Discovering and Measuring Malicious URL Redirection Campaigns from Fake News Domains

Zhouhan Chen
New York University
zc1245@nyu.edu

Juliana Freire
New York University
juliana.freire@nyu.edu

***Abstract*—Malicious URLs are used to distribute malware and launch social engineering attacks. They often hide behind redirection networks to evade detection. Due to the difficulty in discovering redirection traffic in real-time, previous approaches to understanding redirection networks were reactive and passive. We propose a proactive algorithm that is able to uncover redirection networks in real-time given a small set of seed domains. Our method works in three steps: (1) collecting redirection paths, (2) clustering domains that share common nodes along redirection paths, and (3) searching for other domains co-hosted on similar IP addresses. We evaluate our method using real websites that we discovered while auditing 2,300 popular fake news sites. We seeded our algorithm with a subset of 276 fake news domains that redirect, and uncovered three large-scale redirection campaigns. We further verified that 91% of entry point domains were not new, but recently expired, re-registered, and parked on dedicated hosts.**

To mitigate this threat vector, we deployed our system to automatically collect newly re-registered domains and publish new redirection networks. During a five-month period, our threat intelligence reports have received over 50,000 Google Search impressions, and have been recommended by commercial vendor tools. We also reported findings to Google and Amazon Web Services, both of which have acted promptly to remove malicious artifacts. Our work offers a viable approach to continuously discover evasive redirection traffic from re-registered domains.

***Index Terms*—URL redirection, domain registration, fake news, expired domain, redirection campaign, proactive discovery**

I. Introduction

Malicious URLs are evasive, short-lived, and usually hide behind redirection networks [1], [2], making automatic and continuous discovery an important but difficult task. One major challenge is how to design an algorithm that can uncover large number of malicious URLs based on a small set of seeds. For example, previous work used suspicious keywords (e.g., *LV, GUCCI*) as seeds to identify pages that abuse search engine optimization algorithms [3], or used advertisement platforms as seeds to crawl fake advertisements [4].

To continuously discover URL redirection traffic, we focus on a distinct group of domains – those that are expired, re-registered, and redirect upon visit. The motivation came from our investigation that 12.22% (276 out of 2,300) of previously reported fake news domains were subsequently abandoned, and re-registered by actors that use redirection to push unwanted content [5]. Recent work shows that domain re-registration is a large-scale, lucrative, and suspicious business model [6]. For example, [7] concludes that 10% of all *.com* domains are re-registered on the same day they expire, and [8] shows that re-registered domains are associated with higher level of malicious activities. Most work in domain re-registration have so far focused on retrospective trend analysis instead of proactive malware discovery.

We propose and implement a discovery algorithm to uncover large redirection campaigns based on a set of seed domains. Our algorithm has three steps: (1) we crawl and construct redirection path of each seed domain; (2) we cluster domains that share common intermediate nodes together; and (3) for each cluster identified in step two, we expand the search scope to collect other domains co-hosted on the same IP address(es), an infrastructure commonly shared by parked domains [9]. We can then reapply step (1) and (2) to obtain larger clusters. Our approach is **agnostic** to the topics of the websites – it just requires a set of seeds to bootstrap the exploration.

We evaluate our algorithm using 276 fake news domains as seeds, and discover three active URL redirection campaigns. The largest one has more than 4,500 entry point domains, 91% of them re-registered. To improve our understanding of how and where each campaign operates, we contextualize its traffic flow using domain WHOIS records and IP geo-locations.

To continuously surface redirection networks, we deploy our system to crawl newly re-registered domains hosted on popular bullet-proof hosting services. Our project is available at https://zhouhanc.github.io/malware-discoverer. Our daily threat intelligence reports have been recommended by commercial vendor tools such as VirusTotal and RiskIQ. We also reported our findings to affected parties including Google and Amazon Web Services, both of which have acted promptly in response. Our work provides a new path to measure hard-to-discover redirection traffic. Our discoveries also help to reduce the time it takes to mitigate and remediate malicious activities.

II. Redirection campaign discovery algorithm

In this section, we present our **crawl-cluster-expand** strategy that uncovers redirection campaigns in an almost unsupervised manner. On a high level, the input to our system is a small set of suspicious domains that redirect, and the output of our system is one or several larger URL redirection campaigns.

A. Crawl redirection path

Tracing redirection traffic is a challenging task due to cloaking. Cloaking is when a website serves different payloads to different requests. Common cloaking techniques include client-side JavaScript execution, IP rate limiting and IP blocklists (for example, IP addresses from security companies and universities are often blocked). [3] reported that cloaking is highly effective to stop automated crawlers, and is widely deployed by malicious websites to evade detection.

To address these challenges, we run a headless Python Selenium Chrome Browser to emulate human behavior. We route each request through a virtual private network from VyprVPN, which provides servers from more than 70 cities in 5 continents [10]. We use OpenVPN [1] to automatically update hostname and IP address every 10 requests. Empirically, we start to observe IP ban when the number of requests exceeds this threshold. Given a URL, our crawler first visits the home page, then follows all redirections and saves the redirection path. Our crawler also saves each final landing page in text (HTML) and image (PNG screenshot) formats. Our crawler does not perform keyboard operations such as clicking or typing. Domains that require human interaction to redirect is not covered by our system, because it is very challenging and error-prone to automatically select the HTML element that triggers a redirection.

B. Cluster similar domains

To facilitate the discussion of domain clustering algorithm, we introduce the concept of ***tier***. We focus on three tiers of domains:

1) **Tier 1** domain is the first node along a redirection path. It is also referred to as entry point domain.
2) **Tier 2** domain is the second node along a redirection path. Those domains are usually backbones of a redirection network: aggregating requests from tier 1 domains and relaying requests to domains at the next tier. Topologically, tier 2 domains have high in-degrees.
3) **Tier 3** domain is the last node along a redirection path (*not the third node*). They are also referred to as final landing domains.

The domain clustering algorithm aims to identify tier 1 domains that belong to the same redirection cluster. Mathematically, given a set of domains and their corresponding domain-level redirection paths, our algorithm constructs a directed acyclic graph. Each node is a domain, and two nodes are connected if one redirects to another. Empirically, we find that tier 2 domains are widely used as C&C servers and we cluster tier 1 domains together if they share the same tier 2 domains. To identify important networks, we filter out tier 2 domains whose in-degree is smaller than a threshold n, a user-defined cutoff. In our real-world experiment, we set $n = 10$. The output of this step is a list of clusters. Within each cluster, there are at least n redirection paths, all sharing a common tier 2 domain.

[1]https://github.com/OpenVPN/openvpn

In extreme cases, an intermediate domain may redirect back to another intermediate domain, making the redirection graph cyclic. When that happens, our algorithm recursively removes the domain with the lowest in-degree, until there is no cycle. In our real-world study, this situation rarely happens.

C. Search expansion

The previous step generates clusters of domains that share the same redirection infrastructure (in our case, common tier 2 domains). If we only start with a small set of seeds, the size of each cluster will be small. How do we discover domains not in our seed but potentially belong to the same cluster? We decide to expand our search scope by collecting domains co-hosted on the same IP address(es), because IP is a relatively costly resource shared by many malicious URLs [9]. Previous work also explored other infrastructural signals to cluster domains, such as domain name registrar [11]. We find that registrar is a low-quality signal in our real-world data collection. In Section III, we show that domains belonging to one campaign are registered at more than 500 registrars.

Our search expansion algorithm works in the following way: given a set of domains that belong to the same redirection network, we use ViewDNS.info to first search for IP addresses those domains have been hosted on during the past 24 hours. Then for each IP, we search the most recent domains hosted on that IP. We join all co-hosted domains together into one set. Because ViewDNS.info does not support reverse lookup for IPv6, in this paper we only focus on IPv4 address space.

D. From cluster to campaign

Finally, to reveal larger redirection campaigns, we use the newly aggregated set of domains from step C as seeds, and reapply step A and B. We further aggregate clusters together if there is a link from one cluster to another. This aggregation is useful to reconstruct redirection campaigns that use multiple tier 2 domains to deliver malicious payloads hosted on multiple final landing domains.

III. Discovering redirection campaigns from real-world fake news sites

To evaluate our method, we bootstrap our algorithm with a set of fake news domains that redirect. In this section, we first give an intuition of why we study fake news domains, then explain our findings and analyze redirection traffic. We remind readers that our algorithm is agnostic to the topic of websites. In the future, we plan to explore other approaches to generate seeds. In retrospect, fake news domains are short-lived and are likely to become expired, making them good candidates to investigate.

A. Harvesting Seeds from Fake News Domains

Recent research in fake news discovery focuses on sites that pose social risks [5]. Those sites contain misinformation and are shared across online platforms. However, there is another group of fake news domains that pose cyber risks by redirecting users to malware sites.

TABLE I
SUMMARY STATISTICS OF THREE DISCOVERED REDIRECTION CAMPAIGNS, WITH VARYING LEVELS OF SOPHISTICATION.

campaign	example seed domain	number of domains	cloaking	still active (as of February 1, 2021)
1	nycpost.pro	37	no	no
2	cnnews3.com	750+	yes	yes
3	fox-news24.com	4500+	yes	yes

The boundary between social risk and cyber risk is fluid. For example, a site may initially contain fake content, then become abandoned, and subsequently re-registered by actors that use redirection to push unwanted software. Because there is no systematic study of how many fake news sites are abandoned and abused, we decide to audit fake news sites, in the hope of identifying suspicious seeds.

We first compile a list of 2,300 known fake news domains from five sources: Media Bias/Fact Check [12], Politifact [13], BuzzFeed [14], Opensources Fake News Corpus [15] and MIT Fake News Dataset [16]. We then crawl each domain between March 11, 2020 and March 12, 2020. Only 67.6% domains land on home page; **12.22% (276)** sites redirect to another domain; 10.96% domains do not have valid IP addresses; 3.22% domains are on sale; and the rest 5.92% domains return a non-200 response code. Because our discovery algorithm operates on domains that redirect, we use these 276 domains as seeds to bootstrap our algorithm.

B. Uncovering Redirection Networks

We apply our discovery algorithm using the 276 fake news domains as seeds. We collect all redirection paths between March 13, 2020 and March 14, 2020. We filter out campaigns with fewer than ten tier 1 domains, as those small networks are too small and not malicious. In the end, we identify three redirection campaigns with varying network resilience, summarized in Table I. The two largest campaigns are globally distributed, use cloaking to bypass detection, and are still active as of February 1, 2021. We now analyze each campaign in detail.

C. Characteristics of each campaign

Campaign 1 consists of 37 tier 1 domains, all registered at Namecheap and hosted on Bluehost. Tier 2 domains are hosted on DataWeb Global. Tier 3 domains are hosted on content distribution networks including AWS and Cloudflare. Most final landing pages encourage users to install suspicious Chrome extensions. We reported all 37 domains to Bluehost on March 14, 2020. In an email, Bluehost notified us that "we've taken necessary action based on your report. Due to our privacy policy, we can not specify the exact action we've taken." We independently verified that all 37 domains were suspended.

Campaign 2 is a network of more than 750 domains. Both tier 1 and tier 2 domains are hosted on Trellian, an Australian domain monetization company, and Above, a subsidiary of Trellian. We sent our findings to the abuse reporting email provided by Trellian, but did not receive any response. It is not clear if the company is aware that domains hosted on its servers are part of a malicious redirection network. Finally, most tier 3 domains are hosted on AWS and Digital Ocean, and point to malicious Chrome extensions. Campaign 2 deploy cloaking techniques including client-side JavaScript execution and IP rate limiting. Figure 1 visualizes the traffic flow.

Campaign 3 consists of more than 4500 tier 1 domains. Other than cloaking strategies implemented by Campaign 2, domains from Campaign 3 also use fast-flux to change DNS records rapidly, making it difficult to pinpoint server locations [17]. According to our snapshot of collected data, tier 1 domains are hosted on at least 5 providers across the US and Europe, including Sharktech, LeaseWeb, and NForce Entertainment. Certain "bullet-proof" providers, such as LeaseWeb, have previously been identified to host malicious files [11]. Different from Campaign 2, most tier 2 domains from Campaign 3 are hosted on AWS, and are not associated with any domain parking company. Finally, tier 3 domains are mostly hosted on AWS, and point to malicious Chrome extensions, drive-by downloads and unwanted adware, as shown in Figure 2. We report discovered malicious artifacts to affected parties, including Google, Amazon and Apple. We document their responses in Section V.

Insights on redirection campaigns. From an infrastructure point of view, abusers prefer to host tier 1 domains on either domain parking companies or low-reputation hosting providers. In contrast, tier 3 domains are mostly hosted on AWS or Digital Ocean, because abusers can leverage free cloud storage and free subdomain generation to distribute malware. Sophisticated users also host domains on multiple countries. According to IP geo-location data, we find that domains in campaign 1 are all located in the United States, domains from campaign 2 are hosted in Australia and the United States, while domains from campaign 3 are hosted across North America and Europe, making it difficult to take down the entire infrastructure.

D. Are entry point domains re-registered?

Earlier in this section, we mention that previously operational fake news sites were subsequently abandoned, and re-registered by actors that use redirection to push unwanted content. Is this true for all domains we discovered? We focus on two largest campaigns (2 and 3), and use historical DNS records[2] to validate our hypothesis that most tier 1 domains are re-registered.

There are three DNS states: **new** means a domain is registered and added to a DNS server; **transfer** means a registered domain is transferred to another DNS server; **deleted** means a domain is expired and removed from the DNS server. Consequently, there are three types of DNS changes: new→transfer, transfer→transfer, and deleted→new. **Deleted→new** indicates that a domain is expired and re-registered (colloquially referred

[2]We collect public historical DNS records from http://www.hosterstats.com

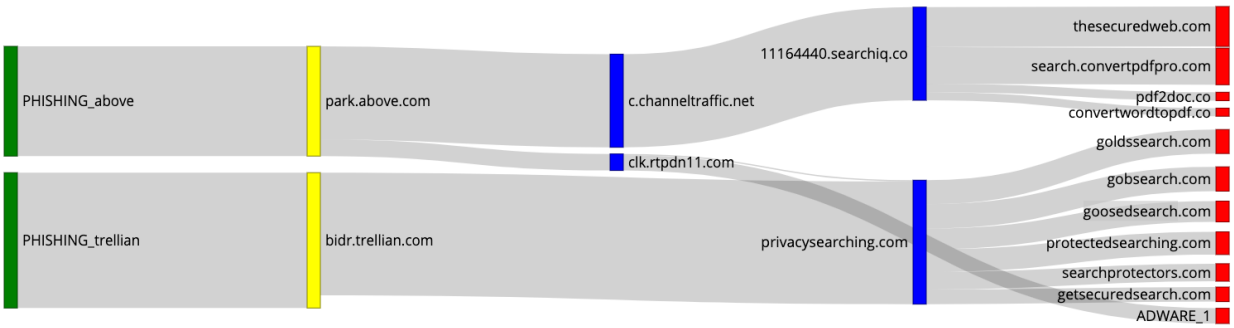

Fig. 1. Redirection network of Campaign 2. We color tier 1, 2, 3 domains as green, yellow, red, and other domains as blue. This network consists of 750 tier 1 domains. To facilitate visualization, we collapse tier 1 domains into one node if they share common tier 2 domains. For example, PHISHING_above refers to all domains that redirect to park.above.com. In Campaign 2, both tier 2 URLs belong to a same domain monetization company.

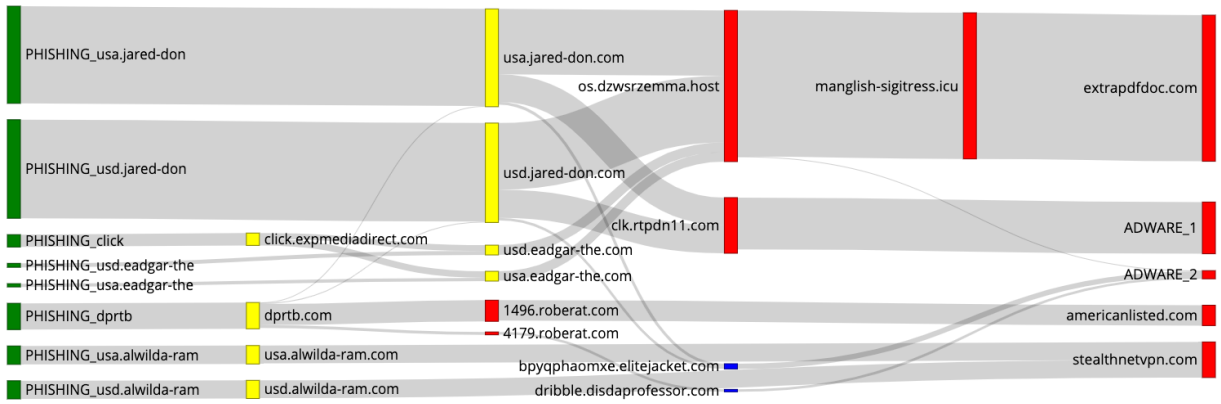

Fig. 2. Redirection network of Campaign 3. This campaign consists of more than 4500 tier 1 domains. Most tier 2 domains have common subdomain structure such as *usa.*, *usd.*. Final landing domains point to adware on S3 buckets (collapsed due to the large number of subdomains) and malicious Chrome extensions.

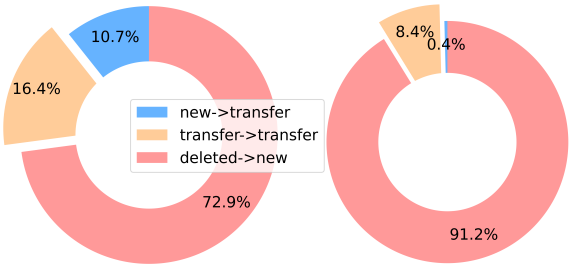

Fig. 3. DNS status change of tier 1 domains in Campaign 2 (left) and Campaign 3 (right). The most common DNS status change is deleted→new, which suggests that most domains are recently expired and re-registered.

to as "drop catch"). According to Figure 3, the vast majority of domains (72.9% in Campaign 2 and 91.2% in Campaign 3) are re-registered. The other two types of change suggest that domain owners often trade domains with each other, or move domain from one DNS server to another.

As a concrete example, Figure 4 illustrates how fox-news24.com, a fake news domain that is now part of Campaign 3, became malicious after it was re-registered. According to WHOIS record, fox-news24.com was initially created in February, 2017, and was immediately used as a channel to spread false information. We find Twitter posts that link to fox-news24.com, as well as articles that fact-check claims made by the website. The last relevant tweet we can find was created in April, 2018. Then in July, 2018, the domain's DNS server was changed, and the site started to redirect users to malicious sites.

When we look at topics of re-registered domains, we do not find evidence that domain buyers are only targeting fake news domains. Instead, they seem to re-register any expired domains available on the market to monetize them. We also cannot identify a single dominant domain name registrar. In fact, the 4,500 tier 1 domains belonging to Campaign 3 come from 500 unique registrars, which is why clustering domains based on registrar feature will not work in our case.

IV. Continuous and Proactive Discovery

Previous models to detect redirection traffic are mostly supervised [11], [18]. In contrast, our discovery protocol is nearly unsupervised (the algorithm only requires a set of seeds). As a result, we often do not have ground truth at the time of discovery. We argue that our approach complements supervised methods – by discovering unseen URL redirection traffic, we can contribute to and enhance current blocklists, and provide threat intelligence in real-time to human analysts. In the rest of this section, we explain how we deploy our system

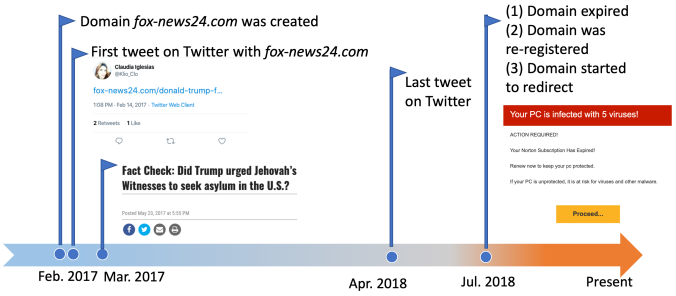

Fig. 4. DNS history and activity associated with fox-news24.com. The domain was initially spreading misinformation on Twitter. After expiration, the domain was re-registered and became an entry point of redirection Campaign 3.

to facilitate real-time exchange of threat intelligence, and the positive impact our system has had on the community.

Before deployment, we first identified seven high-risk IP addresses – 64.32.8.68, 64.32.8.70, 37.48.65.149, 37.48.65.151, 207.244.67.215, 207.244.67.218, and 103.224.182.207. Those addresses belong to the top four "bullet-proof" hosting providers we discovered in Section III – Sharktech (sharktech.net), LeaseWeb (leaseweb.com), NForce Entertainment (nforce.com), and Above (above.com). Each address hosts a dynamic list of re-registered or parked domains.

We also modify the execution order of our algorithm: for each IP address, we first use ViewDNS.info to get all domains hosted on that IP during the past 72 hours. We find that domains timestamped outside of 72 hours become stale and inaccessible. We then crawl each domain's redirection path, cluster domains together if they share the same tier 2 node, and aggregate clusters into one campaign. If there are more than one campaign, we pick the largest connected component. We repeat this protocol four times, using four different user-agents: Chrome, Safari, Android, and IPhone[3]. This allows us to extract a variety of malicious artifacts. Finally, our system produces a contextualized report that includes top-abused domains, IP addresses, visualization of the redirection traffic, and screenshots of final landing sites. We publish our daily report on Github.

From October 2020 to March 2021, our Github repository has received more than 50,000 Google search impressions and 1300 clicks. Some of the most often searched queries are questions such as "*what is allbestsecureus.com*" (allbestsecureus.com is a domain we discovered that links to malicious mobile apps). Our reports have also been recommended by antivirus engines such as Virustotal and RiskIQ.[4].

V. Responsible Disclosure

To better understand how redirection campaigns target both desktop and mobile users, we re-ran experiments in Section III: for every entry point domain in each campaign, we re-crawled its redirection path four times, using four user-agents (Chrome, Safari, Android, and IPhone). We studied the final HTML pages, and identified three major types of malicious artifacts: browser extensions, drive-by-downloads, and fleeceware, which are mobile applications that trick users to pay high subscription fees upon installation [19]. Accordingly, we contacted the following three companies:

Google. A large number of desktop requests landed on websites that ask users to add an extension to Chrome. We used regular expression to extract hyperlinks with subdomain *chrome.google.com*. In total, we found 14 extensions with an aggregated downloads of 330,000. Those extensions claim to enhance a user's search experience. However, under the hood they set aggressive permissions, including the ability to read all web requests and execute JavaScript from arbitrary domains. We reported all extensions to Google Safe Browsing Team. We verified on December 2020 that all extensions have been taken down.

Amazon. Amazon Web Services (AWS) is targeted by abusers in two ways. First, Amazon Elastic Compute Cloud (EC2) instances are used as intermediate servers to redirect traffic to malicious sites. Second, Amazon Simple Storage Service (S3) is used to host drive-by downloads. We reported our findings to the AWS Trust and Safety Team and presented them in July 2020. When we followed up in October 2020, the Team informed us that "the abuse case created by us is closed." We independently verified that the URLs we reported no longer exist.

Apple. We identified nine active fleeceware on Apple Store. To extract app ID, we used regular expression to match hyperlinks that start with *apps.apple.com*. According to SensorTower, those fleeceware have an aggregated monthly download of 800,000 times, and an aggregated monthly revenue of $1,000,000. We reported all apps to Apple App Review Team, but did not receive any response. All fleeceware are still available as of March 2021.

VI. Related Work

Our research weaves together multiple abuse vectors including domain re-registration, URL redirection, and malicious software distribution. We show that these problems do not exist in isolation, but overlap with each other in complex ways.

Domain re-registration. Domain re-registration is when expired domains are resold on the Internet. Though the business model is legitimate, it is often abused: [6], [8] observe that many domain re-registrations happen soon after expiration, and [7] estimates that 10% of all .com domains are re-registered on the same day as their old registration is expired. Furthermore, [9] infiltrates into the domain parking network, and observes the presence of click fraud, traffic spam and traffic stealing.

URL redirection and malware distribution. [3] points out that URL redirection is a salient feature of websites that want to evade detection. Multiple data sources have been explored: [18] collects redirection chains from browser history, [20] uses honey pots to harvest malicious redirect URLs, and

[3]The exact user-agent string comes from: https://www.whatismybrowser.com/guides/the-latest-user-agent/

[4]More details are available on Github: https://zhouhanc.github.io/malware-discoverer/image/virustotal_recommendation.png, https://zhouhanc.github.io/malware-discoverer/image/riskiq_recommendation.png

[21] analyzes drive-by downloads by deploying sensors at a university network. Previous research also uses network-based approaches to uncover malware campaigns [11] and to measure illicit traffic monetization [22]. Our proposed algorithm combines a high-yield data source (re-registered domains) with network-based clustering, and is more proactive than existing supervised methods.

VII. Limitations and next steps

Certain assumptions we make limit the type of redirection networks we detect. For example, if every entry point domain has a unique tier 2 domain, then our discovery algorithm will not find any meaningful cluster. To address this issue, we can modify step 2 and cluster redirection paths based on the third, fourth, or final landing domain. Another countermeasure against our detection is to host every entry point domain on a unique IP address. If that is the case, we can modify step 3 (search expansion) to look for domains sharing other signals, such as DNS server, IP subnet (blocks of IP spaces), shared IP prefixes or shared autonomous system numbers (ASNs). One challenge of using a weaker or fuzzier signal is that we might trace more benign domains. In the future, we plan to explore more robust strategies to improve recall (number of discovered malicious sites) without compromising accuracy.

VIII. Conclusion

In this paper we present an algorithm that proactively uncovers malicious URL redirection networks. We evaluate our method on a list of fake news domains that redirect, and discover three large-scale redirection campaigns. We visualize redirection traffic and offer key insights. We then deploy our algorithm to continuously discover redirection campaigns. We disclose our findings to affected technology companies, and plan to keep sharing intelligence with the security community. Our work points to an urgent need to enforce stricter policies against deceptive software distribution via URL redirection.

References

[1] T. Nelms, R. Perdisci, M. Antonakakis, and M. Ahamad, "Towards measuring and mitigating social engineering software download attacks," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 773–789.

[2] B. Srinivasan, A. Kountouras, N. Miramirkhani, M. Alam, N. Nikiforakis, M. Antonakakis, and M. Ahamad, "Exposing search and advertisement abuse tactics and infrastructure of technical support scammers," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018, p. 319–328.

[3] L. Invernizzi, K. Thomas, A. Kapravelos, O. Comanescu, J. Picod, and E. Bursztein, "Cloak of visibility: Detecting when machines browse a different web," in *2016 IEEE Symposium on Security and Privacy (SP)*, 2016, pp. 743–758.

[4] P. Vadrevu and R. Perdisci, "What you see is not what you get: Discovering and tracking social engineering attack campaigns," in *Proceedings of the Internet Measurement Conference*, ser. IMC '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 308–321.

[5] Z. Chen and J. Freire, "Proactive discovery of fake news domains from real-time social media feeds," in *Companion Proceedings of the Web Conference 2020*, ser. WWW '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 584–592.

[6] C. Lever, R. Walls, Y. Nadji, D. Dagon, P. McDaniel, and M. Antonakakis, "Domain-z: 28 registrations later measuring the exploitation of residual trust in domains," in *2016 IEEE Symposium on Security and Privacy (SP)*, 2016, pp. 691–706.

[7] T. Lauinger, A. Chaabane, A. S. Buyukkayhan, K. Onarlioglu, and W. Robertson, "Game of registrars: An empirical analysis of post-expiration domain name takeovers," in *Proceedings of the 26th USENIX Conference on Security Symposium*, ser. SEC'17. USA: USENIX Association, 2017, p. 865–880.

[8] T. Lauinger, K. Onarlioglu, A. Chaabane, W. Robertson, and E. Kirda, "Whois lost in translation: (mis)understanding domain name expiration and re-registration," in *Proceedings of the 2016 Internet Measurement Conference*, ser. IMC '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 247–253.

[9] S. Alrwais, K. Yuan, E. Alowaisheq, Z. Li, and X. Wang, "Understanding the dark side of domain parking," in *23rd USENIX Security Symposium (USENIX Security 14)*. San Diego, CA: USENIX Association, Aug. 2014, pp. 207–222.

[10] "Vyprvpn openvpn setup for linux (ubuntu)," 2020. [Online]. Available: https://support.vyprvpn.com/hc/en-us/articles/360037721812-VyprVPN-OpenVPN-Setup-for-Linux-Ubuntu-

[11] Z. Li, S. Alrwais, Y. Xie, F. Yu, and X. Wang, "Finding the linchpins of the dark web: a study on topologically dedicated hosts on malicious web infrastructures," in *2013 IEEE Symposium on Security and Privacy*, 2013, pp. 112–126.

[12] "Media bias fact check," 2021. [Online]. Available: https://github.com/drmikecrowe/mbfcext

[13] Politifact, "Politifact's fake news almanac," 2017. [Online]. Available: https://www.politifact.com/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/

[14] BuzzFeedNews, "Analysis of fake news sites and viral posts," 2018. [Online]. Available: https://github.com/BuzzFeedNews/2018-12-fake-news-top-50

[15] "Opensources fake news corpus," 2017. [Online]. Available: https://github.com/several27/FakeNewsCorpus

[16] H. Allcott and M. Gentzkow, "Replication data for: Social media and fake news in the 2016 election," 2017. [Online]. Available: https://www.openicpsr.org/openicpsr/project/113992/

[17] A. Caglayan, M. Toothaker, D. Drapeau, D. Burke, and G. Eaton, "Real-time detection of fast flux service networks," in *2009 Cybersecurity Applications Technology Conference for Homeland Security*, 2009, pp. 285–292.

[18] G. Stringhini, C. Kruegel, and G. Vigna, "Shady paths: Leveraging surfing crowds to detect malicious web pages," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, ser. CCS '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 133–144.

[19] J. Chandraiah, "Don't let fleeceware sneak into your iphone," *Sophos*, 2020. [Online]. Available: https://news.sophos.com/en-us/2020/04/08/iphone-fleeceware/

[20] M. Akiyama, T. Yagi, T. Yada, T. Mori, and Y. Kadobayashi, "Analyzing the ecosystem of malicious url redirection through longitudinal observation from honeypots," *Computers & Security*, vol. 69, pp. 155 – 173, 2017, security Data Science and Cyber Threat Management.

[21] T. Nelms, R. Perdisci, M. Antonakakis, and M. Ahamad, "Webwitness: Investigating, categorizing, and mitigating malware download paths," in *24th USENIX Security Symposium (USENIX Security 15)*. Washington, D.C.: USENIX Association, Aug. 2015, pp. 1025–1040.

[22] B. Liu, Z. Li, P. Zong, C. Lu, H. Duan, Y. Liu, S. Alrwais, X. Wang, S. Hao, Y. Jia, Y. Zhang, K. Chen, and Z. Zhang, "Traffickstop: Detecting and measuring illicit traffic monetization through large-scale dns analysis," in *2019 IEEE European Symposium on Security and Privacy (EuroS P)*, 2019, pp. 560–575.