Understanding the Dark Side of Domain Parking

Sumayah Alrwais, *Indiana University Bloomington and King Saud University;* Kan Yuan, *Indiana University Bloomington;* Eihal Alowaisheq, *Indiana University Bloomington and King Saud University;* Zhou Li, *Indiana University Bloomington and RSA Laboratories;* XiaoFeng Wang, *Indiana University Bloomington*

https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/alrwais

This paper is included in the Proceedings of the
23rd USENIX Security Symposium.

August 20–22, 2014 • San Diego, CA

ISBN 978-1-931971-15-7

Understanding the Dark Side of Domain Parking

Sumayah Alrwais[1,2], Kan Yuan[1], Eihal Alowaisheq[1,2], Zhou Li[1,3] and XiaoFeng Wang[1]

[1]Indiana University, Bloomington
{salrwais, kanyuan, ealowais, xw7}@indiana.edu
[2]King Saud University, Riyadh, Saudi Arabia
[3]RSA Laboratories
zhou.li@rsa.com

Abstract

Domain parking is a booming business with millions of dollars in revenues. However, it is also among the least regulated: parked domains have been routinely found to connect to illicit online activities even though the roles they play there have never been clarified. In this paper, we report the first systematic study on this "dark side" of domain parking based upon a novel infiltration analysis on domains hosted by major parking services. The idea here is to control the traffic sources (crawlers) of the domain parking ecosystem, some of its start nodes (parked domains) and its end nodes (advertisers and traffic buyers) and then "connect the dots", delivering our own traffic to our end nodes across our own start nodes with other monetization entities (parking services, ad networks, etc) in-between. This provided us a unique observation of the whole monetization process and over one thousand *seed* redirection chains where some ends were under our control. From those chains, we were able to confirm the presence of click fraud, traffic spam and traffic stealing. To further understand the scope and magnitude of this threat, we extracted a set of salient features from those seed chains and utilized them to detect illicit activities on 24 million monetization chains we collected from leading parking services over 5.5 months. This study reveals the pervasiveness of those illicit monetization activities, parties responsible for them and the revenues they generate which approaches 40% of the total revenue for some parking services. Our findings point to an urgent need for a better regulation of domain parking.

1 Introduction

Consider that you are a domain owner, holding a few domain names that you do not have a better use of. Then one thing you could do is to "park" them with a domain parking service to earn some extra cash: whenever web users type in those domain names (probably accidentally) in the browser's address bar, the parking service resolves the domains to advertisement laden pages, the revenue generated in this way is then split between the parking service and the domain owner. Such *domain parking monetization* is a million-dollar business [29], offering a unique marketing channel through newly acquired, underdeveloped domains, or those reserved for future use. However, with a large number of parked domains being monetized through those services (`Sedo` reported to have 4.4M parked domains in 2013 [29]), what becomes less clear is the security implications of such activities, particularly whether they involve any illicit operations, a question we attempt to answer.

Dark side of domain parking. Prior research shows that once malicious domains (e.g., those hosting a Traffic Distribution System, TDS) have been discovered, they often end up being parked [17], i.e. temporarily hosted by some domain parking services. Those domains come with a large number of backlinks through which they can still be visited by the victims of the malicious activities they were once involved in, such as compromised websites. Web traffic from those backlinks is clearly of low quality but apparently still used by the parking services to make money [17]. More problematically, a recent study reveals suspicious redirections performed by some parking services could be related to click spam [10], though this could not be confirmed. This finding echoes what we observed from the redirection chains collected by crawling parked domains, some of the URLs on the chains carried URL patterns related to ad click delivery even though our crawler did not click on any link at all (Section 2.3). Also malicious web content has long been known to propagate through parked domains [14, 20].

With all such suspicions raised, it is still challenging to determine whether some parking services are indeed involved in illicit activities, and if so the types of roles they play there. The problem is that parking services' monetization decisions and strategies cannot be directly observed from the outside. Therefore, in the absence of information about the nature of input traffic to those services and the actual ways it has been monetized, all we have is guesswork. For example, even though we did observe redirection chains seemingly related to click delivery, without knowing a parking service's interactions with its ad network, we were still left with little evidence that what we saw was indeed click fraud. Note that parking services do not need to play conventional tricks, such as running click bots [21] or using hidden iframes embedded in a compromised page, to generate fraudulent clicks, and therefore will not be caught by standard ways in click fraud detection. In the presence of such technical challenges, little has been done so far to understand the illicit activities that can happen during the monetization of parked domains.

Our study. In this paper, we report the first attempt to explore the dark side of domain parking and uncover its security implications. This expedition is made possible by an innovative methodology to infiltrate parking services. More specifically, we purchased a set of domains and parked them with those services. Those domains, together with our crawler that continuously explored parked domains, enable us to control some inputs to the parking services. On the receiving end, we launched ad campaigns and made purchases of web traffic through the ad network or the traffic sellers associated with those parking services. By carefully tuning parameters to target web audience we controlled, we were able to connect the dots, receiving the traffic generated by our crawler going through our parked domains and onto our campaign websites. This placed us at a unique vantage point, where we could observe complete monetization chains between the start and the end nodes we controlled.

By analyzing such monetization chains, we were surprised to find that domain parking services, even highly popular ones such as PS5[1], shown in Table 1, are indeed involved in less-than-legitimate activities that should never happen: our ad campaigns were charged for the "click" traffic produced by our own crawler that never clicked and the traffic we purchased turned out to have nothing to do with the keywords we specified; also interestingly, we found that for all the visits through our parked domains and hitting our ads, only some of them were reported to our domain-owner's account (i.e. their revenues shared with us), though all of them were billed to our ad account.

To further analyze the scope and magnitude of those problems, which we call *click fraud*, *traffic spam* and *traffic stealing* respectively, we fingerprinted those confirmed illicit monetization chains with a set of salient features called *stamps*, and utilized them to identify monetization activities on 24M visits to parked domains not going through our end nodes (i.e. ad/traffic campaign websites). New findings reveal that even leading domain parking services are involved in illegitimate operations. On the other hand, such operations were present in only about 5% of the traffic we observed, which indicates that those services are largely legitimate. Possible motivations behind their opportunistic attacks could be monetizing less reputable (secondary) domains (e.g., taken-down malicious domains) that are difficult to profit from legitimately, or making up for revenue losses. Furthermore, we conducted an economic study to estimate the revenues of such dark-side activities, which we found to be significant.

Contributions. The contributions of the paper are outlined as follows:

• *New methodologies*. We performed the first systematic study of illicit activities in parked domain monetization. This study was made possible by a suite of new methodologies that allowed us to infiltrate domain parking services and collect a set of complete monetization chains. We further expanded such "seed" chains over a large number of redirection chains collected from parked domains over a 5.5-month period, which laid the foundation for understanding the unique features and the impacts of those activities.

• *New findings*. Our study brought to light a set of interesting and important findings never reported before. Not only did we confirm the presence of illegitimate operations including click fraud, traffic spam and traffic stealing during the monetization of parked domains, but we also reveal the pervasiveness of those activities which affect most leading parking service providers and attribute it to account for up to 40% of their total revenue. Also interesting is the discovery of unique features of those illicit activities and their relations with different monetization strategies and parking service syndicates.

2 Parked Domain Monetization

2.1 Background

Domain parking. As described before, a parked domain is a registered domain name whose owner does not have a better use of it than temporarily running it as an ad portal to profit from the traffic the domain receives. To this end, the owner typically chooses to *park* the domain with a domain parking service, an intermediary between the owner and various monetization options (explained later). This is done by setting up an account with the service, and the owner forwards her domain traffic to the parking service as specified by its regulations. The most common way for doing this is through the Domain Name System (DNS), in which the parked domain's Name Server (NS) or Canonical Name record (CNAME) is set to point to that of the parking service.

[1] Throughout this paper, we anonymize the identities of domain parking services found to be participating in illicit activities due to legal restrictions imposed by Indiana University and RSA.

In this way, the service gains complete control on the parked domain and any traffic it receives. Alternatively, the domain owner can choose to log her domain traffic before redirecting it to the parking service through HTTP redirections.

Parking services provide a domain owner with a platform to manage her parked domains. For example, some parking services let the owner set the keywords to be used for the parked domain monetization. Also, a domain owner can monitor their domain earnings through revenue reports.

Monetization options. A parked domain owner can profit from her domain traffic through a number of monetization options. The most popular ones are search advertising and direct-navigation traffic monetization, as elaborated below:

• *Search advertising*. In search advertising (aka., sponsored search), the advertiser runs a textual ad campaign with a search ad network and selects a set of target keywords for displaying her ads. To serve these ads, the publisher may operate search-related services such as search engines and toolbars, or use in-text advertising techniques (i.e. when the mouse hovers on a target word, the ad is displayed) to identify the right context for advertising: for example, it shows ads associated with the target words that are included in the search terms entered into search engines or the toolbars. Parking services are one of such publishers who provide textual ads relevant to the names of parked domains.

Search advertising is made possible through pay-per-click (PPC) XML feeds as illustrated in Figure 1. A publisher submits a search query for certain keywords (relevant to the domain names, in the case of parked domains) and receives relevant ads in the XML format, which also include the bidding price per ad from the advertisers. The publisher in turn picks up a set of ads to display. Once a user opts to click on an ad, the click traffic is bounced through a number of hosts such as click servers before reaching the advertiser's web page. This click is paid for by the advertiser and the revenue generated in this way is shared between the publisher and the ad networks.

Popular search ad networks such as `Google AdWords` and `BingAds` are considered to be top-tier (premium), while other less reputable ones are 2nd-tier or lower. Compared with other ad networks, top-tier networks offer a higher rate per click (CPC) to the publisher and a better click fraud detection to the advertiser.

• *Direct navigation traffic (PPR)*. Direct navigation traffic (aka., *type-in* traffic) is generated when the web user enters a domain name as a query and expects to be redirected to a related domain. For example, one may type in "findcheaphotels.com" in the address bar and land at `mytravelguide.com`. This is caused by a direct-navigation-traffic purchase in which the owner of `mytravelguide.com` purchases through a direct navigation system the traffic related to keywords "travel" and/or "hotels", for example. Parked domains can serve such a direct navigation system by redirecting type-in traffic to them who ultimately redirect it to traffic buyers like `mytravelguide.com`. This monetization option is called Pay-Per-Redirect (PPR) or zeroclick.

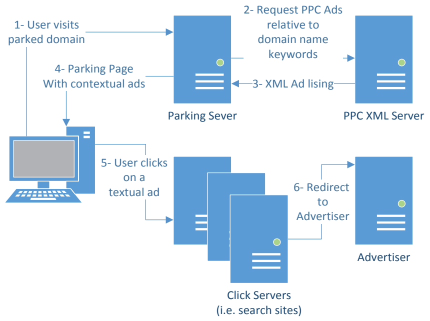

1- User visits parked domain
2- Request PPC Ads relative to domain name keywords
4- Parking Page With contextual ads
3- XML Ad lising
Parking Sever
PPC XML Server
5- User clicks on a textual ad
6- Redirect to Advertiser
Advertiser
Click Servers (i.e. search sites)

Figure 1: Legitimate PPC XML feed operation.

2.2 Ecosystem and Illicit Monetization

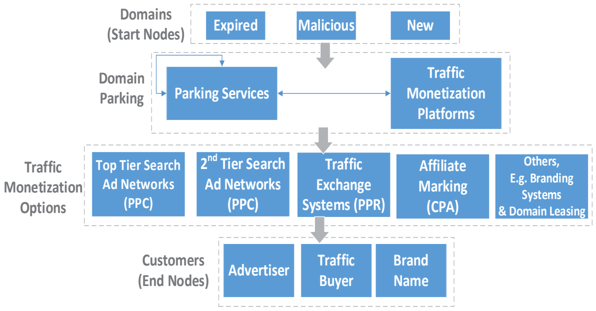

Domains (Start Nodes)
Expired
Malicious
New
Domain Parking
Parking Services
Traffic Monetization Platforms
Traffic Monetization Options
Top Tier Search Ad Networks (PPC)
2nd Tier Search Ad Networks (PPC)
Traffic Exchange Systems (PPR)
Affiliate Marking (CPA)
Others, E.g. Branding Systems & Domain Leasing
Customers (End Nodes)
Advertiser
Traffic Buyer
Brand Name

Figure 2: Domain Parking Ecosystem

Domain parking services have a significant hold in the domain industry. Table 1 shows some of the most popular parking services investigated in our study. By looking at the ranks of the domain names associated with the companies running these parking services, we find that many of them are ranked among the top 100k by Alexa [2]. Also shown in the table is the total number of parked domains observed on Feb. 25^{th} according to *DailyChanges* [34]. Note that the list maintained by DailyChanges is not comprehensive. Nevertheless, it is quite clear from Table 1 that parking services indeed cover a large number of domains. Here we describe our understanding of its ecosystem, based on our analysis of a large amount of data collected from the domains under leading parking services. Such an understanding also leads to the suspicion of illicit activities within this ecosystem which motivates this research.

Infrastructure. Figure 2 illustrates the infrastructure of the domain parking ecosystem, as revealed by the redirection chains observed during our crawling of parked domains. Those domains are the *start nodes* for the

		Alexa [2]	DailyChanges [34]	Our data set	
#	Parking Service	Global	# Parked	# Parked	# Monetization
		Rank	Domains	Domains	Chains
1	Above	39,901	512,206	17,160	348,534
2	PS1	19,973	1,101,050	11,850	94,794
3	PS2	21,446	1,615,644	9,972	82,258
4	PS3	78,275	261,357	6,447	141,174
5	PS4	59,248	1,275	3,606	174,410
6	PS5	11,357	136,243	2,782	108,956
7	PS6	62,369	330,032	1,020	135,946
8	Rook Media	2,827,350	132,455	562	7,505
9	Fabulous	32,850	406,872	315	53,851
10	InternetTraffic	4,072,159	1,290,732	151	7,764

Table 1: Top 10 Parking services in our data set. The number of monetization chains is not the same as the number of parked domains since each parked domain was crawled multiple times during our collection period. Note that parking services found in this paper to carryout illicit activities are anonymized as "PS#".

whole ecosystem. They include expired domains with back links, blacklisted domains (e.g., exploit servers or TDSes [17]) seized or taken down then repurchased by domain owners or newly acquired domains. As illustrated in the figure, parked domains forward their traffic to parking services, which in turn select the most profitable monetization option in real time, based upon a set of characteristics of the traffic such as its geolocation, browser type and domain keywords. Occasionally, a parking service chooses to forward the traffic to another parking service when the latter offers a higher return on a specific traffic instance. In addition, a parking service may collaborate with *traffic monetization platforms* (e.g. `Skenzo.com`), which monetize different types of traffic such as parking traffic, error traffic (i.e. `404` not found pages) and non-existent domains. Here, we refer to this type of partnership as parking syndication.

The end targets of any traffic monetization option can be either an advertiser, a traffic buyer or a brand name, which are the *end nodes* of the infrastructure.

Potential illicit monetization activities. In our study, we discovered, during our crawling of parked domains, some suspicious activities that call into question the legitimacy of some monetization operations. Specifically, we found that some URLs on the redirection chains initiated by our crawler contain patterns related to the delivery of clicks, for example, "`http://fastonlinefinder.com/ads-clicktrack/click/newjump1.do?`". The problem is that our crawler never clicked on any URLs. It just visited a parked domain and followed its (automatic) redirection chain (see Section 2.3). Also, we observed a lot of "shady" search websites (e.g., `fastonlinefinder.com`), which look like search engines but return low-quality ad results. Those search sites are also observed in prior research [18, 4] and have been presumed to be related to malicious activities like click fraud and malware delivery.

However, confirming the presence of illicit activities in domain monetization is challenging. Take click fraud as an example. We need to determine whether the crawler traffic has indeed been monetized as clicks, which can only be confirmed at the advertiser end. Further complicating this attempt is the observation that some parking services try to make the click delivery look like zeroclick monetization (PPR) by bouncing the traffic through entities with indications of "zeroclick" in the URL: for example, a visit to a parked domain is initially redirected to `http://bodisparking.com/tracking?method=zeroclickrequest` before moving to the click URL. Also, malware scanning cannot find any malicious payloads from the traffic collected from parked domains. Most importantly, given that the traffic for domain monetization goes down a complicated redirection chain, including ad networks and parking-service syndication, it becomes highly nontrivial to identify the party responsible for a malicious activity, even when its presence has been confirmed.

2.3 Overview of Our Study

Here we describe at a high level what we did in our research to understand the suspicious activities that happen within this domain parking ecosystem.

Data collection. As discussed above, the data used in our study was collected from crawling parked domains. For this purpose, we implemented a dynamic crawler as a Firefox extension and deployed it to 29 Virtual Machines (VMs). The crawler is designed to simulate a user's visit to a URL through a browser by rendering its content and running scripts. All such content and HTTP traffic (such as redirections) generated are collected and dumped into a database. In this way, the crawler is able to gather the information produced by execution of dynamic content.

Those crawlers worked on a list of parked domains, which was updated every 3 days during the past 5.5 months (August 1st, 2013 to January 20th, 2014). Those domains were discovered by reverse-lookup for the NS records of known parking services (a list built manually) using the *PassiveDNS* set (DNS record collection) provided by the Security Information Exchange [30]. In order to investigate the monetization activities through those domains, we constructed a *monetization chain* for each URL visit. A monetization chain is a sequence of URL redirections (e.g. HTTP 302, iFrame tags, etc.) observed during a visit to a parked domain, including ad networks and traffic systems related to monetizing the visit.

During each visit, each crawler randomly picked one of 48 user agent strings covering popular browsers, operating systems and mobile devices. Overall, we made about 24M visits to over 100K parked domains. From all those visits, we identified 1.2M (5%) monetization chains including redirections (not direct display of ads).

The leading parking services involved in those chains are presented in Table 1.

Infiltration and expansion. To identify illicit activities involved in the monetization of parked domains and understand the scope and magnitude of the problem, we performed an infiltration study on the domain parking ecosystem to gain an "insider" view about how those parking systems operate. This is critical for overcoming the barriers mentioned in Section 2.2. More specifically, we ran our crawlers to collect data from parked domains and also parked domains under our control with major parking services. Additionally, we launched a few ad campaigns and also purchased traffic associated with some keywords. By carefully selecting the parameters at our discretion, we were able to "connect the dots", linking the start nodes (domains) or traffic sources (crawlers) under our control to our end nodes (ad or traffic purchase campaigns) on monetization chains. Those chains (called *seeds*), together with the accounting information we received from related parking services and ad networks, reveal the whole monetization process with regard to our inputs. This enables us to identify the presence of click fraud, traffic stealing (failing to report monetized traffic) and traffic spam (low-quality traffic). We elaborate this research in Section 3.

To understand the impacts of those fraudulent activities, we extracted from the seed monetization chains a set of fingerprints, or *stamps*, to identify the monetization method used. Once a monetization chain is identified as either PPC or PPR, we infer the presence of illicit activities. Our research shows that our approach accurately identifies illicit monetizations through known ad networks and traffic systems. Most importantly here, this approach helps us *expand* those seeds to a large number of monetization chains collected by our crawlers. Over those chains, we performed a measurement study, which shows the pervasiveness of the problems, their unique features and the profits the parking services get from the illicit activities. The study and its outcomes is reported in Section 4 and Section 5.

Adversary model. In our research, we consider that the parking service is untrustworthy, capable of manipulating the input traffic it receives and its accounting data to maximize its profits at other parties' cost. It also cloaks frequently to avoid being detected by third parties. On the other hand, the service cannot change its interfaces with legitimate ad networks: it needs to make the right calls to deliver its traffic to the networks. In the meantime, some less reputable ad networks (2nd-tier or lower) may not be trustworthy either, which adds complexity to assigning blame to different parties involved in a known fraudulent activity.

In practice, parking services are actually legitimate companies. What we found is that they apparently behave legitimately most of time but are indeed involved in illicit operations occasionally. This adversary is actually very unique, since they blur the lines between fraudulent and legitimate transactions and conduct operations with highly questionable practices.

3 Dark Side of Domain Parking

In this section, we report on our infiltration of the domain parking ecosystem. As discussed before, what we did is to control traffic sources (crawlers), some start nodes (parked domains) and some end nodes (ad campaigns & traffic purchases) of the ecosystem, to get end-to-end monetization chains going through them, as depicted in Figure 3. The figure shows that the chains are as follows: from our parked domains to our end nodes, that is, advertisers or traffic buyers (in black); from other parked domains to our end nodes (in red); from our parked domains to other end nodes (in green) and from our crawlers but not through our domains or end nodes (in blue). Among those chains, the black and red chains connect our traffic source, crawler, to our end nodes through parked domains, which are used as the ground truth for validating our findings (Section 3.3) and the seeds for detecting illicit activities on other chains (Section 4).

Below we describe how we infiltrated the ad networks and direct traffic navigation systems on the end-node side and the parking services on the start-node side.

3.1 Infiltrating End Nodes

Here we walk through our infiltration of the end nodes of the ecosystem, which includes a few steps: we need to identify the right targets (ad networks or traffic systems), register with them, launch ad campaigns and set the right parameters to maximize the chances of receiving our own crawling traffic.

Target identification and registration. To identify the most popular targets, we inspected a sample dataset, including monetization chains collected during the first two weeks of August 2013, to collect a set of the most prevalent top and 2nd-tier ad networks and direct navigation systems. This turned out to be rather straightforward for some targets (e.g., the `Looksmart` ad network with a domain name `looksmart.com`), but not so for others. For example, for some ad networks (e.g. `Advertise`), only the domains of the "shady" search websites they utilized showed up on their click URLs; the "masters" of those search domains were not revealed from their `whois` records, which indicated either an anonymous registration or missing organization names. To uncover those ad networks, what we did include using a domain's Autonomous System Names (ASN) or other domains sharing its IP addresses to determine its affiliation, as well as comparing an ad network's contact

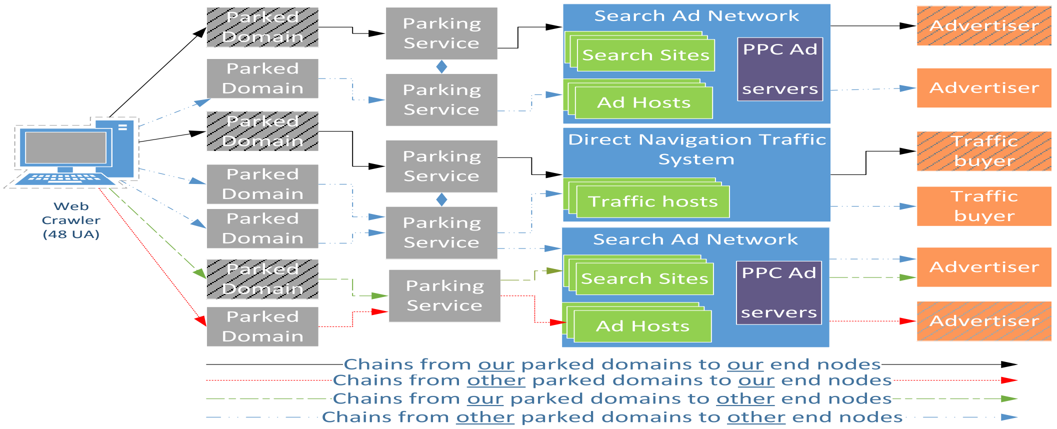


Figure 3: Types of monetization chains captured in our data set where the dashed boxes represent entities under our control.

information with that on the domain's whois record.

Once a set of targets (ad networks and traffic systems) were identified, we attempted to register with them as an advertiser or a traffic buyer. This happened with fake identities whenever possible, for the purpose of concealing our true identity to avoid cloaking activities, but we had to use our real information for some of them, which asked for IDs such as driver's license and a credit card. All together, we successfully affiliated ourselves with 15 out of 25 top ad networks and traffic systems identified. For those we failed to do so, the main cause was that they only accepted large-budget customers.

Campaign creation and parameter tuning. We set up a search-like website (Figure 7 in Appendix) and hosted it on three domains, one for traffic purchasing and the other two for advertising. Using those domains, we created both advertising and traffic purchasing campaigns. Specifically, for each of them; we selected 10 keywords related to *our own parked domains*, and also constructed our target URLs, given to the ad networks or traffic systems, to communicate a set of data (e.g., ad network names, publisher IDs, keyword used, etc.) to be used to identify monetization chains that end at our websites through our crawlers. Table 2 summarizes our infiltration meta-data.

For each campaign, we carefully adjusted its parameters to maximize the chances of getting traffic from our own crawlers, which provided us with the end-to-end monetization chains we were looking for. The strategies for such parameter tuning varied across different ad networks and traffic systems. Specifically, some of them offered geo-targeting, which we leveraged to aim at the city our crawlers were located. When this was not available, we tried to take advantage of other features such as browser type and timing when offered. Particularly, for the browser feature, our campaign opted to target the least common browser type, which our crawlers also used for their user agents. Additionally, the timings of some campaigns were tuned in a way that they only ran when traffic from sources other than our crawlers was minimal, for example, from 12AM-6AM and 10PM-12AM. As an example, the direct navigation traffic system DNTX provided only country-level geo-location and the platforms (iOS, Android, BlackBerry and others) under its mobile/tablet category. In this case, we targeted our campaigns at the country our crawlers were running from and the Blackberry platforms.

Overall, we ran advertising and traffic campaigns through all 15 ad and traffic networks starting from Nov. 22nd last year, which cost us $2,260 in total. The logistics of our end node infiltration campaigns are summarized in Table 3. Among those campaigns, the two with Admanage lasted for only two days before we were locked out from our account for almost a month without giving any explanations. Also, our campaign with Approved Search did not receive any traffic that fit our targeting criteria.

3.2 Infiltrating Start Nodes

To infiltrate start nodes, we went through top 10 most popular parking services identified from our sample dataset (Section 3.1), opened accounts with them and carefully chose our monetization options so as to maximize the chance of observing illicit activities.

Parking service registration. Among all the services we tried, Domain Power asked for real identity information, which we skipped, and trafficZ turned down our application, citing the small volume of traffic our domains received. Other parking services performed some type of authentication, such as sending a PIN number to a valid phone number, and verifying the consistency between the owner information on our domains' whois records and that on our application. We passed those checks and used fake identities to register with 7 parking

Parked domains	FREE-JOBS.INFO, JOBS-BOARD.INFO, REAL-JOBS.INFO, NEWS-CHANNEL.BIZ, NEWS-FEED.BIZ, FAMILY-VACATION.ORG DREAM-VACATION.ORG, COUPONS-FREE.INFO, LOCAL-COUPONS.INFO, SUPERCOUPONS.INFO, CLOTHES-SHOP.INFO DESIGNER-CLOTHES.INFO, TEAMXYZ.INFO, XYZAGENT.INFO, LOWCOST-FLOWERS.COM, EDITING-SOFTWARE.ORG EDUCATION-RESOURCES.ORG, EDUCATION-GUIDE.ORG, MARKETING-EDUCATION.ORG, CITY-CARS.NET, MUSIC-LIVE.ORG SOFWTARE.COM, NEWS-NETWORK.BIZ
Target Keywords	Jobs, Cars, News, Vacation, Coupons, Clothes, Software, Education, Music, Flowers
Target URL example	`http://anwers.net/search.php?s=advertise&c=camp7&type=kw&kw=jobs&aff=63567&geo=us_in_bloomington`

Table 2: Infiltration meta-data. The URL example indicates the ad network is "Advertise", click keyword is "Jobs" and the publisher ID is "63567".

Network	Campaign Type	Total Hits	AVG CPC/CPR	Budget ($)	# Days	Targeting
adMarketplace	PPC	143	0.2	100.13	18	City & Time
Advertise	PPC	18	0.067	125.17	17	City, Device & Time
Google AdWords	PPC	1	1.86	222.71	32	City & Device
Affinity+	PPC	371	0.3	250.36	17	City & Time
Approved Search	PPC	0	0.05	10.5	2	Country & Time
Bidvertiser	PPC	0	0.1	160.13	28	Country
Bing Ads+	PPC	6	0.97	33.06	30	City, Device, Time
Ezanga	PPC	35	0.21	47.92	33	City & Time
Looksmart	PPC	131	0.19	91.1	30	City & Time
Avenue5	PPC	0	0.05	44.25	37	Country & Time
Admanage	PPC	0	0.4	43.6	2	Country
7search+	PPC	3	0.1	297	26	Country, Device
	PPR	146	0.1	203	29	& Time
DNTX	PPC	1	0.058	155.89	13	Country
	PPR	29	0.056	203.5	27	& Device
Adspark	PPR	106	0.2	31.8	5	City, Device & Time
Trellian	PPR	25	0.201	250	43	Country

Table 3: End node Infiltration Logistics. + Networks that required registration with a real identity. "Total hits" represent click/traffic hits received at our web server and initiated by our crawler. Note that `7Search` offered both PPC and PPR campaign types and as such, we created two campaigns with it, one for each type.

services as illustrated in Table 4. Most of these services are indeed popular as shown earlier in Table 1.

#	Parking Service	# Parked Domains
1	PS5	9
2	PS2	2
3	PS6	6
4	PS1	4
5	PS4	3
6	Rook Media	2
7	PS7	2

Table 4: Parking Infiltration Logistics. Note that the total number of domains does not add up to 23 which is due to moving some domains between parking services. Note that PS7 is not shown in Table 1 as it is not in the top 10 parking services.

Domain monetization. We purchased 23 domains under a number of top level domains and parked them with the 7 services. The names of those domains were carefully chosen to match the keywords we targeted in our campaigns (see Table 2). We also set their NS records to point to those of their corresponding parking services, with some exceptions discussed later.

Our preliminary analysis on data crawled from parking services showed that suspicious activities were only observed on redirection chains. Therefore, we tried to avoid situations like PPC listings, where a set of PPC ads are displayed on a parking landing page. Instead, we chose not to have such a page (displaying signs like "for sale"), so the parking services can monetize the traffic our domains received through redirections. Actually, one parking service, `Bodis`, allowed us to explicitly set this monetization option by redirecting the traffic instead of setting the NS record to `Bodis`. For example, we parked the domain `news-network.biz` with `Bodis` by using the GoDaddy forwarding service to send our domain traffic to `http://bodisparking.com/news-network.biz`.

3.3 Findings

Through infiltrating the start and end nodes of the ecosystem and crawling domains hosted by popular parking services, we were able to collect 1,015 monetization chains that link our crawlers to our end nodes (our advertising or traffic purchase websites), sometimes through our start nodes (parked domains). Using those chains as the ground truth, we confirmed the presence of illicit activities during parked domain monetization, including click fraud, traffic spam and traffic stealing, as elaborated below.

Click fraud. Out of all the ad clicks delivered to our advertising websites through parking services, 709 were found to come from our own crawlers. They are clearly fraudulent since our crawlers were designed not to click on any ad (Section 2.3). Table 3 details the number of clicks received from our crawler through each ad network.

Traffic spam. The 4 traffic purchasing campaigns we launched received 306 traffic hits from our crawler through domains parked with parking services. Upon examining the parked domains that served as the start nodes on those monetization chains, we found that 83 of them were totally unrelated to the keywords we purchased from the direct traffic systems. Table 5 provides examples of spam and good-quality traffic.

Keyword	Spam	Not Spam
Music	19jj.com, ib2c.com.cn	thepiatebay.org, itunesstore.de
Software	almacenyhostpublico.com	linuxfab.cx,iphoneos3.com
Coupon	seattleseoforum.com	coupons-free.info
"Others"	brf.no, betovilla.com,gddfg.com 70263.com, facebooki.pl	education-guide.org dolla.com

Table 5: "Spam" and "Not Spam" examples of parked domains appearing in traffic purchase monetization chains leading to our traffic buyer website. "Others" represents instances where the purchased keyword was not propagated through to our traffic buyer URL but it is still evident from the spam examples that they are not related to any of the 10 keywords we purchased.

Traffic stealing. Occasionally, parking services were found to be dishonest with domain owners, failing to inform them for part of the revenue they were supposed to share with the owners. Specifically, we cross-examined the revenues of the domains under our control and the billing reports for the ad/traffic campaigns we launched. This revealed that some monetization chains going through our parked domains were not reported to us (domain owners) but charged to our campaign accounts.

For example, we confirmed the existence of traffic stealing from monetization chains captured by our crawler connecting three of our domains (parked with PS5) and our PPR campaign with 7Search. This was achieved through comparing the billing reports provided by 7Search, the parked domains' revenue reports provided by the parking service and the related monetization chains with the right combination of time stamp, source IP address, referral domain and keyword. It turns out that we, as a campaign owner, were billed for 23 traffic hits by 7Search (see Figure 6(b) in Appendix) but nothing was reported by the parking service (see Figure 6(a) in Appendix) in December 2013. We show the breakdown of the crawlers' traffic in Table 6. Clearly, the parking services kept the rightful share away from us as the owners of the parked domains. Note that not all requests from our crawler were billed by 7Search because they limit the traffic hits by one IP address and a valid visiting period (our campaign was set to run between 12AM-6AM and 10PM-11:59PM). Additionally, we found other monetization chains, captured by our crawler and monetized by the same parking service through other ad networks such as Advertise that have not been reported on our parked domains' revenue reports.

	Traffic Reported by		
Parked Domain	Crawler	Parking Service	Billed by 7search
Coupons-free.info	24	0	16
Real-jobs.info	23	0	5
News-feed.info	21	0	2

Table 6: Traffic stealing through 3 of our parked domains in the month of December, 2013.

4 Fingerprinting Monetization Chains

Through the infiltration study, we confirmed the presence of illicit activities in the monetization of parked domains. What is less clear, however, is the pervasiveness and impact of those activities. Understanding of this issue cannot rely on the 1,015 seed chains (reported in Table 3) whose traffic sources and end nodes were under our control. We need to identify the illicit operations occurring on other monetization chains, particularly those blue and green ones in Figure 3, which do not connect to our ad/traffic campaigns. To this end, we developed a technique that fingerprints the monetization options observed on our seed chains. These fingerprints, which we call *stamps*, were used to "expand" the seed set, capturing the illicit activities on other monetization chains collected by our crawlers over months.

4.1 Methodology

The idea. As discussed above, the problem of detecting illegitimate operations, which we did not have a direct observation of, comes down to identifying the monetization options they involve. More specifically, as soon as we know exactly how a parking service monetizes a visit from our crawler, we can immediately find out whether a fraudulent activity occurred: clearly, the PPC option is a fraudulent click, as our crawler never clicked; when it comes to PPR, we check the consistency between the keywords expected by the end nodes and the names of the parked domains the traffic went through (see Section 5 for details). Therefore, the question here becomes how to determine which options have been used in a given monetization chain.

Actually, even though those options might not be evident on a monetization chain, we know that it must go through a corresponding monetization party (ad networks, traffic systems, etc.) before the traffic gets to the end node. This needs to happen for accounting purposes: for example, if a click has not been sent to a PPC ad network, the ad network never knows about it and therefore will not be able to pay its publisher or bill its advertiser. Also, the last few URLs leading to the end node are clearly related to the monetization party. As an example, let us look at a monetization chain captured by our crawler in Table 7, which was initiated by a visit to a domain parked with PS1 and ended at our advertiser site. Looking backward from our URL along the chain, we can see two URLs from fastonlinefinder.com, a search website. The site turns out to be affiliated with the Advertise ad network. Interestingly, once we compare this path with other chains also through the same ad network, it becomes quite clear that they carry a unique pattern: first, right before the end node, the last two URLs are always similar; second, for these two URLs, even though they vary across different chains in terms of search websites, the remaining part mostly stays constant. This observation was further verified by the click URLs for sponsored ads from the same ad network, which we obtained by registering with Advertise as a publisher (see Appendix for details on search sites and sponsored click URLs).

The above example shows that we can leverage the ordered sequence of URL patterns to determine the presence of a monetization option. Such a sequence is a "stamp" we utilize to expand our seed set to find other illicit monetization chains within the dataset collected by our crawlers. Following we describe the methodology

#	URL	Description
1,2	`http://bastak-taraneh.com/`	Parked domain
3	`http://otnnetwork.net/?epl=...`	Parking service anchor
4,5	`http://67.201.62.155/index2.html?q=...&des=`	AdLux
6	`http://21735.1b2a3r4w5dgp6v.filter.nf.adlux.com/ncp/checkbrowser?key...`	ad network (syndicate)
7	**`http://fastonlinefinder.com/ads-clicktrack/click/newjump1.do?...&terms=ticketsoftware...`**	Search site for
8	**`http://fastonlinefinder.com/ads-clicktrack/click/newjump2.do?terms=ticketsoftware&...`**	Advertise ad network
9	`http://anwers.net/search.php?s=advertise...&kw=software...`	Our Advertiser

Table 7: End-to-End monetization chain example of a visit to a parked domain leading to our advertiser page. For clarity sake, we omit parts of the URLs.

for clustering and generalizing URLs from a monetization party, and extracting the stamps from the URL patterns.

URL-IP Cluster (UIC) generation. A specific URL of a monetization party (ad network, traffic systems, etc.) can be too specific for fingerprinting its monetization activity. An ad network can have many affiliated websites and each site may have multiple domains and IP addresses. To utilize such a URL for generating a stamp, we first need to generalize it across those domains, addresses and potential variations in its file path and other parameters. To this end, we clustered related URLs into *URL-IP clusters* (UIC). A UIC includes a set of IP addresses for related hosts and the invariant part of the URL (without the host name) across all members in the cluster. The former describes the ownership of this set of URLs and the latter represents their common functionality, which together fingerprints a monetization option with regard to an ad network or a traffic system.

To cluster a group of URLs into UICs, we first extracted the host part of a URL, replacing it with all IPs of the domain, and then broke the remaining part of the URL into tokens. A token is either the full path of the URL including file name (which is typically very short for a monetization URL) or an argument. The value of the argument was removed, as it can be too specific (e.g. keyword and publisher ID). Over those IP-token sets, we ran a clustering algorithm based upon Jaccard indices for both IPs and tokens, as follows:

1. Each URL (including an IP set and a token set) is first assigned to a unique UIC.

2. Two UICs are merged together when *both* their IP sets and token sets are close enough (Jaccard indices above their corresponding thresholds).

3. Repeat step 2 until no more UICs can be merged.

A pair of thresholds are used here to determine the similarity of two IP sets (T_{ip}) and two token sets (T_{tok}) respectively. In our research, we set T_{ip} to 0.1 and T_{tok} to 0.5, and ran the algorithm to cluster all the URLs on the 1.2M monetization chains collected by our crawlers. By replacing individual URLs with their corresponding UICs, we obtained 429K unique generalized chains, which were further used to detect illicit activities.

Stamp extraction. The stamps for different monetization options were extracted from seed monetization chains, after generalizing them using the aforementioned UICs. Specifically, we utilized 715 UIC chains (generalized from the 1,015 chains reported in Table 3) to fingerprint 11 ad networks and traffic systems, with one stamp created for all the campaigns associated with a given ad network or traffic system. For this purpose, we applied a 2-fold cross-validation approach to generate stamps and assess their effectiveness. Specifically, we randomly split the UIC chains for each campaign into two equal sized sets, one for stamp extraction (training) and the other for stamp evaluation (test). Over the training set, we determined a stamp by traversing each UIC chain backwards and selecting the sequence of UICs shared by *all* the chains involving a certain ad network or traffic system. Typically the longest sequence identified in this way became the stamp for all the chains going through its related monetization organization. However, for a campaign with a small number of UIC chains (e.g. `BingAds`), we only utilized the last common UIC (right before our advertiser's URL) across all the chains as the stamp.

All together, our approach generated stamps for 11 ad networks and traffic systems. Ad network stamps contained on average two UICs while traffic system stamps were mostly one UIC in length.

4.2 Evaluation

False negative. Using all the 11 stamps generated from the training set, we analyzed all the monetization chains within the test set. For each campaign, its stamp was found to match all of its monetization chains in the test set and thus no false negative was observed.

False positives. We further evaluated the false positive rate that could be introduced by the stamps on a dataset containing 768M redirection chains collected by crawling the top 1M Alexa websites [2] Jan 21-31, 2014. The purpose here is to understand whether a redirection chain not involving clicks or traffic selling can be misidentified as a related monetization chain and whether the monetization chains of one ad network or traffic system can be classified as those of another party. By applying our stamps on the dataset, we flagged 12 chains as matches to click stamps and another 19 chains as matches to traffic stamps. Upon manually analyzing the 12 click chains, we found that all of them were actually fraudulent clicks generated by parking services (10 chains) and a traffic/ad network named CPX24 (2 chains). CPX24 in our case generated clicks on its own hosted ads when traffic from publishers flowed in. For the 19 traffic chains, they were indeed PPR monetization chains. Further, by manually searching for a set of ad network specific domain names such as `affinity.com` within the Alexa dataset, we discovered a number of redirection chains going through the same ad networks fingerprinted by our stamps but *not* involving any clicks (e.g. ad display and conversion tracking URLs). None of them were misidentified by our stamps as click-based monetization chains.

Discussion. Our evaluation shows that the stamps generated over UICs accurately capture all monetization chains associated with a specific ad network or traffic system. However, those stamps are designed for analyzing the traffic through individual organizations, which is enough for our purpose of understanding the scope and magnitude of fraudulent activities, not for detecting those operations on any monetization chains, particularly those belonging to other monetization parties. Also note that we cannot use existing ad-blocking lists such as EasyList [26] to serve our goal, due to its limitations: first, EasyList does not distinguish between a click and an impression (ad display); second, search websites used by ad networks to deliver clicks are not covered by the block list; finally, the list fails to include the hosts and URLs of traffic monetization systems.

5 Measurements

In this section, we report our measurement study on illicit parked domain monetization. This study is based upon a dataset of 1.2M monetization chains collected in a 5.5-month span. Such data were first labeled using the "stamps" generated (Section 4.1) from the seed data to identify the monetization options associated with individual chains, and then analyzed to understand the pervasiveness of illicit monetization practices and its financial impact. Here we elaborate on the outcomes of this study.

5.1 Dataset Labeling

Expansion. To perform the measurement study, we labeled the 1.2M monetization chains collected from crawling parked domains by "expanding" the 1,015 seed chains (the ground truth) to this much larger dataset. Specifically, we generated UICs over those 1.2M chains, generalized the seed chains using those UICs, and then extracted click and traffic stamps from the seeds as described in Section 4.1. Matching those stamps to the generalized UIC chains in the larger dataset (429K UIC chains), we were able to label 120,290 (28.03%) UIC chains corresponding to 212,359 (17.1%) URL monetization chains. The labeled set includes two monetization options, PPC (45.7%) and PPR (56.3%) where 2% of them include both PPR and PPC monetizations on the same chain as explained later.

Unknown set. Although many chains were successfully labeled, there are almost 308K UIC chains in the dataset not carrying any stamps, which were marked as "unknown". Looking into this unknown set through random sampling, we found that it exhibited consistent patterns related to click delivery and traffic selling which can be added to the labeled set if we had verified seed chains. For example, we found many other ad-nets such as `Adknowledge` and `Bidvertiser` (2.9%), and other traffic monetization systems such as `Adrenalads` and `ZeroRedirect` (19.5%). Particularly, `Sendori`, a traffic platform, is widely present, covering 5.4% of the chains in the dataset. 2.6% of those chains actually led to domain name marketplaces such as `SnapNames`, and a large portion (over 19.7%) of them stop at some parking services and traffic monetization platforms (e.g. `Skenzo`) with error messages indicating cloaking behavior.

5.2 Monetization Decisions

Over the labeled dataset, we analyzed the parties responsible for such decisions and the way those decisions were made, based on a categorization of the parked domains involved.

Monetization decision maker. Finding the party that chooses a monetization option is important, as it tells us who is the ultimate culprit for an illicit activity. However, this is challenging, due to the syndication of multiple monetization parties, among parking services, adnets and traffic systems. Within our dataset, we found that these types of syndications are pervasive (49.5%). As an example, `AdLux` in Table 7 is actually a syndicate of `Advertise`, displaying its ads and sharing its click revenue. In the presence of a syndication, a starting node's parking service may not be responsible for the follow-up illicit monetization, which could actually be performed by one of its syndicates. To this end, we identify the parking service of the starting node to be the responsible party of a monetization chain only when the click or traffic stamp appears right after the starting node (i.e. parked domain). When there are other entities between the parked domain and the stamp, we use a *parking-service anchor* as described below.

Typically, a parking service funnels the traffic from its parked domains to a "controller" domain, which we call a parking service (PS) anchor, for choosing a monetization option. Our idea here is to locate the PS anchor right before a click or traffic stamp. When this happens, the owner of the anchor is clearly responsible for the monetization decision. To this end, we picked out the most prevalent second UICs down individual monetization chains (which is expected to cover over 50% of all the chains associated with a specific parking service), and identified its ownership using its `whois` records and Name Server. Such a UIC is considered to be an anchor for the parking service.

In our research, we identified anchors for 4 of the most prevalent parking services in our set. Some parking services such as `PS6` and `PS4` launder all the traffic through direct navigation traffic systems (`DNTX` and `ZeroRedirect` respectively) which are owned by the parent companies of the two parking services. Since those traffic systems are used by other clients, we did not consider them to be the anchors of the parking service. Using the heuristic described earlier (i.e. direct link from a parked domain to a stamp) and the list of anchors, we assigned each monetization chain to the parking service responsible for the selection of its monetization option, as illustrated in Table 8. Here the "unknown" category includes the chains we could not determine the parties responsible for their monetizations, due to the disconnection between the parked domain or PS anchor and the click or traffic stamp, with unknown UICs standing in-between. For example, the chain in Table 7 was marked as "unknown", as the known anchor `http://otnnetwork.net/?epl=` is separated from the ad stamp there.

Impacts of domain categorization. As discussed before, our research focuses on the redirection chains generated by 1.2M out of 24M visits to parked domains. The rest of those visits only resulted in a simple display of PPC ads, which were less likely to be used for illicit monetization. The fact that those redirection chains were so rare to see here can be attributed to IP cloaking. In the meantime, we believe that this is also caused by the way that traffic from different domains is monetized. Specifically, a parking service like `Bodis` often classifies domains into "primary" or "secondary". Primary domains are those accepted by top-tier search networks (e.g. `Google AdWords`) to display their ads while secondary ones are less trusted, including those serving malicious content before taken down and the ones related to typos of trade or brand names. The secondary domains here are much more likely to lead to redirection chains, as discovered in our research (illustrated by Figure 4). Among all domains visited by our crawlers, we found (using [6]) that only 2.9% of them were considered to be secondary, which naturally limits the number of the redirection chains we could observe.

5.3 Illicit Monetization

In this section, we report our findings about the prevalence of illicit monetization practices, particularly click fraud, traffic spam and the malware distribution discovered during our analysis of the labeled dataset. Note that we did not measure traffic stealing, as this activity could only be observed on the monetization chains whose start and end nodes were under our control.

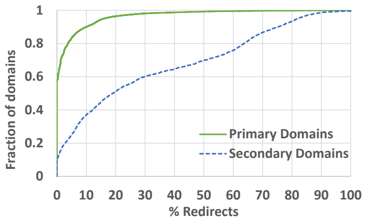


Figure 4: Tendency towards redirect monetization of "secondary" domains. "% Redirects" represents the percentage of redirection chains observed per parked domain.

		PS5	PS6	PS1	PS2	Others	Unknown
PPC Ad-nets	adMarketplace	1		9,526	1670	4	35,704
	Advertise	10,277		2,329		730	3,576
	Affinity						9,904
	Google AdWords						18
	Bing Ads						18,186
	Looksmart					385	143
	Ezanga					3	422
PPR Traffic Systems	7Search -Spam	7,766				666	1,484
	7Search - Malware	43					
	DNTX - Spam	758	21,959	189	2	25,776	2,305
	DNTX - Malware	45	3,217	9	1	2,924	113
	Trellian - Spam	1		9	1	11,781	
	Trellian - Malware					1	
	AdsPark -Spam		1,755				2,183
	AdsPark - Malware						1
Totals		18,803 (11.02%)	22,425 (13.15%)	13,808 (8.1%)	1,673 (1%)	39,872 (23.4%)	73,949 (43.4%)

Table 8: Illicit activities observed by parking services in the labeled set. "Others" refers to some of the parking services shown in Table 1 not necessarily anonymized.

Traffic spam. Using the traffic stamps, we discovered 119K (56.3%) traffic monetization chains. To identify the presence of traffic spam on each of those chains, we compared the keywords associated with its start node domain with those of its end node (assuming that end nodes purchased keywords related to the contents of their domains). This works as follows:

- *Keyword generation*. To generate keywords for both the start and end nodes, we used a keyword suggestion tool by `BingAds` [1], a tool widely used by advertisers to select keywords for ad targeting. This tool automatically created a list of keywords (including typos) for each domain (start and end nodes).

- *Keyword filtering*. At this step, we cleaned the list of keywords, discarding common ones ("www", "com", etc.). Specifically, we calculated the normalized entropy of each word as the prior work did [31] and then removed the 50 words with the lowest entropy (i.e. highly popular). Also dropped from the keyword list were determiners, pronouns, interjection and "wh"-words ("what", "where", etc.), which are unlikely to be related to specific domain content. For this purpose, we filtered out the keywords using *Stanford CoreNLP* [35], a natural language processing tool for part-of-speech tagging and stemming.

• *Keyword matching*. Comparing the keywords of the start node (i.e., parked domain) with those of the end node (traffic purchase website) on a monetization chain, we considered the chain to be traffic spam if individual keywords of its parked domain did not match *any* words associated with its end node. In the case that one of these two domains do not have any keywords, we attempted to match the other domain's keywords to its domain name. If this attempt succeeds, the chain would not be considered as spam, otherwise; it would.

As a result, we found that 70.7% of all PPR monetization chains are traffic spam as illustrated in Table 8 and attributed to each parking service and traffic system. Table 9 provides some traffic spam examples received by popular brand names.

End node	Parked domain examples
Amazon.com	craigslits.com, 14.de, audii.de
Apple.com	acgeo.com, backlinkscenter.info
Coupons.com	4google.com,agendo.com
Sears.com	uasairways.com, cursoblogger.com
Expedia.com	pizzahutjobs.com, financetasksforce.com

Table 9: Examples of end nodes receiving spam traffic.

Click fraud. All labeled 97K (45.7%) PPC monetization chains are clearly fraudulent clicks, as our crawlers never clicked on any ads. Table 8 provides a breakdown of fraudulent clicks observed from each parking service through ad-nets for which we have a click stamp. By taking a close look at the ad-nets involved, we found that none of the fraudulent clicks on the top-tier networks (`Google AdWords` & `BingAds`) could be attributed to a parking service, due to ad-net syndications. Parking services avoid clicking on top-tier ad-nets' ads because they have a better click fraud detection system than 2nd-tier networks and as such they only happen through ad-net syndication. Additionally, 2% of the fraudulent clicks could not be attributed to a parking service due to the presence of a traffic stamp between the start node (i.e parked domain) and the click stamp. For example, domains parked with `PS6` resulted in fraudulent clicks through a traffic system (`DNTX`) which is owned by the same parent company of `PS6`, namely `TeamInternet AG`.

Also interesting is the observation that not only were the clicks delivered through those chains completely fraudulent but they often came from parked domains that had nothing to do with the ad campaigns at the end nodes. Specifically, we applied the keyword generation and matching approach described above to analyze the relations between the parked domains on those chains and their corresponding end nodes. This study reveals that 61.3% of the fraudulent clicks were from parked domains completely unrelated to the end nodes on their chains. Also given the fact that the average cost-per-click (CPC), which is $0.28, is twice as much as the average cost-per-redirect (CPR) that we paid, there is no legitimacy whatsoever in such click-faking activities.

(a) Revenue Estimates

	P_{FPPC}	P_{PPR}	P_{TS}	P_{PPC}	P_{PPA}	$\frac{Rev_{Fraud}}{Rev}$
PS5	0.01	0.01	0.78	0.97	0.0057	40.3%
PS1	0.0015	0.0003	0.77	0.998	0.00003	7.4%
PS3	0.0004	0.004	0.71	0.995	0.0001	9.3%
PS2	0.0001	0.0000004	0.7	0.9997	0.00016	0.8%
PS6	0	0.015	0.66	0.983	0.0015	18.5%
PS4	0	0.0073	0.64	0.976	0.017	10%

(b) Description of variables used.

Legitimate	P_{PPC}	Probaility of monetization through the display of Pay-Per-Click (PPC) ads.
	P_{PPR}	Probability of monetization through Pay-Per-Redirect (PPR).
	P_{PPA}	Probaility of monetization through affiliate marketing, Pay-Per-Action (PPA).
Fraudulent	P_{FPPC}	Probability of monetization through a fraudulent click on a Pay-Per-Click (PPC) ad.
	P_{TS}	Probaility of monetization through traffic spam in Pay-Per-Redirect (PPR).

Table 10: Estimates of illicit monetization revenues for selected parking services.

Malware distribution. Also discovered in our research is parking services' involvement (probably unwittingly) in malware distribution. We found that many PPR monetization chains were leading to malicious content, either through drive-by downloads or through social engineering scams such as FakeAV or flash player updates (see Figure 8 in Appendix). This occurred because the traffic systems involved did not do their due diligence in detecting the traffic buyers who actually disseminate malware. Using content structure clustering, a technique applied by prior research [13], we concluded that *at least* 3.7% of the PPR traffic buyers spread malware. This illicit activity not only hurts the victims visiting a parked domain but also affects the parked domain when it gets blacklisted by URL scanners such as SafeBrowsing [12], which reduces the monetary value of the parked domain when its owner decides to sell it.

5.4 Revenue Analysis

Model. As discussed before, parking services are unique in that their monetization operations involve both legitimate and illegitimate activities. To understand the economic motives behind this monetization strategy, we analyzed their revenues with a model derived from that used in prior research [23]:

$$Rev = Visits \cdot (Rev_{Fraud} + Rev_{Legit})$$

where the total revenue Rev is calculated from the total number of visits and the average revenue for each visit. This average revenue is further broken down into two components, the part from illicit monetization (Rev_{Fraud}) and that from legitimate monetization (Rev_{Legit}). These components were further estimated as follows:

$$\begin{aligned}Rev_{Fraud} &= P_{FPPC} \cdot CPC + P_{PPR} \cdot CPR \cdot P_{TS} \\ Rev_{Legit} &= P_{PPC} \cdot p_{pclk} \cdot CPC + P_{PPR} \cdot CPR \cdot P_{\widetilde{TS}} \\ &\quad + P_{PPA} \cdot p_{pact} \cdot CPA\end{aligned}$$

Intuitively, the above two equations describe five possible situations when a visit to a parked domain is monetized: illicit activities (click fraud or traffic spam) or legitimate ones (legitimate click, direct traffic or affiliate marking monetization). The revenue component from the illicit activities is estimated using the probability of click fraud P_{FPPC} and that of traffic spam $P_{PPR} \cdot P_{TS}$, where P_{PPR} is the probability of direct traffic monetization (PPR) and P_{TS} is the chance of traffic spam when PPR is chosen, together with the revenues for a click *CPC* and a redirect *CPR*. Similarly, the legitimate revenue component comes from a PPC display (with a probability P_{PPC}) given that a user clicks on one of the ads (with a click-through rate p_{pclk}), type-in traffic (P_{PPR}) when it is *not* subject to traffic spam (with a probability $P_{\widetilde{TS}}$), or affiliate marketing (P_{PPA}) when the user performs the operation expected (with a probability p_{pact}). The revenues for those legitimate activities are *CPC*, *CPR* and *CPA* respectively.

Results. In our analysis, we estimated all the probabilities above (P_{FPPC}, P_{PPR}, P_{TS}, P_{PPC} and P_{PPA}) using the *larger* set of all 24M visits to 100K parked domains in a 5.5-month span (Section 2.3). Also, the click-through rate p_{pclk} and a user's probability of taking an action under PPA were both set to 0.02, and *CPA* to \$0.265, all according to the prior work [23], while CPC and CPR were determined as \$0.28 and \$0.14 respectively, based on the average cost for our ad/traffic campaigns.

In the absence of data about the total number of visits per parking service, all we could do is estimate the portion of its income from the illicit activities to its total revenue, based upon our data. The results are shown in Table 10.

Discussion. From the table, we can see that even reputable parking services like PS2 have at least 0.8% of its revenue come from illicit monetizations. For others, this revenue source is even more significant (e.g., 40.3% for PS5 whom we found to be aggressive in its illicit monetizations). Revenue from fraudulent clicks is found to be zero for PS6 and PS4 because, as described earlier in Section 5.3, they are bouncing their traffic through their own traffic systems and as such we can not attribute fraudulent clicks to them. Note that our estimates here are very conservative, due to the cloaking those services played to our crawlers (which used a small set of IP addresses) and the limited scope of our study (which only covered 11 ad-nets and traffic systems). We expect that the ratios of illicit revenues are much higher in practice.

6 Discussion

Domain Parking Regulation. Our study uncovers the illicit monetizations by parking services but the underlying problem is even graver. Currently, there is no regulation on the behaviors of parking services, which allows them to set up arbitrary terms of services accommodating their own benefits. It is worth noting that parking services may have started to exhibit illicit monetization activities due to the decline in their revenues [3, 29]. Also, our research shows they have a tendency to profit from secondary domains illicitly (Figure 4), due to the difficulty in monetizing those domains through a legitimate channel. Protecting the advertisers' and traffic buyers' benefits in the existence of dishonest parking services is challenging because incoming traffic can be manipulated. What complicates the situation more is that ad-nets could be owned by the company who also runs parking services and the advertisers have no fair party to talk to. Further, direct navigation traffic (i.e. zeroclick) is being advocated by parking services and there is no guarantee on the quality of the incoming traffic. Our research discloses the dark side of parked domain monetization which calls for serious policy efforts to regulate parking services.

Here, we suggest several practices that could mitigate many types of illicit monetization activities when enforced. First, the advertisers should be provided with a clearer picture of the monetization activities. For example, the types of publishers should be marked out as well to advertisers besides their publishers' IDs, which helps advertisers in auditing and monitoring traffic coming from parking services. In fact, some ad-nets are already moving to such direction: for example, Affinity, a popular ad-net, distinguishes publishers by assigning them types such as "in-text" and "domain zero click". We also suggest providing traffic buyers with a way to check the integrity of incoming traffic such as passing the domain name of each start node in the referral. Enforcement and compliance of such mechanisms requires the presence of a 3rd party service (i.e. policy enforcer) in the ecosystem.

Legal and ethical concerns. There are several ethical concerns raised during our study, and we carefully designed our experiments to address them. First, we crawled our own parked domains which is problematic if we earn profit from it. To address this issue, we avoid cashing in the revenues we earned from the 7 parking services hosting our parked domains (\$81.06 in total). Second, we ignore *robots.txt* served by parking services when crawling since we focused on their illicit behaviors. Other studies on malicious activities also ignore the *robots.txt* file [17, 33, 19, 7]. Third, one may question that the artificial traffic generated by our crawler could affect the advertisers or traffic buyers. In fact, we crawled parked domains in a moderate speed and parking services have deployed mechanisms to discern artificial traffic and stop charging the advertisers when identified [28]. Lastly, we ran campaigns with ad-nets and traffic systems but there was no actual business running. The websites and advertisements we set up were coherent to the policies of ad-nets and traffic systems. There was no damage to visitors and we did not collect any Personally Identifiable Information (PII) from them.

7 Related Work

Parking services. Although domain parking services have been here for years, little has been done to understand their security implications. What comes close are the works on typo-squatting [37, 22, 9], which reveals that domain owners utilize this technique for profit. Also, prior research shows that malicious domains tend to be parked once detected [27, 17]. Most related to our research is the study on click spam [10], which focuses on click-spam detection and also mentions the possible involvement of one parking service (Sedo) in such activities based on its JavaScript. Such code was not found in our research. Compared with the prior work, what we did is a systematic study on the illicit activities of parking services, which has never been done before. This is made possible by the new infiltration analysis we performed. Our study not only confirms the presence of illicit operations within parking services but also brings to light their scope and magnitude.

Illicit activities in online advertising. Ad-related illicit activities have been extensively studied. Examples include click-fraud [21, 4, 5, 11, 25], drive-by-download [18], trending-term exploitation [23] and impression fraud [32]. Such prior work all looks at a conventional adversary who performs malicious activities whenever possible. The parking services, however, are very different: they run legitimate business with advertisers and ad networks. However, our study reveals that a significant portion of their revenues actually come from illicit activities, which raises the awareness about this completely unregulated business.

Infiltration into malicious infrastructure. To understand how underground businesses work, a lot of studies attempt to infiltrate their business infrastructure. Examples include the work on Spam [16, 15], CAPTCHA solving [24], blackhat SEO [36] and Pay-per-install networks [8]. Different from such prior research, we need to infiltrate the parking monetization process without disrupting its operations. This was achieved using a new approach through which we controlled some nodes on both ends of the monetization ecosystem and managed to link them together.

8 Conclusion

This paper reports the first systematic study on illicit activities in parked domain monetization. To demystify this "dark side" of parking services, we devised an infiltration analysis to gain control of some start nodes and end nodes of the parking ecosystem, and then connect the dots, sending our crawling traffic across the nodes under our control on the both ends, with the monetization entities (domain parking services, ad networks) in-between. This analysis provided us a unique observation of the whole monetization process, which enabled us to confirm the presence of click fraud, traffic spam and traffic stealing. We further expanded those seed chains to millions of monetization chains collected over 5.5 months, using the stamps of their monetization options. Over such data, our study revealed the pervasiveness of the illicit monetization practices and their revenues, which calls for policy efforts to control those illicit operations.

Acknowledgements

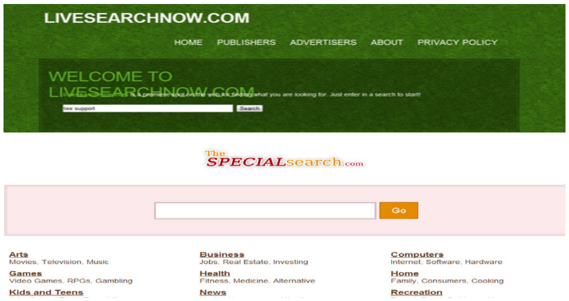
We thank our shepherd Damon McCoy and anonymous reviewers for their insightful comments and suggestions. We thank Yinglian Xie and Fang Yu from Microsoft Research for their valuable comments. This work is supported in part by NSF CNS-1017782, 1117106, 1223477 and 1223495. Any opinions, findings, conclusions or recommendations expressed in this paper do not necessarily reflect the views of the NSF.

Appendix

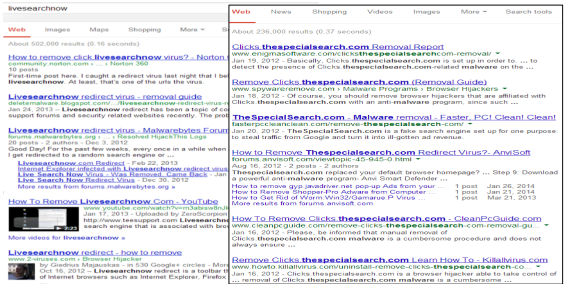
The true operation of the shady search sites. fastonlinefinder.com is one of a large number of search sites we refer to as "shady". They are shady in that they rarely display organic results and emphasize on sponsored ads. Moreover, they have been reported in previous works on click fraud [18, 4] and have been presumed malicious to some extent. Additionally, many victims have often complained about a "redirect" malware hijacking their traffic and redirecting to these search sites as shown in Figure 5.

Through our empirical investigations, we discovered the actual role they play. We found their true operation was to act as click servers for search ads (similar to traditional click servers of other none search advertisements) and as such they are owned and operated by ad-nets. Another use of those search sites was to set the click referral and as such, the advertiser will assume their ad was displayed on the referring search site.

It is important to note here, that the use of such search sites is not illegal. It is only misunderstood due to their abuse by ad-net publishers. A fraudulent publisher will use a malware or Trojan to generate clicks on their ads and since the clicks lead to an ad-net's search sites, the


LIVESEARCHNOW.COM
HOME
PUBLISHERS
ADVERTISERS
ABOUT
PRIVACY POLICY
WELCOME TO LIVESEARCHNOW.COM
is a premiere spot on the web for finding what you are looking for. Just enter in a search to start!
tee support
Search
The SPECIALsearch.com
Go
Arts
Movies, Television, Music
Games
Video Games, RPGs, Gambling
Kids and Teens
Business
Jobs, Real Estate, Investing
Health
Fitness, Medicine, Alternative
News
Computers
Internet, Software, Hardware
Home
Family, Consumers, Cooking
Recreation

(a) Screen shots of two ad-net search sites: `livesearchnow.com` with Advertise & `thespecialsearch.com` with Affinity.



About 502,000 results (0.16 seconds)

How to remove click.livesearchnow virus? - Norton
community.norton.com › ... › Norton 360
10 posts
First-time post here. I caught a redirect virus last night that I bel
livesearchnow. At least, that's one of the urls the virus.

Livesearchnow redirect virus - removal guide
deletemalware.blogspot.com/.../livesearchnow-redirect-virus-r
Jan 24, 2013 – Livesearchnow redirect has been a topic of co
support forums and security related websites recently. The prob

Livesearchnow redirect virus - Malwarebytes Forum
forums.malwarebytes.org › ... › Resolved HijackThis Logs
20 posts - 2 authors - Dec 3, 2012
Good Day! For the past few weeks, every once in a while when
I get redirected to a random search engine or ...
Livesearchnow.com Redirect - Feb 22, 2013
Internet Explorer infected with Livesearchnow redirect virus
Live Search Now Virus - Was Removed, Came Back - Jan
Live Search Now Redirect Virus - Dec 30, 2012
More results from forums.malwarebytes.org »

How To Remove Livesearchnow.Com - YouTube
www.youtube.com/watch?v=m3abisw8nJk
Jan 17, 2013 - Uploaded by ZeroScorpion
http://www.teesupport.com Livesearchno
search engine that is associated with brov
More videos for livesearchnow »

Livesearchnow redirect - how to remove
www.2-viruses.com › Browser Hijacker
by Giedrius Majauskas - in 530 Google+ circles - More
Oct 16, 2012 – Livesearchnow redirect is a toolbar th
of Internet browsers such as Internet Explorer, Firefox

About 236,000 results (0.37 seconds)

Clicks.thespecialsearch.com Removal Report
www.enigmasoftware.com/clicksthespecialsearchcom-removal/
Jan 19, 2012 - Basically, Clicks.thespecialsearch.com is set up in order to. ... to detect the presence of Clicks.thespecialsearch.com-related malware on the ...

Remove Clicks.thespecialsearch.com (Removal Guide)
www.spywareremove.com › Malware Programs › Browser Hijackers
Jan 18, 2012 - Of course, you should remove browser hijackers that are affiliated with Clicks.thespecialsearch.com with an anti-malware program, since such ...

TheSpecialSearch.com - Malware removal - Faster, PC! Clean! Clean!
fasterpccleanclean.com/remove-thespecialsearch-com/
Jan 20, 2012 - TheSpecialSearch.com is a fake search engine set up for one purpose: to steal traffic from Google and turn it into ill-gotten ad revenue.

How to Remove Thespecialsearch.com Redirect Virus?- AnviSoft
forums.anvisoft.com/viewtopic-45-945-0.html
Aug 16, 2012 - 2 posts - 2 authors
Thespecialsearch.com replaced your default browser homepage? ... Step 9: Download a powerful anti-malware program- Anvi Smart Defender ...
How to remove gyp.javadriver.net pop-up Ads from your ... 1 post Jan 26, 2014
How to Remove Shopper-Pro Adware from Computer ... 1 post Jan 21, 2014
How to Get Rid of Worm:Win32/Gamarue.P Virus ... 1 post Mar 21, 2013
More results from forums.anvisoft.com

How To Remove Clicks.thespecialsearch.com - CleanPcGuide.com
www.cleanpcguide.com/remove-clicks-thespecialsearch-com-removal-gu...
Jan 16, 2012 - Please, be informed that manual removal of Clicks.thespecialsearch.com malware is a cumbersome procedure and does not always ensure ...

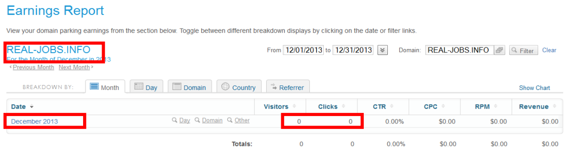
Remove Clicks.thespecialsearch.com Learn How To - Killallvirus.com
www.howto.killallvirus.com/uninstall-remove-clicks-thespecialsearch-co...
Jan 16, 2012 - Clicks.thespecialsearch.com is a browser hijacker able to take control of ... removal of Clicks.thespecialsearch.com malware is a cumbersome ...

(b) Google search results showing malware complaints for the two search sites.

Figure 5: Ad-net search site examples showing screen shots of the search sites and malware related complaints by users.

search site become wrongly accused as the malicious party.

Evidence Survey. Along our infiltration we collect a set of evidence to support our findings. We start with Figure 6 which confirms traffic stealing by one parking service that is not reporting traffic, as shown in 6(a), which have been monetized through `7search` and verified by our payment for the traffic as shown in 6(b).


Earnings Report

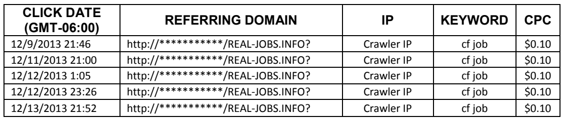
View your domain parking earnings from the section below. Toggle between different breakdown displays by clicking on the date or filter links.

REAL-JOBS.INFO
For the Month of December in 2013
‹ Previous Month Next Month ›

From 12/01/2013 to 12/31/2013 Domain: REAL-JOBS.INFO Filter Clear

Date		Visitors	Clicks	CTR	CPC	RPM	Revenue
December 2013	Day Domain Other	0	0	0.00%	$0.00	$0.00	$0.00
	Totals:	0	0	0.00%	$0.00	$0.00	$0.00

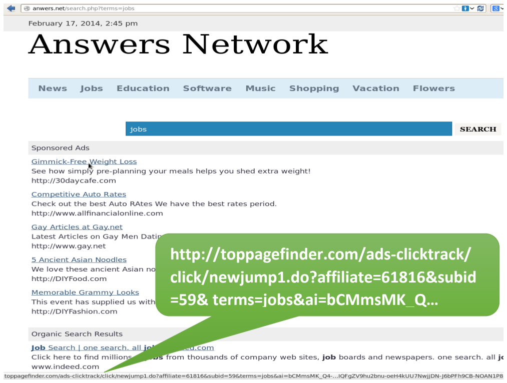
(a) Screen shot of our parked domain revenue report with the parking service in question.


CLICK DATE (GMT-06:00)	REFERRING DOMAIN	IP	KEYWORD	CPC
12/9/2013 21:46	http://***********/REAL-JOBS.INFO?	Crawler IP	cf job	$0.10
12/11/2013 21:00	http://***********/REAL-JOBS.INFO?	Crawler IP	cf job	$0.10
12/12/2013 1:05	http://***********/REAL-JOBS.INFO?	Crawler IP	cf job	$0.10
12/12/2013 23:26	http://***********/REAL-JOBS.INFO?	Crawler IP	cf job	$0.10
12/13/2013 21:52	http://***********/REAL-JOBS.INFO?	Crawler IP	cf job	$0.10

(b) Billing report by 7Search shows 5 billed traffic hits from our parked domain. Part of the referral is removed to anonymize the parking service.

Figure 6: Traffic stealing observed on our parked domain

Additionally, We verify the association of search sites to ad-nets by registering with two ad-nets (`Advertise` & `Bidvertiser`) as a publisher interested in displaying their sponsored ads. We set up our website with a search service that pulls organic search results from `Google` and sponsored ads from the two ad-nets we registered with. By pulling sponsored ads from the ad-nets, we verified the use of search sites as the click URLs as shown in Figure 7 which shows one click URL by `Advertise` that has the same URL tokens as the URLs in Table 7. Actually, the same website, `toppagefinder.com`, appeared also in our data set and as such was in the same UIC which was a correct association. Note that the same website used here for our publisher was also used for our advertising and traffic buying campaigns.



February 17, 2014, 2:45 pm

Answers Network

Sponsored Ads

Gimmick-Free Weight Loss
See how simply pre-planning your meals helps you shed extra weight!
http://30daycafe.com

Competitive Auto Rates
Check out the best Auto RAtes We have the best rates period.
http://www.allfinancialonline.com

Gay Articles at Gay.net
Latest Articles on Gay Men Datin
http://www.gay.net

5 Ancient Asian Noodles
We love these ancient Asian no
http://DIYFood.com

Memorable Grammy Looks
This event has supplied us with
http://DIYFashion.com

Organic Search Results

Job Search | one search. all jo ed.com
Click here to find millions bs from thousands of company web sites, job boards and newspapers. one search. all jo
www.indeed.com

toppagefinder.com/ads-clicktrack/click/newjump1.do?affiliate=61816&subid=59&terms=jobs&ai=bCMmsMK_Q4-...lQFgZV9hu2bnu-oeH4kUU7NwjjDN-J6bPFh9CB-NOAN1P8

http://toppagefinder.com/ads-clicktrack/click/newjump1.do?affiliate=61816&subid=59& terms=jobs&ai=bCMmsMK_Q...

Figure 7: Our Advertiser, publisher and traffic buyer website.

Finally, in Figure 8 we show examples of visits to domains parked with `PS5` leading to malware downloads through two traffic systems, namely `DNTX` and `ZeroRedirect`.

References

[1] ADS, B. Bing ads api. `https://developers.bingads.microsoft.com/`.

[2] ALEXA. Alexa top global sites. `http://www.alexa.com/topsites`, February 2014.

[3] ALLEMANN, A. Sedo reports continuing decline in domain parking. `https://domainnamewire.com/2013/11/12/sedo-reports-continuing-decline-in-domain-parking/`, November 2013.

[4] ALRWAIS, S. A., GERBER, A., DUNN, C. W., SPATSCHECK, O., GUPTA, M., AND OSTERWEIL, E. Dissecting ghost clicks: Ad fraud via misdirected human clicks. In *Proceedings of the 28th Annual Computer Security Applications Conference* (New York, NY, USA, 2012), ACSAC '12, ACM, pp. 21–30.

[5] BLIZARD, T., AND LIVIC, N. Click-fraud monetizing malware: A survey and case study. In *Malicious and Unwanted Software (MALWARE), 2012 7th International Conference on* (Oct 2012), pp. 67–72.

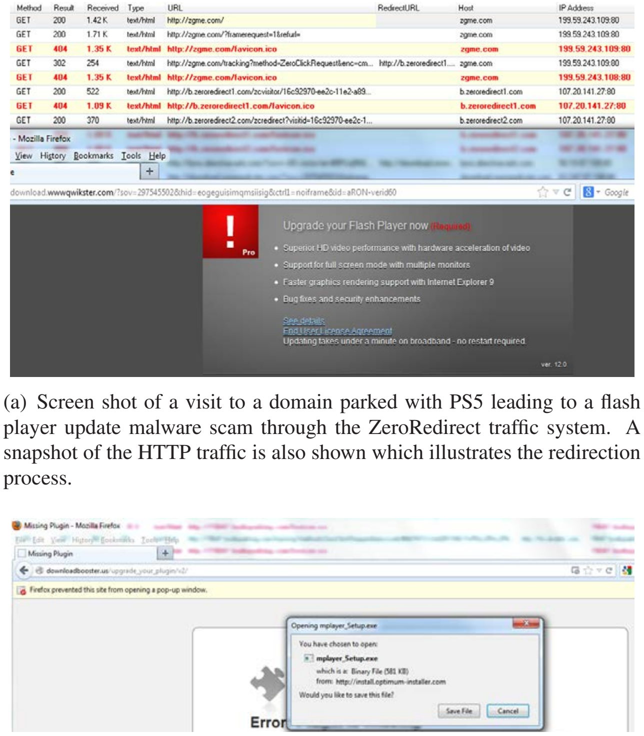
[6] BODIS. Javascript and xml api. `https://www.bodis.com/news/javascript-and-xml-api`.

[7] BORGOLTE, K., KRUEGEL, C., AND VIGNA, G. Delta: Automatic Identification of Unknown Web-based Infection Campaigns. In *Proceedings of the ACM Conference on Computer and Communications Security* (2013), CCS '13, ACM.

[8] CABALLERO, J., GRIER, C., KREIBICH, C., AND PAXSON, V. Measuring pay-per-install: The commoditization of malware distribution. In *Proceedings of the 20th USENIX Conference on Security* (Berkeley, CA, USA, 2011), SEC'11, USENIX Association, pp. 13–13.

[9] COULL, S., WHITE, A., YEN, T.-F., MONROSE, F., AND REITER, M. Understanding domain registration abuses. In *Security and Privacy Silver Linings in the Cloud*, K. Rannenberg, V. Varadharajan, and C. Weber, Eds., vol. 330 of *IFIP Advances in Information and Communication Technology*. Springer Berlin Heidelberg, 2010, pp. 68–79.

[10] DAVE, V., GUHA, S., AND ZHANG, Y. Measuring and fingerprinting click-spam in ad networks. *SIGCOMM Comput. Commun. Rev. 42*, 4 (Aug.



Method	Result	Received	Type	URL	RedirectURL	Host	IP Address
GET	200	1.42 K	text/html	http://zgme.com/		zgme.com	199.59.243.109:80
GET	200	1.71 K	text/html	http://zgme.com/?framerequest=1&refurl=		zgme.com	199.59.243.109:80
GET	404	1.35 K	text/html	http://zgme.com/favicon.ico		zgme.com	199.59.243.109:80
GET	302	254	text/html	http://zgme.com/tracking?method=ZeroClickRequest&enc=cm...	http://b.zeroredirect1....	zgme.com	199.59.243.109:80
GET	404	1.35 K	text/html	http://zgme.com/favicon.ico		zgme.com	199.59.243.108:80
GET	200	522	text/html	http://b.zeroredirect1.com/zcvisitor/16c92970-ee2c-11e2-a89...		b.zeroredirect1.com	107.20.141.27:80
GET	404	1.09 K	text/html	http://b.zeroredirect1.com/favicon.ico		b.zeroredirect1.com	107.20.141.27:80
GET	200	370	text/html	http://b.zeroredirect2.com/zcredirect?visitid=16c92970-ee2c-1...		b.zeroredirect2.com	107.20.141.27:80

- Mozilla Firefox

View History Bookmarks Tools Help

download.wwwqwikster.com/?sov=297545502&hid=eogeguisimqmsiisig&ctrl1=noiframe&id=aRON-verid60

Upgrade your Flash Player now (Required)

- Superior HD video performance with hardware acceleration of video
- Support for full screen mode with multiple monitors
- Faster graphics rendering support with Internet Explorer 9
- Bug fixes and security enhancements

See details
End User License Agreement
Updating takes under a minute on broadband - no restart required

ver. 12.0

(a) Screen shot of a visit to a domain parked with PS5 leading to a flash player update malware scam through the ZeroRedirect traffic system. A snapshot of the HTTP traffic is also shown which illustrates the redirection process.

Missing Plugin - Mozilla Firefox

Missing Plugin

downloadbooster.us/upgrade_your_plugin/v2/

Firefox prevented this site from opening a pop-up window.

Opening mplayer_Setup.exe

You have chosen to open:

mplayer_Setup.exe
which is a: Binary File (581 KB)
from: http://install.optimum-installer.com

Would you like to save this file?

Save File Cancel

Error

(b) Screen shot of a visit to a domain parked with PS5 leading to malware download through the DNTX traffic system

Figure 8: Visits to parked domains leading to malware distribution.

2012), 175–186.

[11] Dave, V., Guha, S., and Zhang, Y. Viceroi: Catching click-spam in search ad networks. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer &#38; Communications Security* (New York, NY, USA, 2013), CCS '13, ACM, pp. 765–776.

[12] Google. Safe browsing api google developers. `https://developers.google.com/safe-browsing/`.

[13] Hachenberg, C., and Gottron, T. Locality sensitive hashing for scalable structural classification and clustering of web documents. In *Proceedings of the 22Nd ACM International Conference on Conference on Information &#38; Knowledge Management* (New York, NY, USA, 2013), CIKM '13, ACM, pp. 359–368.

[14] HUANG, W. Parked domain numbers and traffic, and more on the exploits served. `http://blog.armorize.com/2010/08/parked-domain-numbers-and-traffic-and.html`, August 2010.

[15] Kanich, C., Weavery, N., McCoy, D., Halvorson, T., Kreibichy, C., Levchenko, K., Paxson, V., Voelker, G. M., and Savage, S. Show me the money: characterizing spam-advertised revenue. In *Proceedings of the 20th USENIX conference on Security* (Berkeley, CA, USA, 2011), SEC'11, USENIX Association, pp. 15–15.

[16] Levchenko, K., Chachra, N., Enright, B., Felegyhazi, M., Grier, C., Halvorson, T., Kanich, C., Kreibich, C., Liu, H., McCoy, D., Pitsillidis, A., Weaver, N., Paxson, V., Voelker, G. M., and Savage, S. Click Trajectories: End-to-End Analysis of the Spam Value Chain. In *Proceedings of 32nd annual Symposium on Security and Privacy* (May 2011), IEEE.

[17] Li, Z., Alrwais, S., Xie, Y., Yu, F., and Wang, X. Finding the linchpins of the dark web: a study on topologically dedicated hosts on malicious web infrastructures. In *Proceedings of the 2013 IEEE Symposium on Security and Privacy* (Washington, DC, USA, 2013), SP '13, IEEE Computer Society, pp. 112–126.

[18] Li, Z., Zhang, K., Xie, Y., Yu, F., and Wang, X. Knowing your enemy: understanding and detecting malicious web advertising. In *Proceedings of the 2012 ACM conference on Computer and communications security* (New York, NY, USA, 2012), CCS '12, ACM, pp. 674–686.

[19] Lu, L., Perdisci, R., and Lee, W. Surf: detecting and measuring search poisoning. In *Proceedings of the 18th ACM conference on Computer and communications security* (New York, NY, USA, 2011), CCS '11, ACM, pp. 467–476.

[20] Mahjoub, D. A look at the relationship between parked domains and malware. `http://labs.umbrella.com/2013/03/20/discovery-of-new-suspicious-domains-using-authoritative-dns-traffic-and-parked-domains-analysis/`, March 2013.

[21] Miller, B., Pearce, P., Grier, C., Kreibich, C., and Paxson, V. What's clicking what? techniques and innovations of today's clickbots. In *Proceedings of the 8th international conference on Detection of intrusions and malware, and vulnerability assessment* (Berlin, Heidelberg, 2011), DIMVA'11, Springer-Verlag, pp. 164–183.

[22] Moore, T., and Edelman, B. Measuring the perpetrators and funders of typosquatting. In *Proceedings of the 14th International Conference on Financial Cryptography and Data Security* (Berlin, Heidelberg, 2010), FC'10, Springer-Verlag, pp. 175–191.

[23] Moore, T., Leontiadis, N., and Christin, N. Fashion crimes: Trending-term exploitation on the web. In *Proceedings of the 18th ACM Conference on Computer and Communications Security* (New York, NY, USA, 2011), CCS '11, ACM, pp. 455–466.

[24] Motoyama, M., Levchenko, K., Kanich, C., McCoy, D., Voelker, G. M., and Savage, S. Re: Captchas: Understanding captcha-solving services in an economic context. In *Proceedings of the 19th USENIX Conference on Security* (Berkeley, CA, USA, 2010), USENIX Security'10, USENIX Association, pp. 28–28.

[25] Pearce, P., Grier, C., Paxson, V., Dave, V., McCoy, D., Voelker, G. M., and Savage, S. The zeroaccess auto-clicking and search-hijacking click fraud modules. Tech. Rep. UCB/EECS-2013-211, EECS Department, University of California, Berkeley, Dec 2013.

[26] Petnel, R. Easylist. `https://easylist-downloads.adblockplus.org/easylist.txt`.

[27] Rahbarinia, B., Perdisci, R., Antonakakis, M., and Dagon, D. Sinkminer: Mining botnet sinkholes for fun and profit. In *Presented as part of the 6th USENIX Workshop on Large-Scale Exploits and Emergent Threats* (Berkeley, CA, 2013), USENIX.

[28] Sedo. Domain parking terms and conditions. `https://sedo.com/us/about-us/policies/domain-parking-terms-and-conditions-sedocom/?tracked=1&partnerid=38758&language=us`.

[29] Sedo Holding. Sedo holding ag 6-month report. `http://www.sedoholding.com/fileadmin/user_upload/Dokumente/English/Reports_2013/Sedo_Holding_6M_Report_2013.pdf`, 2013.

[30] SIE, I. Security information exchange (sie) portal. `https://sie.isc.org/`.

[31] Sinka, M. P., and Corne, D. W. Towards modernised and web-specific stoplists for web document analysis. In *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on* (2003), IEEE, pp. 396–402.

[32] Springborn, K., and Barford, P. Impression fraud in online advertising via pay-per-view networks. In *Proceedings of the 22Nd USENIX Conference on Security* (Berkeley, CA, USA, 2013), SEC'13, USENIX Association, pp. 211–226.

[33] Stringhini, G., Kruegel, C., and Vigna, G. Shady paths: Leveraging surfing crowds to detect malicious web pages. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer &#38; Communications Security* (New York, NY, USA, 2013), CCS '13, ACM, pp. 133–144.

[34] Tools, D. Daily dns changes and web hosting activity. `http://www.dailychanges.com/`, February 2014.

[35] Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (2003), Association for Computational Linguistics, pp. 173–180.

[36] Wang, D. Y., Savage, S., and Voelker, G. M. Juice: A longitudinal study of an seo botnet. In *NDSS* (2013), The Internet Society.

[37] Wang, Y.-M., Beck, D., Wang, J., Verbowski, C., and Daniels, B. Strider typo-patrol: Discovery and analysis of systematic typo-squatting. In *Proceedings of the 2Nd Conference on Steps to Reducing Unwanted Traffic on the Internet - Volume 2* (Berkeley, CA, USA, 2006), SRUTI'06, USENIX Association, pp. 5–5.