Wide Web have emerged as a new web architecture, called the *Dark Web*. The Dark Web has been notorious for harboring cybercriminals abusing anonymity. At the same time, the Dark Web has been a last resort for people who seek freedom of the press as well as avoid censorship. This anonymous nature allows website operators to conceal their identity and thereby leads users to have difficulties in determining the authenticity of websites. Phishers abuse this perplexing authenticity to lure victims; however, only a little is known about the prevalence of phishing attacks on the Dark Web.

We conducted an in-depth measurement study to demystify the prevalent phishing websites on the Dark Web. We analyzed the text content of 28,928 HTTP Tor hidden services hosting 21 million dark webpages and confirmed 901 phishing domains. We also discovered a trend on the Dark Web in which service providers perceive dark web domains as their service brands. This trend exacerbates the risk of phishing for their service users who remember only a partial Tor hidden service address.

Our work facilitates a better understanding of the phishing risks on the Dark Web and encourages further research on establishing an authentic and reliable service on the Dark Web.

CCS CONCEPTS

• **Security and privacy** → **Phishing**; • **Information systems** → *Data extraction and integration*; • **Networks** → *Network privacy and anonymity*.

ACM Reference Format:
Changhoon Yoon, Kwanwoo Kim, Yongdae Kim, Seungwon Shin, and Sooel Son. 2019. Doppelgängers on the Dark Web: A Large-scale Assessment on Phishing Hidden Web Services . In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3308558.3313551

1 INTRODUCTION

The *Web* is the most popular, worldwide, and accessible platform for sharing and disseminating information across the globe. However, there is not only a bright side of the Web, but there is also a dark side. A set of platforms that host websites whose owners and users remain anonymous is now referred to as the **Dark Web**, whereas the *Surface Web* hosts regular websites. The Dark Web hosts websites whose formats and appearances are the same as those of the Surface Web. However, the way to access the Dark Web is different from that of the Surface Web. It demands the use of an anonymity network service for Dark Web service providers and their visitors to hide their identities on the Web.

The definition of the Dark Web has not officially been established [60], but it is often referred by the popular press and security community to emphasize illicit activities that abuse anonymity networks [15]. In this paper, we use the term "Dark Web" to refer to the collection of hidden Web services built on anonymous networks.

The Dark Web has become a major distribution channel for delivering and advertising malicious content. Silkroad [29] and Hansa-Market [13] are well-known Dark Web marketplaces that sell drugs, illegal weapons, and even malware. In addition, researchers have revealed that the Dark Web contained a considerable amount of harmful content [27, 53], and their findings have been confirmed by government investigative agencies [6] as well.

Conversely, the Dark Web offers a last resort for people who want to avoid censorship, to abide freedom of the press, and even to minimize tracking risks for their privacy. For instance, Venezuela experiencing the recent financial turmoil have blocked accessing political and social content on the Web, thus leaving Tor as the only option to access the restricted content [22].

Motivation. Phishing is one of the most effective threats that harvest users' privacy-sensitive information [64]. There thus exist the previous investigative studies that emphasize the severity of phishing attacks and the prevalence of phishing campaigns on the Surface Web [44, 54, 59, 61, 64]. Thomas et al. assessed the severity of Web phishing campaigns. Their study showed that phishing websites on the Surface Web emulating Gmail, Yahoo, and Hotmail logins had managed to steal 1.4 million credentials [61].

Conversely, the phishing threats on the Dark Web are understudied. Relatively little is known about the prevalence of phishing websites across the Dark Web universe. This trend stems from the absence of the oracle telling whether a given dark website is a phishing site or not. There is sufficient information to obtain the identity of owners operating websites on the Surface Web, which include HTTPS certificates, WHOIS, and DNS records. On the other hand, most dark websites' owners seek to hide their identities, which naturally makes it improbable for users to distinguish an authentic dark website from its phishing websites on the Dark Web.

We argue that it is crucial to address these phishing risks. On the Dark Web, users have no practical way to check the authenticity of

*Corresponding author

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313551

Web services except leveraging out-of-band clues, thus exacerbating the severity of phishing threats. To the best of our knowledge, mo previous study has investigated the phishing risks on the Dark Web. Prior work focused on analyzing illegal contents or unexpected activities [26, 42, 43, 46, 51], uncovering illegal activities [24, 25, 29, 34, 58, 63], and analyzing the popularity of the content on the Dark Web [27, 28].

Contributions. We conducted an in-depth analysis to identify phishing websites on the Dark Web. To do this, we collected more than 21 million webpages from 100K Tor hidden services for seven months. Our dataset represents the most up-to-date and comprehensive characteristics of the Dark Web.

We start with identifying phishing candidates whose website contents are almost identical to other websites. In other words, we investigate how many domains with distinct content and their duplicates constituted the Dark Web. We employ a carefully designed content grouping algorithm that classifies onion domains into content-wise distinct website groups based on the text and the title in each Tor onion domain's homepage. We observe that only 5,718 website groups exhibiting distinct content are available on the Tor network. Interestingly, the content of the top two website groups is duplicated in over 200 domains, which calls into question the authenticity of these domains.

For each website group consisting of multiple domains, we analyze all of the identified duplicates and confirm the presence of abundant phishing websites among the duplicates. Specifically, we identify 791 phishing websites that target five major Dark Web services including dark marketplaces and Bitcoin mixing services. We further analyzed how often users would encounter such phishing domains by counting cross references from other domains. In general, an authentic dark website is more frequently referenced by other dark domains than its phishing websites. However, interestingly, the most referenced website of Dream Market, a popular black market, is a phishing domain, which demonstrates that the attacker diligently spreads phishing domains.

To find further phishing websites on the Dark Web, we leverage "gray website." *Gray website* refers to a website that provides their identical services on both of the Surface and Dark Web. By leveraging the same ownership of a gray website and its corresponding surface website, we find 297 phishing websites on the Dark Web that target gray websites including Facebook. Our study is the first large-scale investigative study that confirms prevalent phishing sites on the Dark Web. Biryukov et al. [27] briefly mentioned the presence of one phishing website that mirrored Silkroad.

We also analyze the common trend in onion domain addresses on the Dark Web. Each website on the Tor network is represented by an onion domain name, a cryptographic string computed from the owner's public key. Considering that even picking an arbitrary five character-long prefix of an onion domain is computationally expensive and the visitors should still type the entire domain, we expect no meaningful prefix on onion names. However, we observe that the majority of the Dark Web service providers intentionally generate them to have meaningful prefixes that are at least five characters long. We conclude that the onion domain name itself is perceived as a brand for the service providers on the Dark Web. This trend has led to dark websites being vulnerable to phishing attacks because victims will depend on a partial memorable prefix instead of its entire onion domains to determine the authenticity of a service.

In this paper, we systematically studied unique/duplicate textual contents based on the comprehensive Tor anonymity network dataset, which is larger than any previous research. We observe abundant duplicate websites that target phishing victims and manifest their characteristics. We also confirm close correlations between website content and onion domains, which exacerbate the phishing risks on the Dark Web. Our work facilitates a better understanding of the phishing websites on the Dark Web and invites further research.

2 RELATED WORK

The *Dark Web* is a privacy-centric web environment built on anonymity networks such as Tor [33]. In the name of anonymity, the Dark Web has been aiding and abetting illegal activities. According to Biryukov et al. [27], most dark websites hosted content devoted to adult films (17%), drugs (15%), counterfeit items (8%), and weapons (4%) in 6,579 HTTP(S) Web services. Moreover, another study [53] has shown that the most popular Tor hidden services were botnet command and control (C&C) servers, and a large proportion of the content on the Dark Web is of questionable legality.

The notoriety of the Dark Web has triggered national authorities to seize many websites that are directly involved in illegal activities or that indirectly help users commit crimes [6]. Therefore, there is a growing interest among the authorities as well as researchers to understand the Dark Web, which demands discovering, classifying, and even deanonymizing dark websites.

Unlike the Surface Web, where domain names are well indexed by search engines, the Dark Web domain names are unknown unless they are explicitly announced via public channels. Biryukov et al., however, have demonstrated that it was possible to reveal all the Tor hidden services domain names by exploiting a design flaw of Tor's hidden services protocol [28]. Specifically, they modified the Tor relays (Onion routers) to sniff out the domain names. Today, the Tor administrators are actively detecting and banning such misbehaving onion routers [38, 56]. In addition, a recent study [45] showed that it is also possible to gather some of the onion domain names using conventional search engines that crawl and index hidden services via Tor proxy services [37].

Deanonymization of Tor hidden services has also been studied in previous work. Øverlier and Syverson [51] were the first to demonstrate the deanonymization of Tor hidden services, and there have been other studies [26, 42, 43, 46, 50] until today. More recently, there have been attempts to automate the analysis of the Dark Web; CARONTE [46] is an automatic Tor hidden service deanonymization system, and ATOL [35] is an automated Tor hidden service categorization framework.

Christin [29] and Soska et al. [58] have performed an extensive analysis of anonymous marketplaces on the Dark Web. Several related studies [24, 34] have also assessed the volumes of these marketplaces. On the customer side of dark markets, Van Hout et al. [63] and Barratt et al. [25] have provided notable insights into the participants of drug marketplaces. Biryukov et al. [27, 28] and Owen et al. [53] have analyzed the content and popularity of the Tor hidden services in general. Recently, Sanchez-Rola et al. [57]

measured the structural connection between the Dark Web and the Surface Web, and they also assessed the usage of tracking scripts on the Dark Web.

Phishing has remained a critical security threat [31, 32, 36, 40, 41, 48, 61, 64]. M. Cova et al. [31] investigated the diverse phishing methods of prevalent phishing kits. X. Han et al. [36] proposed a defense sandboxing technique to protect victims from phishing kits. Thomas et al. [61] used a Google dataset to assess the severity of currently on-going phishing campaigns and found that 12.4 million people are potential phishing victims, and 1.9 billion credentials have been leaked due to previous data breaches.

We conducted a content-based analysis to identify phishing websites on the Dark Web. Due the anonymous nature of the Dark Web, identifying phishing websites itself imposes a unique challenge. To the best of our knowledge, our study is the first investigative study that systematically examines the presence of phishing websites on the Dark Web and their distinctive characteristics.

3 DATA COLLECTION

The Dark Web refers to websites accessible via all the anonymity networks, including Tor [33], I2P [39], and Freenet [30]. Our study targets *hidden Web services*, which are the hidden services accessible via HTTP(S) on the most prevalent anonymity network, Tor.

We initially collected 10K Tor hidden service addresses (or onion domain names) from the popular Tor hidden service search engines and directories: Ahmia [23] and FreshOnion [10]. One of the most effective onion domain collection strategies, which Biryukov et al. [28] introduced, would capture all the onion domains available on the Tor network. They introduced a method of deploying modified onion relays that sniff out hidden service descriptors from anonymity network traffic. However, because the Tor policy now bans such misbehaving onion routers [9], we did not employ this method for ethical and respectable research.

We implemented a Web crawler based on Scrapy [18] and Splash [20] to collect not only static content but also dynamically changing content. To bypass anti-crawler/bot mechanisms that dark websites may employ, we used the latest Tor browser user agent for the crawler and distributed HTTP(S) requests towards the same domain via different Tor circuits to avoid being blacklisted due to sending numerous requests.

We deployed 16 web crawler instances and began by exhaustively exploring each of the 10K seed onion domains to collect all of the Tor HTTP(S) hidden services. We configured the crawlers not to follow any hyperlink to the Surface Web, but only to traverse .onion links until there were no more links to follow. We regularly updated our seed onion addresses by visiting the sources above to maximize our data coverage. Because both Ahmia and FreshOnion provide a full list of their indexed hidden services, the collected seeds were not biased to our selections of search keywords. We only collected textual data to avoid downloading any illegal content.

During the data collection period (7 months), we collected 28,928 unique onion domains with 21 million webpages. Table 1 summarizes our dataset coverage. Notably, of the 28,928 onion domains, there were 13,326 distinct second-level domains (2LDs). Our crawlers were initially deployed with 10K seed onion addresses, and by the end of our data collection period, we found 100K distinct Tor hidden services (or 2LDs). Although we were able to collect 100K distinct 2LDs, in the end of our data collection period, our webpage dataset included the webpages collected from 13,326 2LDs. This discrepancy resulted from the fact that not all Tor hidden services were alive during the data collection period (7 months) and a partial fraction of the Tor hidden services provided their Web services.

Table 1: Dark Web data collection

Collection period	January 2017 ~July 2017
Number of domains	28,928
Number of 2LDs	13,326
Number of webpages	21,537,119

Note that previous research [27] covered 100% of the Tor hidden services (or 39,824 2LDs) in 2013, and they reported that *only 3,741 of them provided web (HTTP) services*. We emphasize that the number of domains and webpages of our dataset demonstrates extensive coverage of the Dark Web.

We use *domain* and *2LD* to refer to an onion domain and a second-level onion domain, respectively. The *2LD* of an onion domain is an 80-bit number in base32 encoding and generated from a public key, which represents a unique client who owns the corresponding private key. For a domain not on the Tor network, we explicitly use the term:*surface domain*. The *homepage* of a given domain is the final destination page when visiting the domain with a browser.

4 IDENTIFYING CONTENT DUPLICATES

A phishing adversary lures victims to her website of which content is almost identical to the one of a legitimate website, thus expecting victims to provide sensitive information. Since it is infeasible to vet the authenticity of websites on the Dark Web by its anonymous nature, we exploit this intrinsic characteristic to identify phishing website candidates.

To accurately compute phishing candidates, we cluster a large number of domains into one website group when their homepages showed near-identical text contents. We then analyze how many onion domains share the equivalent contents and how many distinct contents exist on a per-website-group basis. Section 4.1 elaborates on our grouping algorithms, and Section 4.1.3 shows the result.

We further investigate multiple domains that share near-identical content in each website group. We leverage external clues nudging authentic service domains and investigate domains not matching the clues in each website group. Section 4.2 describes our findings and the characteristics of observed phishing websites on the Dark Web.

Several surface websites have been hosting their services on Tor hidden services [16] for privacy-concerned users who refuse to be tracked by the service providers. Such websites that appear on both sides are not truly *dark* because the service providers of these websites are no longer anonymous. We refer to such websites as *gray websites* and analyze their unique characteristics. We then identity gray website groups, each shares an almost identical content. In each group, we excluded the gray website that matches a surface webpage and then investigated the remaining domains to manifest their phishing risks. Section 4.3 describes our algorithm for finding gray websites and discusses our findings.

4.1 Website Classification

Non-text data (e.g., videos or images) published on the Dark Web may contain illegal or abusive content [27, 53]. We refrained ourselves from analyzing visual content on the Dark Web, but focused on storing and analyzing textual data on the Dark Web. For this reason, we only scraped and stored HTTP(S) responses with textual content.

We group the domains whose homepages share near-identical text contents into one group. Our classification algorithm has two phases. Phase I groups domains based on visible texts and titles collected from homepages. For each dynamic homepage, which Phase I has overlooked, Phase II extracts a set of the words that appears on the homepage and finds the website group, of which representative word set best covers the word set from the homepage. Each website group represents a set of homepages that have different domains but share almost identical contents.

4.1.1 Classification Phase I. Phase I groups domains based on *visible texts* and *titles* on their homepages. It starts by grouping the domains that show identical texts on their homepages. In the next step, it merges the previously computed groups when any homepages from two different groups share the same title. We continue to merge groups until there is no overlapping titles. Then, each merged group is called a *website group* that represents a set of onion domains that provide almost identical textual contents on their homepages.

Our approach is a heuristic with limitations. A website group combined based on excessively common title or text does not provide meaningful insights into the group. For example, the Apache Web server default page (e.g., It Works!) could be a homepage of many onion domains. We analyzed all the titles and texts of each group and then identified 46 different common titles and texts. We deliberately ignored such texts and titles because they contain no meaningful content.

We manually analyzed dark website groups and verified whether the homepages in the same group indeed have the same content. We confirmed that the Phase I algorithm produced no error such that no website has text content different to its group's one. However, this analysis was incomplete because there were 4,816 domains left unclassified. Phase I only groups websites sharing an identical title and text content; hence, it cannot correctly group website homepages with dynamic contents (e.g., homepages with a visitor counter or a clock).

4.1.2 Classification Phase II. Phase II completes the grouping of homepages with dynamic contents. From each website group C from Phase I, a group of onion domains, we retrieved a static word set w_C that appear in common. Afterward, for each unclassified homepage, we compute its word set w_h and then find a group C s.t. $w_C \subset w_h$.

The suggested algorithm has a limitation. Phase II becomes ineffective when the representative word set for each group is quantitatively small. Therefore, we selectively compared each ungrouped homepage h with each website group C from Phase I. If the static word set size $|w_C|$ for group C is relatively smaller than the homepage word set size $|w_h|$, we excluded this grouping candidate pair, $\{C, h\}$.

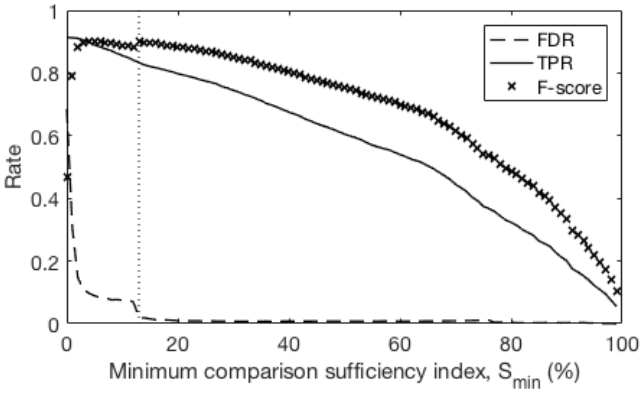

Figure 1: True positive rate (TPR), False discovery rate (FDR) and F-score of the Phase 2 classification by varying the minimum comparison sufficiency index S_{min}.

Quantitatively, for the pair $\{C, h\}$, if the comparison sufficiency index, $S_{C,h} = \frac{|w_C|}{|w_h|}$, is lower than the minimum comparison sufficiency index, S_{min}, we ignored grouping this pair. To determine the optimal S_{min}, we randomly sample 20% of the homepages from each group of onion domains from Phase I and reclassified them using different S_{min}. For each reclassification trial, we measured the true positive rate (TPR) [1] and false discovery rate (FDR) [2] by varying S_{min} as shown in Figure 1.

Our experiment shows that the F-score, the harmonic mean of TPR and 1−FDR, is highest when S_{min} was 13% (dotted line in Figure 1) and the FDR was only 1.9% based on our dataset (the TPR was 83%). We thus disregard any comparison pair whose $S_{C,h}$ is less than 13% in the final grouping phase. Phase II classified 3,397 domains into groups with only one domain member and the remaining 1,419 domains were assigned into the groups with multiple domains. For 200 sampled domains from those 1,419 domains, we manually checked their groups and confirmed that about 84% of websites had their correct grouping.

4.1.3 Classification Result. Table 2 summarizes our final grouping result. From the collected 28,928 onion domains, we found that there were 5,718 text-based distinct website groups. Of the 5,718 website groups, approximately 60% of them were single-domain website groups, and the rest (about 40%) of them were multi domain website groups. This means that 40% of the onion domains had at least one more domain whose homepage text contents were almost identical.

In the rest of this paper, a group of onion domains, which share almost identical texts on their homepages, is referred to as a *website group*. Moreover, we also counted the number of distinct 2LDs for each website group, and the result was not very different from the previous findings (approximately 40% of the website groups were associated with more than one 2LD).

[1] The proportion of websites truly in the group that are correctly classified, a.k.a. recall
[2] The proportion of websites classified to be in the group that are not originally in the group.

Table 2: Distinct website groups (WG) on the Dark Web

# of domains	1	2-50	51-100	101-200	201-300	Total
# of WG (%)	3,397 (59.31)	2,307 (40.35)	7 (.12)	5 (.09)	2 (.03)	5,718 (100)

# of 2LDs	1	2-50	51-100	101-200	201-300	Total
# of WG (%)	3,423 (59.86)	2,286 (39.98)	4 (.07)	4 (.07)	1 (.02)	5,718 (100)

4.2 Phishing Websites

A website group contains all websites that share almost identical texts. However, all websites in such a group are not necessarily the same website managed by one operator. If so, *why do onion domains in such a website group point to different websites although they share the identical content?* In this section, we analyze such domains and manifest 791 phishing onion domains that have targeted six popular dark websites.

There is no known effective method of identifying phishing websites on the Dark Web without an external clue. Dark website owners seek their anonymity and the nature of the Tor hidden network guarantees such anonymity at the protocol level. Therefore, it is improbable to decide whether two websites sharing the identical content belong to the same entity, thus hindering the inference of whether one of them is a phishing website. For the rest of this section, we use different kinds of *external clues* to identify 791 phishing domains. It means that the number of the identified phishing domains is a lower bound in 5,718 website groups, which may contain more phishing dark websites.

We identified the phishing domains that have targeted five (infamously) well-known dark websites (Table 3): Bitcoin Fog, Bitcoin Blender, Helix, AlphaBay Market and Dream Market. We specifically selected these sites, because they provide their authentic domains, which serve as clues for identifying phishing sites. However, there is a technical challenge. The five websites published their authentic domains on forums or their own websites on the Dark Web. It makes difficult for us to verify the authenticity of such information because anyone on the Dark Web can claim that his/her own website is an authentic domain for any of those five websites.

Table 3: Dark websites with known authentic domain names

Service type	Name	Authentic domains (#)	Phishing domains (#)
Bitcoin Mixers	Bitcoin Fog	1	276
	Bitcoin Blender	1	53
	Helix (Grams)	1	187
Dark Marketplaces	AlphaBay Market	16	165
	Dream Market	4	110
Total		24	791

4.2.1 Bitcoin mixers. Although all Bitcoin transactions are pseudonymous, they are public and traceable [49]. Hence, it is possible to link who has been involved in which transactions by analyzing the public blockchain [47, 55]. For this reason, Bitcoin mixing services that anonymize a user's Bitcoin transfers have gained significant popularity. Bitcoin Fog [3], Bitcoin Blender [1] and Helix [12] have been three of the most popular Bitcoin mixing services [62].

One challenge in this study is to identify the authentic onion domains of the three services, which are available exclusively on the Dark Web. Because the operators of these services conceal their identity, it is difficult for them to establish a reliable public channel to announce and promote their services. From our investigation, we learned that Bitcoin-related services are often promoted on a popular online forum called Bitcoin Forum [5], where the operators can anonymously promote their services and build up their reputations. Bitcoin Fog and Bitcoin Blender have first announced and released their authentic domain names on Bitcoin Forum, and both of these services are still currently active [2, 4]. In the case of Helix (or Grams[3]), they have done the same on Reddit [7]. Moreover, the authentic domain names of all three services have also been confirmed in a previous work [62].

By leveraging the authentic domain names of the Bitcoin mixing services and our website classification results, we were able to identify the phishing onion domains. The authentic domain of Bitcoin Fog was found in a website group with 277 different 2LDs, and we were able to conclude that the other 276 domains are phishing domains. To confirm this finding, we manually visited each domain in the Bitcoin Fog website group and investigated their content and behaviors.

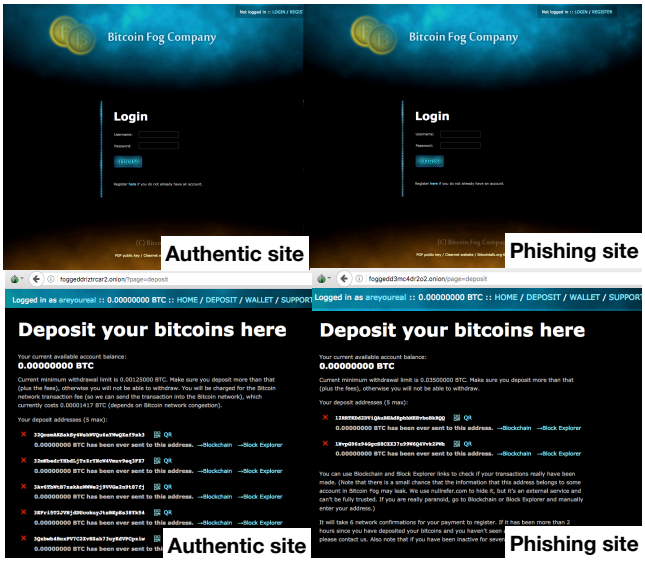

Figure 2: BitcoinFog: Authentic vs. phishing sites

The phishing websites of Bitcoin Fog were surprisingly sophisticated. As shown in Figure 2 (Top), the homepages of both the authentic and phishing sites were a login page and they were visually identical. The behaviors and functionalities of the websites

[3]Grams was a Dark Web marketplace search engine, and they additionally provided Bitcoin mixing services later on. Currently, Grams, including Helix, is discontinued. [12]

were also the same. Figure 2 (Bottom) shows that both the authentic and phishing sites allow users to add or remove Bitcoin deposit accounts. However, the authentic site and the phishing sites did not share the same backend user databases. As demonstrated in Figure 2 (Bottom), we were able to use the same username to create an account for both the authentic and phishing sites, but the Bitcoin deposit addresses linked to the accounts were different.

Bitcoin Blender's authentic domain belonged to a website group with 54 2LDs. Knowing that there is only one authentic 2LD, as shown in Table 3, we could infer that 53 2LDs in the group are phishing domains. We also manually verified each of the 53 phishing domains. Their websites sophisticatedly imitated the authentic website. The authentic website offered more functionalities, such as "Quick Mix" (Bitcoin mixing without an account) and password recovery; however, without knowing that the fact that the authentic website provides such features, it is highly likely that users will not be able to tell which one is the authentic site. Furthermore, the authentic site required users to solve CAPCHA, while the phishing sites did not.

In the case of Helix, its website group contained 188 2LDs, while only one of them was authentic. We also manually analyzed the websites linked to the domains, and we confirmed that the phishing domains led us to non-authentic websites.

4.2.2 Dark Marketplace. Another popular type of dark web services is a dark marketplace, which provides an e-commerce platform to anonymously trade illegal goods or services. Since Silk Road [29] has been terminated, there have appeared other dark marketplaces on the Dark Web. We analyzed two major dark marketplaces that have been active during our data collection period: AlphaBay and Dream Market. Compared to the other cases introduced above, it was even more difficult to identify the authentic domain names of AlphaBay and Dream Market because these sites have been using multiple number of domain names and even frequently changing them over time.

To discover the authentic domain names of AlphaBay, we took advantage of its takedown incident that happened in July 2017 [6]. According to FBI [6], "multiple computer servers used by the AlphaBay website were seized worldwide". Because the incident happened in the last month of our data collection period, we were able to capture the official seizure notice published on the seized websites. In our AlphaBay website group, there were 181 distinct 2LDs. 16 of the 2LDs have been shown the seizure notice since July 2018. Therefore, we inferred such 16 2LDs are authentic AlphaBay domain names, which makes the other 164 domains phishing.

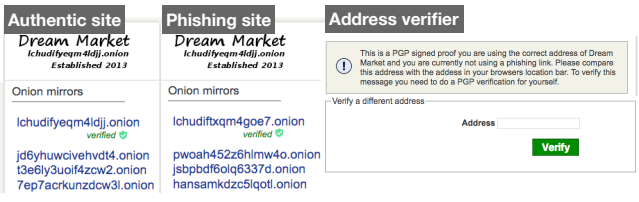

Figure 3: Dream Market: Authentic vs. phishing sites

Dream Market is one of the largest dark marketplace that are currently active, and this market also has multiple mirrors as shown in Figure 3 (Left). Although Dream Market has published a list of the verified domain names to cope with its phishing sites, these phishing sites have also been doing the same by announcing their official domain names (Figure 3-Middle). In addition to actively promoting its authentic domain names, Dream Market provides a domain name verifier (Figure 3-Right), which allows users to check if a given onion domain name is one of the authentic Dream Market domain names or not. By using this verifier, we tested 114 2LDs and found that only four of them are authentic, which suggests that the other 110 domains are phishing domains.

4.2.3 Domain name popularity vs. authenticity. We used different types of clues to identify phishing domains for the five popular dark web services. For these identified phishing domains, we measured the popularity of such domain names on both the Dark and Surface Web. For the popularity, we counted the number of distinct dark webpages that mention each domain name, while excluding any self-references. We used 2LD names that are 16 characters long as keywords to search not only the hyperlinks but also the text that mention them from our dark webpage data collection. To measure the 2LD name popularity on the Surface Web, we counted the number of the search results returned by Google custom search API [11] for each 2LD name.

Figure 4 (a)-(e) shows the popularity of the authentic and phishing 2LD names on the Dark Web for each the service. Each point indicates the number of distinct dark webpages that mention each 2LD name, and two types of markers are used to show the authenticity of the domain names (X markers for phishing domains, O markers for authentic domains). Figure 4 (f)-(j) illustrates the popularity of the 2LD names on the Surface Web, and each point indicates the number of the search results returned from Google when searching for each 2LD name.

In the case of the three Bitcoin mixing services, each of which has the only one authentic domain, the authentic domains have been more popular than the phishing domains on both the Dark and Surface Web except for Bitcoin Blender. One of Bitcoin Blender's phishing domain names has been mentioned more than its authentic domain name on the Dark Web (Figure 4-b). On both of the web environments, the authentic domain names of the dark marketplaces have been referenced more than the phishing domain names; however, on the Dark Web, the most popular domain name of Dream Market was a phishing domain name (Figure 4-e). Furthermore, the phishing domain names have been more frequently mentioned on the Dark Web than on the Surface Web. This difference demonstrates that phishing attackers diligently spread fake information on the Dark Web by abusing their anonymity. In most cases, no result was returned by Google when searching for the phishing domain names of the five major services.

4.3 Gray Websites

A gray website is a dark website that satisfies two conditions: (1) it has the corresponding surface website providing and identical service, and (2) both websites are managed by the same entity. A gray website is usually the mirroring version of a surface website whose customers want to remain anonymous on the Web. We assumed that there are abundant phishing websites on the Dark Web that target such gray website visitors.

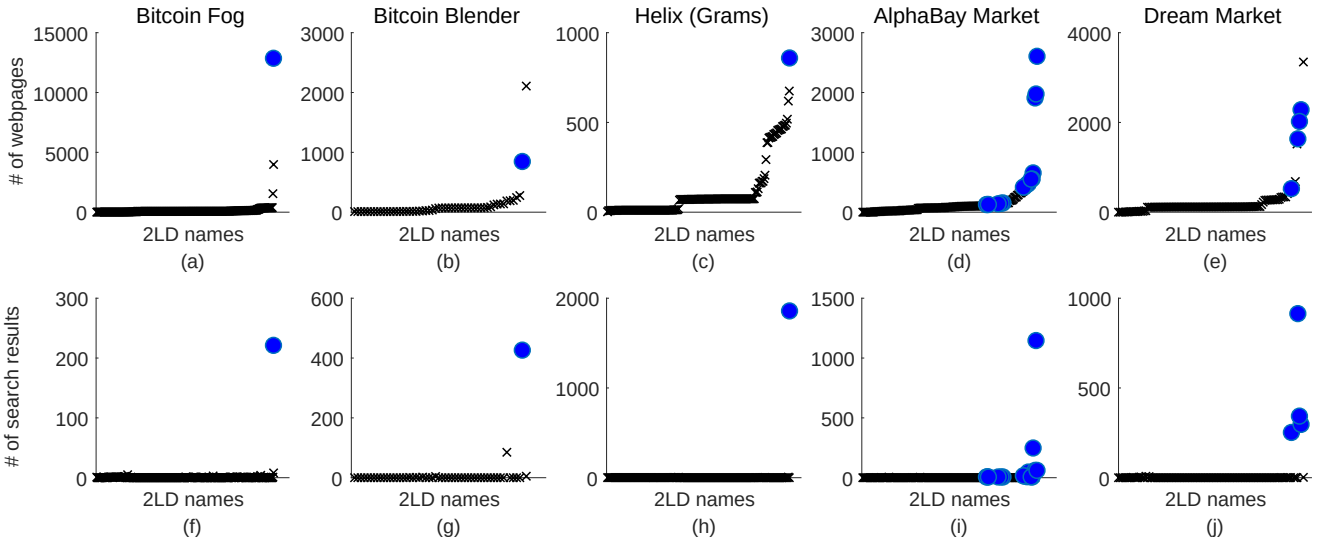

Figure 4: The popularity of five different Dark Web service domain names on the Dark (a-e) and Surface Web (f-j): The authentic domain names (in circular markers) have been mentioned more often than the phishing domain names (in x markers).

To identify such phishing websites, we first find authentic gray websites, each of which has a surface website with identical content. For each authentic gray website, we then compute its website group consisting of content duplicates. Finally, by leveraging information on each surface website paired with an authentic gray website, we confirmed phishing websites using content duplicates.

We started by identifying gray websites by partially applying the algorithm introduced by Matic et al. [46]. Our gray website identification algorithm extracts four types of clues from our dataset and then retrieves the same clues from surface websites. The extracted clues are dark website title, surface domain (e.g., non-onion domain names) in the title, Google Analytics tracking ID, and AdSense publisher ID.

We assumed that the four different clues are effective in deciding whether different websites are managed by the same entity. A homepage title is usually a representative string that shows the characteristics of the website. Because the homepages are already well-indexed by search engines, retrieving a title obtained from a dark homepage is an effective way of checking its existence on the Surface Web. Furthermore, if the title of a dark homepage includes a surface web domain, it is a strong indicator that the surface domain is the entry point of the same website. We also explored the methods of using both Google Analytics tracking and AdSense publisher IDs, which are assigned to each user. Usually, an owner of both dark and surface websites wants to maintain a unified view of his or her services. Therefore, such identifiers are compelling clues to check the same ownership of dark and surface websites.

We used a regular expression that matches surface domain names and matched 5,776 domains from the dark website titles. We then validated the matching surface domains by querying DNS servers, which confirmed 145 valid surface domains. We applied the same technique to extract the other clue types and obtained 276 Google Analytics IDs and 1,171 Google AdSense publisher IDs.

Table 4: Gray websites (GW) by each clue type

Clue type	# of search results	# of potential GW group	# of GW group
Domain in title	N/A	145	82
Homepage title	120,677	1,840	320
Analytics ID	1,241	1,049	80
AdSense ID	17,765	14,720	12

From the three clue types extracted from dark websites, we searched for surface websites using Microsoft's Bing Web Search API [17]. As Table 4 shows, we extracted surface domain candidates for each clue type. Each surface website candidate is then verified whether there exists a dark website whose body texts were similar to the contents of the surface website candidate. We used the same classification technique described in Section 4.1 to decide whether both dark and surface website candidates showed near-identical texts. As a result, we confirmed 383 gray website groups that were also available on the Surface Web.

In addition, we further investigated the contents of what had been identified as gray websites to understand their unique characteristics. We classified the website groups into three major categories: user-privacy-sensitive, provider-privacy-sensitive, and non-privacy-sensitive. A user-privacy-sensitive website is a website whose visitors would want anonymity. Such sites include pornography, file sharing, online gambling, forums, cryptocurrency services, etc. A provider-privacy-sensitive website indicates a website whose owners want to remain completely anonymous because of their illegal activities. Of the 383 gray website groups, about 35% of them were user-privacy-sensitive, about 10% of them were provider-privacy-sensitive, and the rest were non-privacy-sensitive as shown in Table 5.

Among 383 gray website groups, we further identified phishing websites. Note that each gray website group has one authentic

Table 5: Gray website categories

Privacy	Topic type	Percentage (%)	
User	adult	0.86	35.06
	political	2.01	
	forums	3.74	
	file sharing	5.17	
	cryptocurrency	6.9	
	gamble	7.18	
	communications	8.05	
	news	1.15	
Provider	illegal market	9.77	9.77
Irrelevant	academic	3.16	55.17
	software	6.9	
	security company	9.2	
	personal	17.53	
	others	18.39	

gray website with a corresponding surface webpage. We checked whether such a surface webpage explicitly mentions their authentic gray website onion domains. When a gray website group contains other domains that are not promoted by the surface website, we classify other domains as phishing websites. Table 6 summarizes the confirmed gray phishing websites. We confirmed 297 phishing domains from 383 gray website groups consisting of 734 domains. We classified 218 domains in 170 gray website groups as potential phishing websites because their surface websites did not specify their authentic onion domains.

Table 6: Phishing gray websites

Phishing	WG (#)	Authentic gray website domains (#)	Phishing domains (#)
Confirmed	213	219	297
Potential	170	N/A	N/A
Total	383	219	297

One of the phishing gray websites that we have identified targets Facebook. Since Facebook has announced the official onion domain name (facebookcorewwwi.onion) of their dark website [16], the other domains that belong to its website group were obviously phishing domains. As shown in Figure 5, the landing pages of the two domains are visually identical. One notable difference is that the Facebook's official dark website (Figure 5-Top) provides their HTTPS certificate, while the other site (Figure 5-Bottom) does not. We further analyzed the phishing site, and as shown in Figure 5 (Bottom), we confirmed that they were stealing the Facebook login credentials.

4.4 Discussion

By classifying the onion domains into content-wise identical website groups, we confirmed that about 40% of the website groups were associated with more than one onion domain. Then, *why would many website groups have more than one onion domain for their own service?*

Network balancing. One possible explanation is that many websites on the Dark Web could have been using multiple subdomains (lower than second-level) to provide identical content. However,

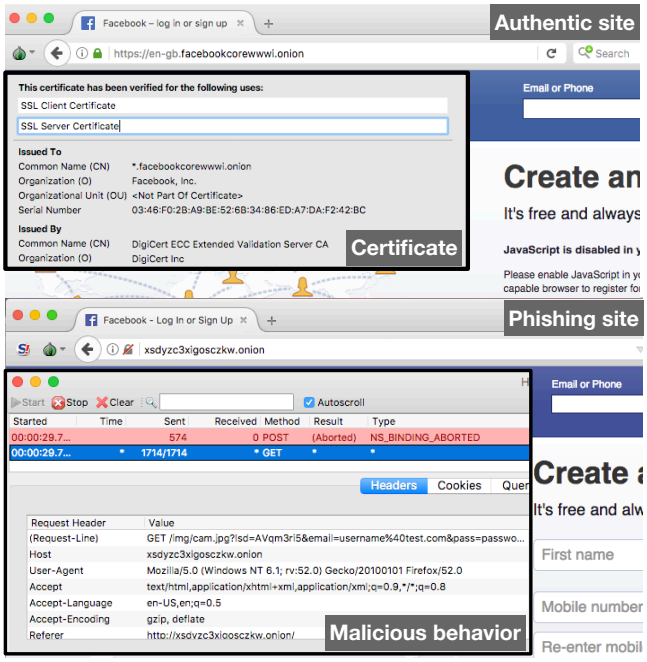

Figure 5: While the official Facebook site (Top) supports HTTPS with a valid certificate, the phishing site (Bottom) does not support HTTPS.

as shown in Table 2, the ratios of single-domain website groups to multi-domain website groups between domains and 2LDs remained almost the same. Hence, we excluded this explanation. Another explanation is that the dark websites could have deployed multiple numbers of Tor clients (or relays) to address the known scalability problems in the Tor hidden service architecture [14, 52].

One scalability problem arises when there are many concurrent requests to a 2LD. A 2LD is bounded only to a unique Tor client because it is computed from the public key assigned to the client. Consequently, one Tor client is solely responsible for the network traffic directed to and originated from its onion address. In such an architecture, the Tor clients that provide hidden services can easily become network bottlenecks as they lack load balancing options [14]. To avoid such a problem, the service providers operating a popular hidden service could have deployed multiple Tor clients, each of which has its own 2LD, to distribute the visitors.

The other scalability problem lies in the *introduction point*, as Tor announced [14], "The current hidden services architecture does not scale well due to introduction points hammered by many clients." An introduction point is a regular Tor relay, chosen by a Tor hidden service for setting up a *rendezvous point*. Thus, accessing such an introduction point is a preceding step to access the hidden service. When the service handles many concurrent requests, their service reliability depends on how many requests their introduction points can handle. This problem was addressed in a previous study [52]; however, Tor has not deployed the proposed solution.

Meanwhile, this problem is easily solvable by deploying multiple 2LDs (or Tor clients). Because multiple 2LDs have more introduction points than one 2LD, each introduction point becomes having less incoming requests of notifying rendezvous points. This simple trick

explains our observation that having multiple 2LDs is common across many dark websites to improve their service availability.

Prevalent phishing websites. We also confirm that phishing is another cause of abundant domains for each website group. We observed numerous phishing websites with content duplicates of authentic dark websites. In particular, we found that at least 791 onion domains have been hosting the phishing websites of the five well-known Dark Web services during our data collection period. After adding 297 gray phishing domains, we confirmed 901 unique phishing domains on the Dark Web.

We observed that many dark websites have been employing different strategies to cope with the phishing threat. One strategy is to leverage a popular social platform on the Surface Web. For instance, Bitcoin Fog and Bitcoin Blender have been building up their reputations on Bitcoin Forum which has more than 2M users. Bitcoin Forum serves as a reliable communication channel among their dark website visitors to establish the authenticity of their onion domains.

The other strategy is to provide an additional service telling whether a given domain is an authentic dark website service. Dream Market, for instance, has been actively announcing their official domain names, and they even provide an address verification service (Figure 3-Right) that tells whether a given domain name is one of their authentic domains or not. Interestingly, Dream Market offers a two-factor authentication (2FA) to protect their users from phishing attacks. Because their users also seek the anonymity as well, instead of requiring the email addresses of their users, they leverage the PGP public/private key pair of a user with 2FA. A user first registers her or his PGP public key. For the following authentications, the website requests the user to decrypt a cipher-text which is encrypted with her or his PGP public key. A phishing attacker is unable to access the Dream Market website even with the stolen credential of a victim unless the attacker has the victim's PGP private key.

We also measured the popularity of the authentic and phishing domain names of the six well-known Dark Web services. We counted the number of surface and dark websites that contain the authentic and phishing domains in their homepages. In most cases, the authentic domain names were relatively more popular than the phishing ones on both the Dark and Surface Web. It shows that the popularity of an onion domain name can be an indicator to determine its dark website authenticity.

Another interesting finding is that the phishing domains are rarely mentioned on the Surface Web while the authentic names are relatively popular in both of the environments. It demonstrates that the anonymous environment of the Dark Web is more susceptible to the fake information on the authentic domains than the Surface Web, which unfortunately has been abused by phishing attackers.

Gray websites. There have been a small number of popular surface websites that are also available as Tor hidden services. Such gray websites are mostly user-privacy-sensitive websites, and each aims to achieve a different goal by allowing users access to the site via the Tor network. For example, The New York Times has been available via nytimes3xbfgragh.onion [21], and as a news company, it aims to allow people around the world to avoid any Internet censorship and read its articles via the Tor network. Facebook has also launched facebookcorewwwi.onion [16] in a similar manner.

We observed that 213 surface websites released their authentic gray website domains. At the same time, we confirmed 297 phishing domains which have lured victims who wish to use gray websites. For instance, DeepDotWeb is a news site covering events and dark website reviews, accessible via www.deepdotweb.com and deepdot35wvmeyd5.onion. Our analysis discovered eight phishing domains cloning the homepage of DeepDotWeb. These phishing websites promoted phishing domains for prevailing dark marketplaces by showing forged user comments and ratings on these websites. Any visitors without knowing the authentic surface website of DeepDotWeb has no practical way of distinguishing its authentic gray website from the phishing websites.

5 DARK WEB DOMAIN NAMES

This section describes our findings regarding how much effort the service providers put into picking their choices of onion domain names. We then explain that this trend leads to many onion domain names containing keywords that reflect the contents of the dark websites. At the same time, this trend exacerbates phishing risks for users who leverage a partial word in a domain name to vet the authenticity of the current dark website.

5.1 Domain Naming Trends

A domain name is a valuable asset especially when it comes to Internet business. For e-commerce or social network services, their domains are often considered as the brands of their businesses. Likewise, a Dark Web service provider may want to obtain a desired domain name; however, onion domain names are fundamentally different from surface domains. An onion domain name must have ".onion" as its top-level domain (TLD), and its 2LD is an 80-bit number in base32 generated from a service provider's public key. Therefore, it is computationally expensive to obtain the exact domain name that a service provider asks for.

A service provider may conduct a brute-force search to find a public key that will generate the desired 2LD. However, finding the desired one (fully matching 16 characters) is computationally infeasible. The author of a brute-force onion address generation tool [19] reported that it would take approximately 2.6 million years to obtain an address that only matches the first 14 characters of the desired domain. Therefore, a service provider might prefer to use a shorter domain prefix, which drastically reduces the domain searching time. Using the same tool, if a service provider looks for a domain whose first five characters matches his/her chosen word, it would only require approximately one minute [19].

We conducted the study to understand how Dark Web service providers perceive their onion addresses. Figure 6 visualizes the domain name similarity heat map. For horizontal and vertical axes, we lexicographically sorted all the second-level onion domain names in our dataset. We then marked a cross point when two corresponding onion domains belonged to the same website group addressed in Section 4.1. We deliberately marked no point when two onion domains on both axes are the same. Finally, we clustered marked points within each 10 by 10 size cell into a color point, which depicts the number of marked cross points. That is, a color point in the figure represents the number of onion domain pairs that share the near-identical content among 100 pairs. Therefore, the existence of

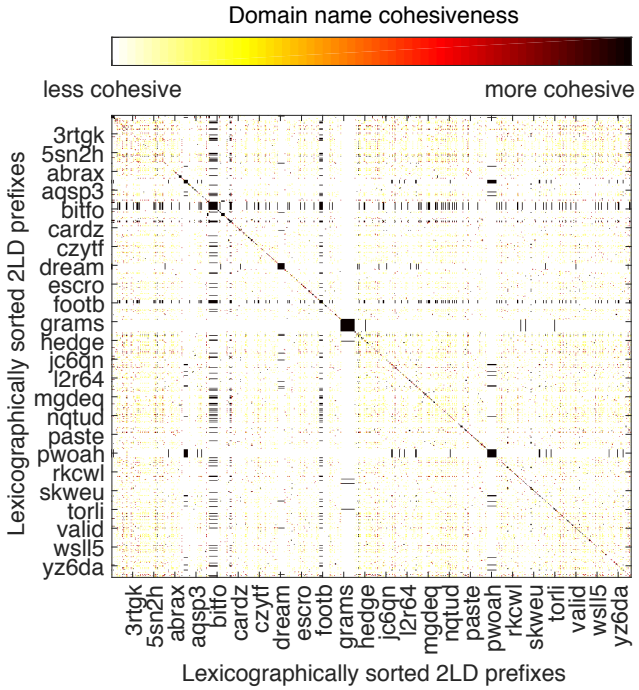

Figure 6: Name similarity heat map of 2LDs.

a clear diagonal line across the heat map of Figure 6 suggests that domain names with text-wise similar content share the same 2LD prefix.

Knowing that many similar 2LDs existed, we further examined whether such domain names were intentionally generated. We reused the website groups that were described in the previous section. We compared the 2LDs to one another for each website group and computed the longest common prefix (LCP) of each domain name combination pair. Therefore, for each website group, there are multiple LCPs out of all the possible domain name pairs. Considering that a 2LD is generated from a random public key, having a long LCP between two domain names is highly unlikely.

We found that at least 67% of the 5,718 website groups had LCPs of at least five characters long. This means that those website groups used brute-force to generate their onion domain names with their preferred prefixes. Interestingly, eight website groups selected domain names whose LCPs were 10 characters long. Generating such onion domains using Shallot [19] on a 1.5GHz processor, could take up to approximately 40 years. Based on this approximation, if a website owners uses 72 cores of Intel Xeon 2.3GHz CPU from Amazon cloud resources [8], it takes approximately 131 days and costs around $9,500.

We further analyzed the LCPs with at least five characters long. There were 184 common words (e.g., crypto, count), and 231 words were from the titles of onion domain homepages (e.g., grams7, alphabayww). Considering that these LCPs are computed from a website group with content duplicates, service providers choose a prefix of their 2LDs based on their content. In summary, a non-negligible portion of Dark Web service providers puts significant computational resources into assigning meaningful prefixes to their onion domain names.

5.2 Discussion

Each onion address is cryptographically bound to the Tor client who owns the private key corresponding to the onion address. Therefore, a correct onion domain that a user wants to visit leads to the webpage owner, who has the private key matching the domain. However, the onion domain itself is not memorable. We have observed that many service providers brute-force generate *partially memorable* onion domain names with human-readable prefixes to make their domains more intuitive.

These trends make phishing easier because people now check the authenticity of a dark website using its partial domain name. Consider two different onion domain names; bitcoinfoganfnad.onion and bitcoinfogansdse.onion. Because both of them share the common prefix "bitcoinfog", the visitors would presume that the addresses will lead them to the BitcoinFog website. However, neither of them is the authentic BitcoinFog website. To introduce one case of this type of abuse, our BitcoinFog website group has 277 domains, while one of them is the official BitcoinFog domain (foggeddriztrcar2.onion). In this website group, there were 109 onion domain names beginning with "bitfog," 40 domains with "fogged," and 19 domains with "btcfog". In the case of our AlphaBay website group (181 domains), where one of the authentic onion addresses is pwoah7foa6au2pul.onion, there were 98 domains with the common prefix "pwoah" and one domain had the longest common prefix "pwoah7foa." One light mitigation for such a phishing attack is to use a completely random onion domain so that its visitors give up on memorizing the domain and use a bookmark instead.

6 CONCLUSION

We conducted a comprehensive investigation of the phishing risks in the Tor anonymity network. We demonstrated that only 5,718 website groups with distinct contents exist among 28,928 onion domains and confirmed 901 phishing domains with content duplicates. We investigated how much effort Dark Web service providers put into selecting their onion domains and addressed that this trend brings an unfortunate side-effect of having many phishing sites with shared onion domain prefixes. We carefully designed the analysis algorithms to identity the phishing risks on the Dark Web and presented the severe phishing risk on the Dark Web, which urges further research to provide users with authentic dark Web services.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their concrete feedback. This research was supported by the Engineering Research Center Program through the National Research Foundation of Korea (NRF) funded by the Korean Government MSIT (NRF-2018R1A5A1059921).

REFERENCES

[1] Bitcoin Blender. http://bitblendervrfkzr.onion.
[2] Bitcoin Blender thread on Bitcoin Talk forum. https://bitcointalk.org/index.php?topic=436467.0.
[3] Bitcoin Fog. http://foggeddriztrcar2.onion.

[4] Bitcoin Fog thread on Bitcoin Talk forum. https://bitcointalk.org/index.php?topic=50037.0.
[5] Bitcoin Forum. https://bitcointalk.org.
[6] Darknet Takedown: Authorities Shutter Online Criminal Market AlphaBay. https://www.fbi.gov/news/stories/alphabay-takedown.
[7] DarkNetMarkets community on Reddit. https://www.reddit.com/r/DarkNetMarkets.
[8] EC2 Instance Pricing - Amazon Web Service (AWS). https://aws.amazon.com/ec2/pricing/on-demand/?nc1=h_ls.
[9] Ethical Tor Research: Guidelines. https://blog.torproject.org/ethical-tor-research-guidelines.
[10] FreshOnion: Tor Hidden Service Directory. http://zlal32teyptf4tvi.onion/.
[11] Google Custom Search JSON API. https://developers.google.com/custom-search/json-api/v1/overview.
[12] Grams, The Google Of The Dark Web Has Shuttered Operations. https://www.forbes.com/sites/zarastone/2017/12/16/grams-the-google-of-the-dark-web-has-shuttered-operations/#1ccbc5257624.
[13] Hansa Market. https://www.deepdotweb.com/marketplace-directory/listing/hansa-market/.
[14] Hidden Services need some love. https://blog.torproject.org/hidden-services-need-some-love.
[15] International Raids Target Sites Selling Contraband on the Dark Web. https://www.nytimes.com/2014/11/08/world/europe/dark-market-websites-operation-onymous.html.
[16] Making Connections to Facebook more Secure. https://www.facebook.com/notes/protect-the-graph/making-connections-to-facebook-more-secure/1526085754298237/.
[17] Microsoft Cognitive Services - Bing Web Search API. https://azure.microsoft.com/en-us/services/cognitive-services/bing-web-search-api/.
[18] Scrapy. https://scrapy.org.
[19] Shallot. https://github.com/katmagic/Shallot/.
[20] Splash - A javascript rendering service. https://splash.readthedocs.io/en/stable/.
[21] The New York Times is Now Available as a Tor Onion Service. https://open.nytimes.com/https-open-nytimes-com-the-new-york-times-as-a-tor-onion-service-e0d0b67b7482.
[22] Venezuela is blocking access to the Tor network. https://www.theverge.com/2018/6/25/17503680/venezuela-tor-blocked-web-censorship.
[23] Ahmia. Ahmia: Tor Hidden Service Search Engine. https://ahmia.fi/.
[24] J. Aldridge and D. Decary-Hetu. Not an'Ebay for Drugs': The Cryptomarket'Silk Road'as a Paradigm Shifting Criminal Innovation. 2014.
[25] M. J. Barratt, J. A. Ferris, and A. R. Winstock. Use of Silk Road, the online drug marketplace, in the United Kingdom, Australia and the United States. *Addiction*, 109(5):774–783, 2014.
[26] A. Biryukov and I. Pustogarov. Bitcoin over Tor isn't a good idea. In *Proceedings of the 2015 IEEE Symposium on Security and Privacy (SP)*, pages 122–134. IEEE, 2015.
[27] A. Biryukov, I. Pustogarov, F. Thill, and R.-P. Weinmann. Content and popularity analysis of Tor hidden services. In *Proceedings of the 2014 IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pages 188–193. IEEE, 2014.
[28] A. Biryukov, I. Pustogarov, and R.-P. Weinmann. Trawling for tor hidden services: Detection, measurement, deanonymization. In *Proceedings of the 2013 IEEE Symposium on Security and Privacy (SP)*, pages 80–94. IEEE, 2013.
[29] N. Christin. Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd international conference on World Wide Web*, pages 213–224. ACM, 2013.
[30] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong. Freenet: A distributed anonymous information storage and retrieval system. In *Designing privacy enhancing technologies*, pages 46–66. Springer, 2001.
[31] M. Cova, C. Kruegel, and G. Vigna. There is no free phish: An analysis of "free" and live phishing kits. In *Proceedings of the Conference on USENIX Workshop on Offensive Technologies*. USENIX Association, 2008.
[32] R. Dhamija, J. Tygar, and M. Hearst. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2006.
[33] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. Technical report, Naval Research Lab Washington DC, 2004.
[34] D. S. Dolliver. Evaluating drug trafficking on the Tor Network: Silk Road 2, the sequel. *International Journal of Drug Policy*, 26(11):1113–1123, 2015.
[35] S. Ghosh, P. Porras, V. Yegneswaran, D. Ken, and N. Ariyam. ATOL: A Framework for Automated Analysis and Categorization of the Darkweb Ecosystem. In *AAAI Workshop on Artificial Intelligence for Cyber Security (AICS)*. AAAI, 2017.
[36] X. Han, N. Kheir, and D. Balzarotti. Phisheye: Live monitoring of sandboxed phishing kits. In *ACM Conference on Computer and Communications Security*. ACM, 2016.
[37] HERMES Center for Transparency and Digital Human Rights. Tor2web: Browse the Tor Onion Services. https://tor2web.org/.
[38] The New Research from Northeastern University. https://blog.torproject.org/blog/new-research-northeastern-university.
[39] The Invisible Internet Project (I2P). https://geti2p.net/.
[40] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Commun. ACM*, 2007.
[41] M. Jakobsson and S. Myers. *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft*. Wiley-Interscience.
[42] R. Jansen, F. Tschorsch, and A. Johnson. The sniper attack: Anonymously deanonymizing and disabling the Tor network. In *Proceedings of the 21st Annual Network and Distributed System Security Symposium (NDSS)*. The Internet Society, 2014.
[43] A. Kwon, M. AlSabah, D. Lazar, M. Dacier, and S. Devadas. Circuit fingerprinting attacks: Passive deanonymization of tor hidden services. In *Proceedings of the 24th USENIX Security Symposium*, pages 287–302. USENIX Association, 2015.
[44] K. Levchenko, A. Pitsillidis, N. Chachra, B. Enright, M. Félegyházi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, et al. Click trajectories: End-to-end analysis of the spam value chain. In *Security and Privacy (S&P 2011)*, pages 431–446, 2011.
[45] K. Li, P. Liu, Q. Tan, J. Shi, Y. Gao, and X. Wang. Out-of-band discovery and evaluation for tor hidden services. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC)*, pages 2057–2062. ACM, 2016.
[46] S. Matic, P. Kotzias, and J. Caballero. Caronte: Detecting location leaks for deanonymizing tor hidden services. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 1455–1466. ACM, 2015.
[47] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 127–140. ACM, 2013.
[48] T. Moore and R. Clayton. Discovering phishing dropboxes using email metadata. In *eCrime Researchers Summit*, 2012.
[49] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.
[50] R. Overdorf, M. Juarez, G. Acar, R. Greenstadt, and C. Diaz. How Unique is Your. onion? An Analysis of the Fingerprintability of Tor Onion Services. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017.
[51] L. Overlier and P. Syverson. Locating hidden servers. In *Proceedings of the 2006 IEEE Symposium on Security and Privacy (SP)*, pages 100–114. IEEE, 2006.
[52] L. Overlier and P. Syverson. Valet services: Improving hidden servers with a personal touch. *Lecture Notes in Computer Science*, 4258:223–244, 2006.
[53] G. Owen and N. Savage. Empirical analysis of Tor hidden services. *IET Information Security*, 10(3):113–118, 2016.
[54] A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. ACM, 2006.
[55] T. Ruffing, P. Moreno-Sanchez, and A. Kate. P2p mixing and unlinkable bitcoin transactions. In *NDSS*, 2017.
[56] A. Sanatinia and G. Noubir. Honey onions: a framework for characterizing and identifying misbehaving tor hsdirs. In *Proceedings of the 2016 IEEE Conference on Communications and Network Security (CNS)*, pages 127–135. IEEE, 2016.
[57] I. Sanchez-Rola, D. Balzarotti, and I. Santos. The Onions Have Eyes: A Comprehensive Structure and Privacy Analysis of Tor Hidden Services. In *Proceedings of the 26th International Conference on World Wide Web (WWW)*, pages 1251–1260. ACM, 2017.
[58] K. Soska and N. Christin. Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem. In *Proceedings of the 24th USENIX Security Symposium*, pages 33–48. USENIX Association, 2015.
[59] B. Stone-Gross, T. Holz, G. Stringhini, and G. Vigna. The underground economy of spam: A botmaster's perspective of coordinating large-scale spam campaigns. In *Proceedings of the Conference on Large-scale Exploits and Emergent Threats*. USENIX Association, 2011.
[60] P. Syverson and G. Boyce. Bake in. onion for tear-free and stronger website authentication. *IEEE Security & Privacy*, 14(2):15–21, 2016.
[61] K. Thomas, F. Li, A. Zand, J. Barrett, J. Ranieri, L. Invernizzi, Y. Markov, O. Comanescu, V. Eranti, A. Moscicki, D. Margolis, V. Paxson, and E. Bursztein. Data breaches, phishing, or malware? understanding the risks of stolen credentials. In *ACM Conference on Computer and Communications Security*. ACM, 2017.
[62] M. Tran, L. Luu, M. S. Kang, I. Bentov, and P. Saxena. Obscuro: A bitcoin mixer using trusted execution environments. *IACR Cryptology ePrint Archive*, 2017:974, 2017.
[63] M. C. Van Hout and T. Bingham. âĂŸSilk RoadâĂŹ, the virtual drug marketplace: A single case study of user experiences. *International Journal of Drug Policy*, 24(5):385–391, 2013.
[64] S. Zawoad, A. Dutta, A. Sprague, R. Hasan, J. Britt, and G. Warner. Phish-net: Investigating phish clusters using drop email addresses. In *APWG eCrime Researchers Summit*, 2013.