Measurement and Analysis of Traffic Exchange Services

Mobin Javed
U.C. Berkeley
mobin@cs.berkeley.edu

Cormac Herley
Microsoft Research
cormac@microsoft.com

Marcus Peinado
Microsoft Research
marcuspe@microsoft.com

Vern Paxson
U.C. Berkeley, International
Computer Science Institute
vern@cs.berkeley.edu

Abstract

Traffic exchange services enable members to bring traffic to their websites from a diverse pool of IP addresses, in return for visiting sites of other members. We examine the world of traffic exchanges to characterize their makeup, usage, and monetization. We find that the ecosystem includes a range of services, from manual exchanges where participants must solve CAPTCHAs between successive page views, to exchanges that provide tools that automatically surf without requiring any user action. By "milking" a sample of these exchanges, we analyze month-long datasets to examine the nature of URLs that members submit to them. We find a wide prevalence of URLs for services that pay users in return for views to their content, and at least 30% of the requested impressions are for pages that clearly participate in a class of impression fraud called *referrer spoofing*. We also analyze the size and composition of a sample of these exchange networks by making purchases, finding that the exchanges delivered visits from roughly 200K unique IP addresses, and that in some exchange networks, the majority of visits came from cloud hosting services.

Categories and Subject Descriptors

C.2.0 [**General**]: Security and protection

General Terms

Measurement; Advertising fraud

Keywords

Traffic exchanges; Impression fraud; Click fraud

1. INTRODUCTION

An eco-system has arisen to provide free traffic from a diverse pool of IP addresses using the principle of *exchange*: members visit each other's websites, in return receiving visits to their own websites. Exchanges of significant size can potentially enable members to effectively "launder" clicks to their sites, making them appear of a much richer nature than in reality.

IMC'15, October 28–30, 2015, Tokyo, Japan.

ACM 978-1-4503-3848-6/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2815675.2815708.

Understanding the significance and likely evolution of such exchanges requires identifying the motives of the participants. One obvious motive—though not the only one, it turns out—concerns forms of advertising fraud. Internet advertisers generally desire "organic" visitors, i.e., truly engaged users. Unscrupulous publishers of Internet content, however, can seek to receive cheap inorganic traffic in order to fraudulently earn from purported impressions and clicks received on the advertisements they display. Advertisers can identify simple forms of bogus traffic by observing impressions or clicks disproportionately due to the same IP addresses over a short period of time. Fraudsters in turn will seek ways to spread out the apparent pool of users visiting their pages to thwart such detection.

Indeed, over the past decade different schemes for generating fraudulent traffic from a diverse pool have arisen. By their sheer size, botnets are well suited to generate traffic from disparate sources, as evidenced by impression and click-fraud botnets such as ZeroAccess, Fiesta, and 7cy [31, 33]. Another approach, *pay-per-view networks*, embeds websites as pop-unders in other websites, seemingly delivering the visitors of one website to another [34].

A third approach, one that our study examines in detail, is the use of traffic exchanges to pool clicks, allowing participants to bring disparate traffic to their sites on-the-cheap. In such exchange networks, participants performing N clicks for the exchange, all coming from the participant's IP address, in return reap credits to receive a different N visits (or somewhat fewer, depending on the exchange's structure) delivered to URLs of their choosing. Since the exchange has many members, the traffic arriving at each member's target site appears quite dispersed.

More broadly, we examine the world of traffic exchanges, finding a wide variety of offerings, and usage paradigms beyond just ad fraud. The offerings range from manual exchanges, where participants must solve CAPTCHAs between successive clicks, to exchanges that provide tools that automatically surf without requiring any user action, and allow customization of time-on-page, User-Agent, and Referer fields. Some exchanges are free, while others charge for certain levels of service. We study several prominent exchanges in each category, casting light on this little-studied phenomenon and the monetization it supports.

By lurking as a non-active exchange participant, we observe both the origin (which IP addresses) and the destination (what websites) of traffic in circulation. This allows us to estimate the size of exchanges and understand the strategies pursued by those using them. We find that some exchanges deliver traffic from tens of thousands of IP addresses. A large fraction of the traffic that flows through the *autosurf* exchanges is generated by bots running in cloud hosting services, and terminates on websites that generate a multitude of impression requests to ad networks. Such *referrer spoofing* ad fraud makes up one of the major categories we observed [35, 25]. In contrast, members of the *manual* exchanges appear to indeed reflect the sought-after demographic of the ads displayed there.

Two factors distinguish our work from previous studies on traffic generation. First, while there have been studies of services that deliver bulk traffic for money, ours is the first large-scale examination of exchange networks primarily designed for *free* exchange of traffic amongst participants, though they also offer traffic for purchase. Second, the targets of traffic exchanges potentially differ in character from those for paid sites; since participants barter for their traffic with reciprocal traffic, they may pursue different monetization strategies from those who pay for traffic.

The rest of this paper proceeds as follows: § 2 covers related work and § 3 provides basic background on the different categories of exchanges. § 4 describes the characteristics of the exchanges we study. § 5 describes our measurements of popular target websites in the exchanges. In § 6 we measure the size of the exchanges in terms of unique participants and provide estimates of the impressions these exchanges serve per day. In § 7 we discuss our findings and provide rough estimates of the volume of money flowing through the exchanges.

2. RELATED WORK

Traffic generation and advertising fraud have received considerable study. Work on advertising fraud lies in three categories: (i) impression fraud and markets for buying traffic, (ii) characterization studies on click fraud, and (iii) detection of fraudulent clicks.

Zhang et al. study the quality of traffic delivered by traffic providers who promise bulk visitors for a monthly cost, and compare that to the traffic obtained from traditional keyword auctions and sponsored banner ads in search results [37]. They purchase traffic from these providers, and based on characterization of mouse activity, sub-page views, and timing of visits on their websites, find that these providers mostly deliver inorganic/automated traffic; only one of the providers delivered seemingly organic traffic via a paid-to-view scheme for recruiting visitors.

The work of Springborn et al. on impression fraud appears most relevant to our paper [34]. The authors bought views from multiple traffic sources and studied how the traffic is delivered to their honeypot websites by investigating the referer field. They identify various pay-per-view (PPV) networks, in which participating websites load multiple other websites as pop-unders, i.e., in windows hidden behind the active browser window. This way, whenever a legitimate user arrives at the participating website, they unknowingly also visit the websites loaded as pop-unders. By mining the publicly available Common Crawl Dataset [7], the authors identify $\approx$ 11K unique domains that embed PPV tags, similar to standard ad tags, from the ten networks they study. By looking at the traffic statistics of these websites at MuStat [19], they estimate that these networks deliver on the order of 150M impressions per day to websites, which in turn load up $\approx$ 500M ad impressions per day. Based on this analysis, the authors suggest simple countermeasures, to the effect of checking the referer against a blacklist of PPV networks/publishers that participate in these networks, and checking the viewport dimensions to make sure that the ad is indeed displayed in a visible area in a browser. Pay-per-view networks differ from traffic exchange networks in how they generate traffic and their business model. While traffic exchange networks often offer free traffic using the principle of exchange, pay-per-view networks generate traffic using participating websites, to whom they pay cheap CPM (cost per thousand impressions) rates, and then sell the generated traffic at a profit.

We are unaware of any work on view-fraud botnets in the academic literature. However, at least two botnets intended for view fraud have received blogging attention: (i) The Chameleon botnet estimated to consist of $\approx$ 120K machines targeted a set of 200 websites, and repeatedly cleared cookies on the victim's browser to appear as a new user on every visit to the website [9]; (ii) The TDSS botnet, which leveraged the ClickIce Ad Exchange to deliver bot traffic to various publishers, who in turn defrauded display advertisers [8]. This botnet utilized the pay-per-click model of the ClickIce Exchange to request *text ads* that contained a clickable publisher link. TDSS bots have the following capabilities: (i) deliver traffic to publishers by visiting the clickable link; some publishers optimize their webpages for this bot traffic—if a real user visits, they are shown pages with content, but if the bot requests the same page, a page with only ads is displayed, (ii) use the browser cookies of the infected user to make the ad networks on the publisher's website believe that the ad-slot is for a high-value user, and (iii) spoof mouse, scroll, and click events on the target webpages.

Recent years have seen characterization studies of click fraud bots. Miller et al. studied the Fiesta and 7cy clickbot families to understand the click fraud techniques they employ [31]. For Fiesta, they observe that, similar to TDSS, the pay-per-click model is used to deliver traffic to publishers rather than directly to the advertisers in order to serve as an intermediary layer; the publishers act in an advertiser role on *obscure* ad exchanges in order to capture the traffic generated from the bots. For 7cy, they find traffic directly delivered to the publishers, with the bots mimicking human browsing behavior closely and acting upon C&C instructions to click on ads. Pearce et al. examine the activities of the ZeroAccess botnet, and estimate that it plausibly induced losses of $100K per day until Microsoft's takedown operation in late 2013 [33]. Finally, Stone-Gross et al. document the fraudulent activities in online ad exchanges, as observed from the perspective of a participating ad network [35].

Detection of fraudulent traffic has received less attention in the literature, primarily due to the lack of access to data from advertising networks. Haddadi proposed a technique to fight click fraud using "bluff" ads—ads having content that a normal human user is highly unlikely to click, but click-bots would click since they do not base clicking on comprehensibility of the content [27]. Based on this idea, Dave et al. developed a technique to estimate the fraction of click spam from the perspective of an advertiser, independently from the estimates of the ad network [24]. They signed up as an advertiser on ten different networks and collected click data for their ads (26M impressions over a period of 50 days). Analyzing this data, they identified click spam delivered using malware, parked domains, and arbitrage.

Lewis and Rao point out the difficulty of measuring the Return on Investment for advertising campaigns; the noisiness of the efficacy signals available to advertisers makes fraud hard to detect [29].

Click fraud is but one web-based fraud phenomenon that has surged in recent years. Careful measurements have played a large part in furthering our understanding of different types of fraud and abuse. Kanich et al., for example, study the returns from several large spam campaigns [28]. Grier et al. characterize the spam found on Twitter [26]. Christin details activity found on a large underground market for illicit goods [23]. McCoy et al. study the affiliate programs that are commonly used in selling medications online [30]. These measurement studies reveal phenomena that are more complex than they first appear and inform efforts to combat the fraud.

3. BACKGROUND

We surveyed the landscape of traffic exchanges by performing online web searches. We compiled a list of 50 such exchanges, and manually investigated them to understand the features they offer. In our analysis, we observed two categories of traffic exchanges: (i) generic, and (ii) social media promotion, where *generic* refers to an exchange which offers traffic exchange service for any website, whereas *social media promotion* refers to an exchange designed specifically for increasing views, likes, comments, and/or subscribers/friends/followers on social platforms such as Twitter, Facebook and YouTube. The latter category often requires members to use their social network accounts for participation in the exchange.

In this paper, we focus on *generic* exchanges as the destinations to which the participants direct *free* traffic. Determining their reasons for doing so presents interesting puzzles.

We further classify the *generic* exchanges into the following categories based on the quality of traffic they offer:

(i) Manual: Manual surf exchanges require an exchange member to solve a CAPTCHA for each website view. This check allows the exchange to claim that it offers human surfers for website views.

(ii) Basic Autosurf: Basic autosurf exchanges provide the members a surf link that they can open in a browser for autosurf. Javascript on this page automatically fetches the next site to be viewed periodically and opens it inside an iframe within the surf page.

(iii) Advanced Autosurf: Advanced autosurf exchanges provide the members a tool/plugin, that they can download and run on their machines. In addition to automatically fetching and loading the next page to be viewed, this tool can generate webpage view requests with custom *Referer* and *User-Agent*.

We look at a sample of nine exchanges in this paper, which we chose according to their popularity in Google search results, the Internet marketing underground forum *Black Hat World* [5], and various traffic exchange ranking lists. We picked two manual exchanges: EasyHits4U [10] and HitSafari [15]; two basic autosurf exchanges: 247AutoHits [2] and 10Khits [1]; and four advanced autosurf exchanges: HitLeap [14], eBesucher [11], Jingling [16], Otohits [20]. We also look at one of the social media promotion exchanges, EnhanceViews Autowatcher [12], for the purpose of comparing view traffic characteristics.

4. EXCHANGE CHARACTERISTICS

In this section we provide an overview of the characteristics of the exchanges, such as account pricing and the configuration options that each exchange provides for the traffic it delivers. Table 1 summarizes the characteristics of the exchanges we studied. We discuss these in detail below.

4.1 Earning and using traffic exchange credits

All exchanges, with the exception of Jingling, require a user to sign up for an account on their website. The sign-up process is simple, requiring only a valid email address. Some of the exchanges offer premium memberships, which cost extra, but provide a better exchange ratio (of sent to received views), monthly bonus points, or enhanced functionality (see Table 1). The exchanges provide a link or tool to the registered members, which they can view/run on their machine(s) to earn exchange credits. The autosurf link/tool communicates with the exchange Command & Control server periodically to get the list of URLs to view, whereas the manual exchanges require the user to solve a CAPTCHA for each URL fetch from the server. The following are some of the features that differentiate exchanges and membership levels.

- **Account-to-Machine Ratio**: The manual surf exchanges enforce that one account be used only from one IP address at a time. On the other hand, all the autosurf exchanges we studied allow an account to be used from multiple IP addresses simultaneously.

- **Exchange Ratio**: For free accounts, most exchanges offer an exchange ratio of less than 1 (often in the range 0.5–0.7). This means that free members do not receive the same number of visits to their own websites as the number of websites they view. Instead, the exchange takes a cut, thereby enabling it to sell views directly for money. One autosurf exchange offers a ratio greater than 1 for premium levels of membership. The high premium membership cost at this exchange presumably allows the exchange to generate the extra clicks for premium members using its own infrastructure if the order cannot be fulfilled using the exchange members.

- **Credits/Cash-Out**: All the exchanges we investigated offer traffic exchange credits as a reward for surfing/running their autosurf tool. Only EasyHits4U and eBesucher offer a conversion of the credits to cash pay-out (via Paypal). EasyHits4U pays \$0.30 per 1,000 website views, and has a minimum cash-out of \$3. eBesucher pays \$0.02 per 1,000 views, and has a minimum cash-out of ≈\$3 as well. To the best of our knowledge, cashing out on eBesucher involves an identity check, requiring a scanned copy of a valid photo identity document. This means that anyone who wishes to remain anonymous must find an indirect path to monetize their credits. We did not try to cash-out from EasyHits4U, since that would have required significant effort (solving 10K CAPTCHAs).

4.2 Controls offered for views

The exchanges offer a range of capabilities to make the traffic resemble that of an organic user population. Some of these are offered only to premium members, which comes at a monthly membership price of 3–7€, depending on the exchange and the service level (see Table 1).

- **Website Slots:** Website slots are the number of URLs that a member can have simultaneously active for viewing in the exchange (the URLs can be pages of the same website or different websites). The number of slots vary with membership level on most exchanges, with a free account offering between 3–15 slots depending on the exchange, and paid accounts offering a higher number of slots—unlimited in some cases. Some exchanges offer the same number of slots regardless of membership level. For example, Jingling offers 400 slots, and Otohits and EnhanceViews Autowatcher offer unlimited slots for all tiers.

  An exchange user who wishes to direct traffic at a small number of URLs might manage with few slots; however, a user who wishes to generate views to a large number of URLs (for example, different videos on a YouTube channel) might need more than what the free tiers offer.

- **Views Timing**: This refers to how long a website will be viewed for, which can affect whether the view is credited for view count and payment purposes on the target website. Manual and basic autosurf exchanges offer fairly short views

Exchange	Requires Account?	Earning Credits		Delivering Traffic								
		Machines Per Account	Cash Payout	Account Level	Cost	Website Slots	Exchange Ratio	Geo Targeting	Referer		Randomize View Time	Pageview Time
									Anonymous	Custom		
Manual												
HitSafari	Yes	1	No	I	Free	10	0.5	-	-	-	No	6s
				II	$5	30	1.0	-	-	-		
				III	$9.99	40	1.0	-	-	-		
EasyHits4U	Yes	1	Yes	I	Free	15	1.0	✗	-	-	No	10–40s
				II	$7.95	UL	1.0	✓	-	-		
				III	$19.95	UL	1.0	✓	-	-		
Basic Autosurf												
10Khits	Yes	UL	No	I	Free	3	0.5	-	-	-	No	10s
				II	$10	10	1.2	-	-	-		
				III	$29	45	2.0	-	-	-		
247AutoHits	Yes	UL	No	I	Free	5	0.7	-	-	-	No	30s
				II	$6.99	10	0.8	-	-	-		
				II	$9.99	100	1.0	-	-	-		
Advanced Autosurf												
HitLeap	Yes	UL	No	I	Free	3	0.7	-	✗	✗	Yes	10–60s
				II	3€	5	0.8	-	✓	✗		
				III	7€	15	1.0	-	✓	✓		
eBesucher	Yes	UL	Yes	I	Free	15	1.0	✓	✗	✗	No	15s–10m
				II	2.90€	50	1.0	✓	✓	✓		
				III	5.90€	150	1.0	✓	✓	✓		
Jingling	No	UL	No	I	Free	400	1.0	✓†	✓	✗	Yes	20s–40s
				II	6.34€	400	1.0	✓†	✓	✓		20s–2m
Otohits	Yes	UL	No	-	Free	UL	Variable ∓	-	✓	✓	Yes	10s–10m
EnhanceViews	Yes	UL	No	-	Free	UL	1.0	-	-‡	-‡	No	30s–2m

Table 1: The different traffic exchanges offer different levels of service. For example, the Basic service at eBesucher is free, allows geo-targeting, but will not allow anonymous or custom referer-fields, while the premium service costs 2.90€ per month and allows custom and anonymous referers. † indicates that the exchange offers geo-targeting at the granularity of China vs. Non-China. ‡ indicates the exchange does not offer setting a referer value, and we lack visibility to comment on the default value. ∓ indicates the exchange ratio decreases with the number of website "slots" used (see below). UL indicates Unlimited. '–' indicates that the exchange does not offer the feature at any membership level.

(6–20s) for all service levels, and do not offer randomizing the length of the views. Among the advanced autosurf exchanges, Jingling offers the shortest views (view-length of 20–40s) for a free account, but offers longer views at an additional cost. HitLeap offers a maximum of 60s view time, and provides the option of randomizing the time for each view. eBesucher, Otohits and EnhanceViews Autowatcher offer longer views (2m–10m), but all views are of the same length. Longer views cost more on all exchanges.

- **User-Agent**: This option is only offered by advanced autosurf exchanges. Both Otohits and Jingling offer configuring a custom User-Agent. Others do not specifically offer this option. However, based on the traffic we bought (which we discuss later in Section 6), we observe HitLeap and eBesucher, by default, use a variety of User-Agents, whereas Jingling, by default, provides only various versions of Internet Explorer. We lack visibility for EnhanceViews Autowatcher; we did not buy any traffic from this exchange because directing autosurf traffic to YouTube potentially raises ethical concerns.

- **Anonymous/Custom Referer**: This option is also only offered by advanced autosurf exchanges; the manual and basic autosurf exchanges send the link of the exchange itself as referer. All advanced autosurf exchanges except EnhanceViews Autowatcher offer sending a custom *referer* for the view requests at an additional cost (see Table 1 for details). HitLeap and eBesucher also charge for hiding the default referer which indicates the views come from the traffic exchange. Jingling by default does not send a referer. Obviously a user-agent or referer field that indicates the click comes from a traffic exchange would allow a website to easily filter the autosurf traffic.

- **Geo-Targeting**: Geo-targeting refers to advertising the page for viewing to exchange members residing in a specific location. eBesucher and EasyHits4U offer geo-targeting at the country-level as well as at a coarser granularity of continents. Jingling offers geo-targeting at a coarse granularity of China and Non-China, and city-level granularity within China. Others do not offer any geo-targeting.

4.3 Buying views from the exchange

The exchanges also sell traffic views directly, i.e., one can get traffic by paying, instead of by earning exchange credits via surfing/running their autosurf tool. The price per 100K views with a view-length of 15s ranges between $2–60 for the generic autosurf exchanges. Views on manual and social media promotion exchanges cost more; 100K views cost $450 on EasyHits4U, and $1,000 on HitSafari. Similarly, the cost for 100K 15s-long YouTube views on EnhanceViews Autowatcher is $375. (The exchanges sell views in bulk, with discounts for packages with higher views).

5. POPULAR TARGETS IN THE EXCHANGES

In order to develop insight into the popular domains in these exchanges, we captured a snapshot of the websites in circulation in the exchanges. We did so by *milking*: retrieving the instructions from the exchanges on the URLs to view, but not performing the action. In this section, we discuss our milking infrastructure, the

Exchange	Time Span	Milkers	Requested Page Impressions		
			Total	Unique URLs	Unique Domains
EasyHits4U	3 d	-	210	125	121
HitSafari	3 d	-	210	152	141
247AutoHits	30 d	25	0.7 M	3,632	2,195
10Khits	30 d	25	0.2 M	7,209	3,436
HitLeap	30 d	50	5.5 M	98,243	28,734
Jingling	30 d	25	104 M	337,829	61,727
eBesucher	30 d	25	0.6 M	7,323	1,431
Otohits	30 d	50	5.2 M	2,708	1,409
EnhanceViews Autowatcher	7 d	5	8.5 K	2,772	1

Table 2: Dataset obtained by milking the autosurf traffic exchanges, i.e., collecting the lists of URLs to be clicked but not sending clicks. We collected the data for manual exchanges manually, solving CAPTCHAs.

datasets we collected, and analysis on the popular categories of monetization observed in these exchanges.

5.1 Dataset

In all the autosurf exchanges we study, the communication with the exchange is unencrypted, enabling us to easily write our own milker bots that replicate this communication and receive instructions from the exchange. The instructions contain the URL(s) to be viewed, visit-length, and the referer to use in the view request (if offered by the exchange). We ran these milker bots from 25–50 different IP addresses per exchange via a proxy.

Using the above infrastructure, we collected month-long datasets for all autosurf exchanges, except for EnhanceViews Autowatcher. Since the latter only offers visits to a pre-defined list of social networks/user-content websites, we do not carry out the domains analysis for this exchange. Instead, we collected a week-long dataset for EnhanceViews Autowatcher's YouTube Watcher tool to compare the characteristics of video URLs in this exchange with those in the other generic exchanges we study. For the manual exchanges, which require a CAPTCHA to be solved between consecutive page views, we collected smaller datasets by manually surfing on the exchanges.

Table 2 describes the datasets we collected. For the manual exchanges, EasyHits4U and HitSafari, we solved 70 CAPTCHAs each. Each click resulted in the display of the main target page, as well as one display and one text ad. We also manually clicked on the display and text ads (see below), resulting in a total of 210 web page visits on each of the two exchanges. For autosurf exchanges, our milker bots recorded between 8.5K–104M page impression requests per exchange. The average rate of requests and the median view length varies across the exchanges, resulting in a difference in the scale of datasets.

We processed the raw URLs dataset to expand any URLs in the following set of popular URL shortening services: {bit.ly, goo.gl, ow.ly, t.co}.[1] We further processed the URLs to extract the "registered domains", according to the public suffix list maintained by Mozilla [18]. The numbers listed in Table 2 are for the processed dataset. It contains 2.7–337K unique URLs, and 1.4–61K unique domains per exchange. For the rest of the analysis, we work at the granularity of registered domains.

[1]We compiled this list based on their high frequency in our dataset, and unshortened the URLs belonging to them because the shortened versions hide the intended destination. We also observed URLs shortened using paid shortening services in our dataset, but did not unshorten them since they present an interesting monetization avenue, as we sketch later in § 5.3.1.

Category	EasyHits4U	HitSafari
Web traffic	38%	30%
Affiliates and marketing	35%	48%
Other products and services	27%	22%

Table 3: Distribution of website and ad categories we observed on EasyHits4U and HitSafari.

Ethical Considerations: We were cautious to not participate in view/click fraud when collecting our datasets. For the manual exchanges, we collected small-scale datasets by hand. We observed no indication of ad fraud. In addition, since during the process we ourselves viewed the websites circulating in the exchanges—as expected by the exchange and a website owner/advertiser—we view our actions as not contributing to view fraud.

For autosurf exchanges, we only collected the associated URLs, and did not view them in an automated fashion. Our milker bots emulated the timing patterns of requests from the autosurf tools—except for Jingling, for which we could request URLs at a rate five times greater than what a Jingling bot would do. (For other exchanges, requesting at a higher rate did not return any new URLs). We did so after determining, as best as we could from inspecting Jingling's website, that doing so did not violate any terms-of-service.

We note that none of our actions underlied intent to fraudulently earn credits, as prohibited by some of the exchanges in their terms-of-service, though such crediting did occasionally occur as a by-product of milking. The exchanges varied in their detection of milking: HitLeap imposed a penalty on our account after about an hour of milking. They continued to provide us with URLs to view, but we did not earn any credits for those URLs (as in fact, we were not viewing them). eBesucher temporarily blocked our accounts. Milking Jingling did not require accounts, hence we did not earn any credits. Others showed no indications of detection.

5.2 Analysis: Manual Exchanges

For manual exchanges, our datasets are small enough that categorizing them manually was viable. Table 3 summarizes the results.

The first category, web traffic, comprises traffic-related websites. This included other manual exchanges as well as sellers of web traffic and offers of cash for manual clicks (paid-to-promote sites).

The second category, affiliates and marketing, comprises websites that target persons attempting to earn money by selling over the Internet. Many ads offered commissions to affiliates for sale of various products and services. Some websites offered tools and infrastructure for affiliates such as email advertising. Others offered classes and manuals on how to become a successful Internet marketer. For example, one site offered a system for earning $10,000 per month as a Clickbank affiliate for a fee of $37 per month [6].

The large majority of the websites in these two categories appeared to be directly targeted at users of manual traffic exchanges, offering the prospect of cash in exchange for small fees or various types of simple online tasks. Websites of this type made up about three quarters of the total on both exchanges.

The third category, other products and services, contains ads for products and services outside the affiliate marketing and web traffic ecosystem itself. These ranged from web hosting services to legitimate niche products, such as self-published children's ebooks or special chocolate bars, to miracle weight loss programs and supplements, to formulas for large profits in currency and options trading,

Exchange	Top domains by views	Top domains by URLs
247AutoHits	*twistrix.com* (10%)	trck.me (1%)
	hitlink.com (6 %)	trafficadbar.com (1%)
	paragonmailer.com (4%)	wordpress.com (1%)
	paragontraffic.com (4%)	weebly.com (1%)
	vitality.ws (4%)	**adf.ly** (1%)
10Khits	kpopselca.com (10%)	apkmodder.com (3%)
	reportershub.com (2%)	**ziddu.com** (2%)
	ebay.co.uk (2%)	**youtube.com** (2%)
	hitleap.com (2%)	ebay.co.uk (2%)
	apkmodder.com (1%)	seoad.in (1%)
Jingling	**adf.ly** (6%)	**adf.ly** (10%)
	saringan.net (5%)	taobao.com (10%)
	wonosobo.in (4%)	baidu.com (8%)
	trifter.com (2%)	**ziddu.com** (5%)
	gameolosophy.com (2%)	user.qzone.qq.com (3%)
HitLeap	**youtube.com** (8%)	**ziddu.com** (20%)
	ziddu.com (4%)	**youtube.com** (11%)
	ads-host-media.com (2%)	**adfoc.us** (1%)
	ijgbiorjg.com (2%)	**adf.ly** (1%)
	4554fdd56f4.com (2%)	**sh.st** (1%)
eBesucher	mustbuysneakers.com (13%)	admg.info (3%)
	planetfem.com (10%)	dhush.com (3%)
	admg.info (6%)	trends-beauty.de (3%)
	genevaforever.com (6%)	softskills24.eu (2%)
	feltdizzy.com (4%)	job-market24.com (2%)
Otohits	health-spiritual.net (13%)	nomeimporta.es (1%)
	healthspiritual.net (13%)	herreiroimoveis.com.br (1%)
	zenbux.co (10%)	*tripleclicks.com* (1%)
	planete-carpe.com (4%)	weebly.com (1%)
	menintown.net (4%)	my-style.in (1%)

Table 4: Popular domains of the milked URLs in the autosurf traffic exchanges. Bold domains represent direct monetization avenues: the service offers pay-per-view for user content/sharing links. Italic domains are traffic exchanges; members promote exchanges in order to earn activity points, as top 25 active members get a share in the monthly cash pot (worth ≈ $10). Domains colored light-green are backed by a blogging service `triond` that pays users for views to their blog posts. Domains colored red do not have any real content and participate in ad placement fraud via *referer spoofing* on the AppNexus ad exchange.

to a manual on how to fuel cars with water. This category makes up about one quarter of the URLs on each exchange.

The target web sites were mostly free of external ads. We did not observe any indication of ad fraud.

5.3 Analysis: Autosurf Exchanges

For autosurf exchanges, we focus our analysis on the popular domains by views and URLs, since our datasets for this category are extensive. A domain can receive a high number of views from the exchange under three scenarios: (i) popular Internet service domains having a large user base (for example, YouTube), (ii) domains belonging to exchange participants, who are able to earn a large number of exchange credits using a number of machines, and in return direct huge volume of traffic to their own domains, or (iii) someone buying bulk views to a domain from the exchange. We can identify the first of these, but it is hard to differentiate between the latter two, and we treat these as equally interesting as potential targets for monetization/SEO/Alexa rank increase.

We characterize the top five domains by views and URL counts for each exchange in Table 4. We observe four categories of popular domains: (i) direct monetization avenues, (ii) ad fraud domains, (iii) legitimate looking domains, and (iv) traffic exchanges.

5.3.1 Direct Monetization Avenues

This category comprises domains that pay users for views to content they upload, such as YouTube. YouTube inserts video advertisements when playing user videos, and pays the user in proportion to the legitimate views their videos receive. In recent years, another category of services has emerged that does not require users to provide their own content in order to earn. Rather, users can get paid for linking and sharing existing content on the Internet. For example, URL shortening services such as `adf.ly` pay users for visits to their shortened URLs. A visitor to a URL shortened using such a service is first shown an advertisement, and then redirected to the URL that was shortened. Hence, money enters the eco-system from advertisers and flows to the users of the URL shortening services. We next discuss examples of both these monetization avenues as observed in the exchanges we studied.

The URL shortening services `adf.ly`, `adfoc.us`, and `sh.st` are among the top five domains in HitLeap, Jingling and 247AutoHits. Manually inspecting several requested `adf.ly` URLs reveals that 90% of the time the intermediate link solicits downloading a video player. It is possible that in addition to legitimate advertising, some of the malicious advertisers attempt to install malware on user machines. Our observation of malicious advertisements is consistent with the findings of Nikiforakis et al., who investigated malicious aspects of paid URL shortening services [32].

Another domain in this category popular on HitLeap, Jingling, and 10Khits is `ziddu.com`, a content distribution platform that pays for file downloads and sharing news links hosted on its platform [22]. Users with their own content can upload their files and earn for downloads of those files. An alternative (and lower effort) way to earn is to share existing content on Ziddu and direct views to it. Manually analyzing some of the `ziddu.com` URLs in our dataset, we observed that they contain news content lifted from other news websites under the title "Sponsored news", and generate impression requests to multiple ad networks. Further inspecting Ziddu's website, we found that it also offers desktop and mobile apps to *"turn idle cycles into money"*. The app displays advertisements (rather than websites) automatically, indicating that Ziddu apparently defrauds advertisers. In a manual investigation, we found the majority of the advertisements to be malicious, requesting to upgrade a media player or download a software. Ziddu does not list pay rates on its website.

We observe another model for the domains `trifter.com` and `gameolosophy.com`, popular in Jingling. These are niche sites dedicated to travel and online games respectively. The content hosted on these websites is not generated by the website owners. Instead, these websites are backed by `triond.com`, a blogging service, which automatically shares the content submitted by its bloggers to the most relevant site out of the pool of sites it supports. Triond then pays the blogger for views to their content.

YouTube receives the highest percentage of view requests (8%) in HitLeap, and has the second-highest percentage of URLs in this exchange. Note that YouTube *video watch* URLs[2] play videos automatically upon opening the link in a browser or an autosurf tool. Hence, directing exchange traffic to YouTube videos results in the videos being auto-watched. We investigated whether sending exchange traffic from HitLeap to YouTube videos results in increasing video view counts, and consequently provides a monetization avenue. We extracted the list of YouTube videos from our dataset (≈ 10K unique videos) and queried the YouTube API to get their view counts. Per Figure 1, approximately 24% of the videos in circulation on the HitLeap exchange were no longer available on YouTube at the time of querying; they were either taken down by YouTube or the user themselves. 25% had a view count frozen at

[2] These URLs have the form: http://www.youtube.com/watch?v=VIDEOID.

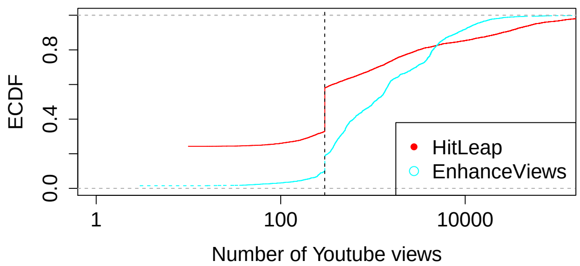

Figure 1: Distribution of view counts of the YouTube video URLs obtained by milking traffic exchanges. We obtained the view counts of the videos by querying the YouTube API provided by Google. The size of the discontinuity at 301 views indicates the fraction of videos stuck at that view count. Note that about 25% of HitLeap but only 10% of EnhanceViews Autowatcher videos are stuck there. The discontinuity at 301 views suggests that YouTube's fraud detection prevents many videos from advancing beyond this point, even if they receive many more clicks.

301 (indicating that YouTube had identified something suspicious with the views [13]). However, ≈ 42% of the videos on HitLeap had view counts above 301, with a maximum of 18 million views.

Recall from Table 1 that HitLeap views are short-lived (60s at maximum), which possibly could be a factor in the detection of its fraudulent traffic. For comparison, we look at the view counts of YouTube videos in the EnhanceViews Autowatcher exchange, which plays the full video for each view. We observe in Figure 1 that EnhanceViews Autowatcher does better than HitLeap, with a frozen view count for only 10% of the videos.

We also queried the YouTube API to get the channel ID and the categories of the videos in our HitLeap dataset. The 10K videos were associated with ≈ 2,500 channels—8% of the videos came from a single channel. The videos in this dominant channel belonged to the categories *Gaming*, *Shows*, and *Film/Animation*, with none of the videos stuck at a view count of 301. Inspecting this channel, we found that the latest video gathered 25K views in one day, though we lack visibility into what fraction of these were fraudulently generated. Generally, we find that 26% of the videos in our dataset belong to the *Gaming* category and 23% in the category *People and Blogs*, with the former mainly screen-captures of people playing video games, and the latter primarily comprising tutorials.

5.3.2 Ad Placement Fraud

We also observed websites that show clear evidence of impression/click fraud. For example, *ads-host-media.com*, *ijgbiorjg.com*, and *4554fdd56f4.com* in HitLeap do not display any content to a visiting user.[3] But these sites generate dozens of ad requests with varying referrers to the AppNexus ad exchange [4]. The URL requests are of the format: `ib.adnxs.com/tt?ttjb=1&bdc=1&bdh=CQLrezBk2vGyQNA7G5993lyZhww.&id=3445029&size=300x250&referrer=carpreserve.com`, with varying referrers sent on each request.

Others have documented this *referrer* spoofing fraud [35, 25]. Stone-Gross et al. observed it in the context of the Yahoo Right Media exchange in their study of fraudulent activities in online ad exchanges. The convoluted online advertising ecosystem fosters this kind of fraud. In particular, the exchange/advertisers take the referrer in the URL parameters at face value. Although the impression requests to an ad exchange contain an `id` value that determines which party gets credited when an ad is displayed on the site in the *referrer* field, the need for reaching broad audiences in online advertising complicates its use for fraud detection. Since impressions can be bought through an intermediary (rather than directly from publishers) to enable wider reach, an ad exchange cannot employ an exclusive website-owner-specific `id` to be credited on impression requests for the particular website [25].

[3] We confirmed that we had not been cloaked against by visiting these domains in the autosurf tool in addition to the browser; we configured a proxy to allow the autosurf tool to visit these domains for a short amount of time.

Exchange	URLs	Impressions
HitLeap	2,606 (2.7%)	1,231,027 (23%)
Jingling	2,571 (0.8%)	10,426,448 (11%)
eBesucher	230 (3.1%)	59,410 (9%)
Otohits	50 (1.9%)	336,634 (6%)
247AutoHits	27 (0.7%)	5,398 (1%)
10Khits	28 (0.4%)	494 (0.2%)

Table 5: Percentage of milked URLs and corresponding impressions served by sites exhibiting referrer-spoofing ad placement fraud in various exchanges.

Given that we observed *referrer spoofing* fraud as one of the major categories in multiple exchanges, we fingerprinted this behavior in order to classify the number of URLs in our dataset that participate in this fraud. We crawled the entire datasets using a distributed infrastructure, recording all the web requests that each URL generates.[4] We then identified the URLs that generate requests to `adnxs.com` with *spoofed* referrers. We identified *spoofed* referrers as those that are for a domain other than on which the ad is actually displayed. Table 5 summarizes our findings. We find ≈ 2.5K URLs in HitLeap and Jingling that participate in this ad fraud, respectively accounting for 23% and 11% of the total impression requests we milked for these exchanges. Other exchanges have fewer URLs, accounting for 1–9% of their impression requests.

5.3.3 Legitimate looking domains

We find some of the other websites harder to judge. While essentially all of them bear advertisements, some of these appear to be niche websites targeted at topics such as cars, sneakers, workout, and news specific to a particular location.

The most popular domain in eBesucher, `mustbuysneakers.com`, is a legitimate-looking niche website, with an '*Advertising: Starboyant Media*' tag in the website footer. Investigating the Starboyant Media website, we found that it offers advertising on three other niche websites: `onlyhiphop.com`, `workouteveryday.com`, and `stylishceleb.com`. We find all of these in our eBesucher dataset, accounting for 22% of the total impression requests we milked on eBesucher.

We find ourselves unable to judge whether these and other legitimate-looking domains are websites with considerable legitimate traffic to which a fraudster also directs traffic, or they have been crafted to make judgment difficult and have little traffic beyond that automatically generated.

5.3.4 Traffic Exchanges

Surprisingly, most of the top domains in the basic autosurf exchange 247AutoHits are other traffic exchanges. On closer inspec-

[4] We opened each website in a browser instance until the browser indicated that the page had finished loading.

tion, we find that these are *splash pages* provided to the users by an exchange itself for circulation in other exchanges. The four exchanges that appear in the top domains—*twistrix.com*, *hitlink.com*, *paragonmailer.com*, and *paragontraffic.com*—were not included in our study. Looking through their websites, we find that these exchanges offer a share in a *cash pot* (worth $\approx$ \$10) at the end of every month to the top 25 active users. Submitting the member-specific exchange promotion link on other exchanges is one of the ways in which members can earn activity points towards a share in the *cash pot*.

6. TRAFFIC QUALITY & SIZE OF EXCHANGE NETWORKS

In this section we study the quality of traffic that the exchange networks deliver, and measure the scale and characteristics of the IP address population that participates in these exchanges. We base our observations on buying bulk views from the exchanges for five consecutive days and recording the characteristics of the visitors.

Infrastructure: We set up eight websites, one per exchange, on Amazon EC2, configured to auto-scale in order to handle large peaks of traffic. The websites themselves were fairly simplistic, and identical in content, appearing to be a photo blog. We employed Javascript instrumentation on the websites to detect (i) mouse movement, and (ii) whether at the time of visit the visitor was logged into any of Google, Twitter, or Facebook.[5] While not perfect, this instrumentation gives us a sense of organic (real) users participating in these exchanges. Where possible, we requested views for a view-length of 30s or greater, in order to allow enough time for the instrumentation Javascript to execute. We also set cookies to study the behavior of returning visitors.

We signed up for two premium accounts on each exchange, and purchased views from the exchanges alternating between the accounts every other day. Premium accounts gave us access to more website slots on the exchanges, enabling us to submit multiple pages of our websites in order to get bulk traffic. Making purchases from two accounts arguably enables us to collect *independent* samples of the IP addresses in an exchange.

Purchases: Table 6 lists the breakdown of traffic we bought from each exchange. For each exchange, we started with a pilot measurement of 100K visitors per day for autosurf and 10K visitors per day for manual exchanges,[6] and scaled down the number of visitors in subsequent purchases, requesting at least twice as many visitors as the number of unique IP addresses observed in the pilot measurement.

Purchase Order Completion: For each purchase, we requested visitors distributed across multiple pages of our corresponding website (at a maximum rate of 250 visitors per hour per page),[7] so that the expected completion time for the purchase order was one day. The exchanges varied in the completion of purchase order. HitSafari, 247AutoHits, and 10Khits on average fulfilled only 50/63/68% of the order, and charged us only for what they were able to fulfill. Our measurements indicate that these three are the smallest of the eight exchanges we measured (on the order of a thousand IP addresses), explaining the inability to fulfill larger orders within a day. Others claimed to fulfill the complete orders, but had some losses—the instrumentation reported by the exchange did not match that recorded by our web server logs. Specifically, Jingling fulfilled only 61% of the order on average; others had a loss in the range 2–12%.

Exchange / IP Addresses	Popular ASNs
eBesucher (33,749)	**Google, US (15%)**
	Hetzner Online AG, DE (5%)
	Scana, UA (5%)
Jingling (110,641)	CHINANET-BACKBONE No.31, CN (26%)
	Amazon, US (12%)
	China169 Backbone, CN (10%)
	Google, US (7%)
	HINET, TW (7%)
HitLeap (72,135)	**Amazon, US (35%)**
	Google, US (19%)
	OVH SAS, FR (10%)
Otohits (3,931)	**Amazon, US (26%)**
	Google, US (24%)

Table 7: ASNs having at least 5% representation in the exchanges. We highlight major cloud hosting providers in bold. A significant percentage of the IP addresses in the advanced autosurf exchanges comes from Google and Amazon.

6.1 Quality of delivered views

The views were delivered from a variety of User-Agents by default, on the order of several hundred for most of the exchanges. Most notably, Jingling delivered traffic from $\approx$ 33K different user-agents (dominated by different versions of Internet Explorer). Since we did not request a custom `Referer`, the `Referer` we received was the link of the traffic exchange for all the exchanges except Jingling, for which it was not present.

The page components (stylesheet, Javascript) hosted on the same domain were fetched for 92–100% of IP addresses of the different exchanges. An image embedded from a different domain was fetched for 86–100% of IP addresses for all exchanges except Jingling, for which only 1% of the IP addresses fetched the image.

We found that Javascript was executed for 83–97% of the IP addresses on six of the exchanges. On Jingling, the number is much lower (37%)—this could be due to high load on the machines that run Jingling, because it initiates multiple threads and views multiple websites in parallel. For 10Khits, the Javascript instrumentation fired for only 20% of the IP addresses. Further, upon investigation, our web logs indicated that for the IP addresses where it did fire, the POST request paths were malformed, hence no data was logged in our database for the mouse movement and social-network login instrumentation. This is presumably because at the time of our data collection, the 10Khits website indicated it was in a transitional phase, soon to launch their own autosurf tool. Perhaps the hits to our websites were delivered by a buggy version of the tool.

6.2 Characterization of View Sources/IP addresses

Over the five days, we received visits to our websites from $\approx$ 204K unique IP addresses across the eight exchanges, corresponding to $\approx$ 93K unique /24 network prefixes. 77–95% of the /24 prefixes have only one IP address, suggesting that a significant fraction of these IP addresses belong to individual machines (or NATs), with little IP aliasing.

6.2.1 Manual Exchanges

We recorded 1K IP addresses on HitSafari from 73 countries, and 6K on EasyHits4U from 128 countries. Of these, 40% on HitSafari

[5] We used the methodology described at http://www.tomanthony.co.uk/blog/detect-visitor-social-networks/.

[6] We made smaller purchases from manual exchanges due to higher cost of manual traffic.

[7] The exchanges set upper bounds on the number of visits per URL per hour, e.g., 500 for HitLeap and 1,000 for eBesucher. We requested a limit lower than the upper bound based on a pilot run that found the upper bound did not get delivered in a timely fashion.

Exchange	Purchase Order (Per Day)				Avg. Order Fulfilled	Exchange Size & Characteristics Of IP Addresses (Measured over five days)						
	Views	Cost	Duration	CPM		IP Addresses	User Agents	Page Component Fetches		Javascript Executed	Social Network Logins	Mouse Movement
								Same Domain	Other Domain			
HitSafari	2.5K	$25	6s	$10	50%	1,018	179	92%	86%	84%	40%	10%
EasyHits4U	10K	$75	30s	$7.5	98%	6,191	699	99%	96%	96%	65%	21%
10Khits	10K	$10	30s	$1	68%	1,578	260	99%	95%	20%	–	–
247AutoHits	10K	$7	30s	$0.7	63%	1,420	219	95%	85%	83%	43%	6%
Otohits	50K	$3	30s	$0.06	98%	3,931	336	100%	97%	97%	13%	4%
eBesucher	100K	$17	30s	$0.17	88%	33,749	1,752	96%	95%	95%	12%	3%
HitLeap	100K	$44	30s	$0.44	97%	72,135	871	100%	100%	97%	–	0.1%
Jingling	100K	$41	40s	$0.41	61%	110,641	33,222	99%	1%	37%	0.1%	–

Table 6: Our purchases from various exchanges to study the quality of traffic. We purchased traffic in varying quantities, and recorded the HTTP requests, page rendering, and organic-visitor instrumentation over five days.

and 65% on EasyHits4U had an account logged into a social network, indicating that a significant fraction of views came from real users (or compromised machines). Our instrumentation recorded mouse movements for 10% and 21% of IP addresses on HitSafari and EasyHits4U respectively. The low percentage is presumably because the users do the bare minimum required to get paid.

Finding these organic visitor indicators, we next investigate whether this traffic comes from real users in low-wage countries who click intentionally in the hope of earning money. We looked at the top ASNs, and find that the traffic mostly comes from residential networks, with no single dominant contributor. Surprisingly, geo-location (using the MaxMind database) indicates that 52% of the IP addresses in EasyHits4U come from Europe, with Asia the second major contributor (25%). The IP addresses from Europe are dominated by Russia, Romania, and Ukraine with each having 5–6% representation. IP addresses from Asia are dominated by India and Indonesia. In HitSafari, North America has a 40% contribution, followed by Europe (31%), and Asia (22%). We find that the country distribution in HitSafari differs from that in EasyHits4U, with UK and Germany having the top representation (but only 4–5% each) in Europe, and China and Singapore dominating IP addresses from Asia.

In light of this geographic diversity and the significant contributions from North America and Western Europe, click farms in low-wage countries do not appear to be the principal source of clicks on manual exchanges.

6.2.2 Autosurf Exchanges

Our measurements find that the basic autosurf exchanges, 247AutoHits and 10Khits, are small in size, on the order of 1.5K IP addresses each. However, in the advanced autosurf category Otohits has $\approx$ 4K IP addresses, while the other three are an order of magnitude larger.

Geolocation: We received hits from 80–147 different countries across the various autosurf exchanges. The spread over countries for the smaller networks differed from that of the larger ones: eBesucher is highly concentrated (80%) in Europe, with Russia, Ukraine and Germany predominant. Jingling has the highest representation in Asia (70%), with China contributing 56% of the IP addresses. The second major contributor in both eBesucher and Jingling is North America, with 15% and 28% representation respectively. HitLeap's network is predominantly based in North America (63%), with 60% of IP addresses from the US. 20% of HitLeap's IP address space is based in Europe, with Italy and Germany the top contributors (but only 4–5% IP addresses each). Similar to HitLeap, Otohits has representation in North America (47%) and Europe (26%), with the US predominant (45%), and Taiwan, Vietnam and Germany each at 5%.

For the smaller networks, 247AutoHits and 10Khits, Asia and Europe have 74% representation. The country distribution is dispersed, with Vietnam, Indonesia, India, Germany, and the UK as the popular countries (4–8% representation).

Organic Visitors: The traffic recorded from 247AutoHits had the highest percentage of visitors from machines where a user was logged in to a social network (43%), but only 6% indicated a mouse movement. For the advanced autosurf exchanges, it is possible that the user uses a browser different than that used by the autosurf tool. Hence we treat the estimates we discuss below as a lower bound. We find that on average $\approx$ 12% of the visitors from eBesucher, and $\approx$ 13% on Otohits, were from machines where a user was logged in to a social network.[8] In contrast, the number is much lower for Jingling; only 0.1% of the visitors appear organic. We recorded mouse movement for only 1–4% of IP addresses on these exchanges.

We could not carry out this analysis for HitLeap and 10Khits, because (i) HitLeap's autosurf tool uses a built-in browser rather than a browser on the user's machine, and thus none of the visitors appeared to have logged in, when in fact they might have been, and (ii) as described in Section 6.1, the 10Khits surf tool had malformed URL request paths, due to which we did not log any data in our database.

Popular ASNs: We next take a look at the popular entities/organizations to which the IP addresses of exchange networks belong. We obtain the ASNs of the IP addresses via Team Cymru's IP-to-ASN mappings database [36]. Table 7 lists the popular ASNs (those having at least a 5% representation), and the corresponding percentage of IP addresses observed for each exchange. We find that a striking percentage of the IP addresses in the HitLeap and Jingling exchanges belong to Amazon, totaling $\approx$ 38K IP addresses. Similarly, a significant portion of IP addresses in the autosurf exchanges comes from Google's cloud, totaling $\approx$ 27K IP addresses.

We also observe other smaller cloud providers and Internet hosting/data centers in the top five ASNs. For example, the second-highest contribution in eBesucher's network comes from Hetzner Online (5%), an Internet hosting provider and data center operator based in Germany. Similarly, HitLeap addresses also appear from other smaller cloud providers such as Limestone Networks [17] and OVH [21].

6.2.3 Overlap across exchanges

We find very little overlap across the IP address space of the manual and autosurf exchanges. However, within these categories, we observe significant overlaps. 17% of HitSafari's IP address space

[8] Conceivably, there is a small chance that there is a fraudster running social media promotion tools in parallel; we do not have a method for verifying whether the user on a machine is genuine.

Exchange	% Overlap with other exchanges	Top overlap contributors ($\geq$ 5% contribution)
HitSafari	17%	EasyHits4U
EasyHits4U	9%	–
10Khits	43%	HitLeap, 247AutoHits, eBesucher
247AutoHits	43%	HitLeap, 10Khits, eBesucher
Otohits	74%	HitLeap, Jingling
eBesucher	8%	HitLeap
HitLeap	31%	Jingling
Jingling	18%	HitLeap

Table 8: Overlap of IP address space across various exchanges.

Exchange	IP addresses		Impressions per bot	Estimated total impressions
HitSafari	329	(σ=94)	–	–
EasyHits4U	1,786	(σ=414)	–	–
10Khits	483	(σ=109)	359	173 K
247AutoHits	507	(σ=376)	1,420	719 K
Otohits	2,178	(σ=378)	3,654	7.9 M
eBesucher	18,164	(σ=707)	1,066	19 M
HitLeap	41,711	(σ=507)	3,607	150 M
Jingling	40,227	(σ=5,200)	19,630	789 M

Table 9: Impression estimates per day. σ represents standard deviation across five days of measurements.

overlaps with that of EasyHits4U. 43–74% of the address space of smaller autosurf networks is also seen in other exchanges. For the larger networks, 31% of HitLeap's space overlaps with that of Jingling.

6.3 Impressions Per Day Estimates

In this section we develop a handle on the number of impressions these exchanges serve per day. We base our estimates on the daily number of IP addresses recorded in our logs. However, we note that two factors can lead to overestimating/underestimating these numbers: (i) IP address aliasing (the same machine switching IP addresses), and (ii) multiple machines behind NATs.

We analyzed cookie data recorded in our web logs to potentially adjust for these two factors. For multiple visits from the same IP address, we find that 85–95% of IP addresses in manual exchanges, 80–85% in basic autosurf, and 20–45% in advanced autosurf returned the same cookie set on the first visit. Others request and return multiple cookies. Investigating this further, we find that a high fraction of the IP addresses in advanced autosurf exchanges clear cookies after every visit.

Due to the limitation of the cookie data, we report our population estimates at the granularity of IP addresses. We use these population sizes to estimate the total number of impressions the exchanges serve per day. To this end, we compute the average number of impression requests that our milker bots received on average per day,[9] and multiply it by the population size of the exchange. Table 9 lists the impressions per day estimates. On the lower end, we estimate that 10Khits serves 173K impressions per day, and on the higher end Jingling serves on the order of 789 M impressions per day.

7. DISCUSSION

In the previous sections we measured a number of facets of traffic exchanges. In this section, we attempt to draw plausible inferences based on what we observed.

7.1 Selection pressure

We find evidence of traffic fraud responding to selection pressure. The very existence of exchanges, a tool to disperse the originating IP addresses of traffic, indicates that advertisers actively look for fraud; clearly, clicking endlessly from a single IP address seldom achieves the desired result. The differences between the free and premium versions of the advanced autosurf tools (shown in Table 1) largely reflect the ability to target certain geographic locations, and to set custom User-Agent and Referer fields. There would be little reason for exchange participants to pay for this functionality if it did not bring meaningful advantage. It seems likely that these fields are used by sites to detect fraudulent traffic, and the ability to vary them will increase profits. The large fraction of YouTube videos stuck at 301 credited views in Figure 1 strongly suggests a fraud-detection barrier that blocks a large fraction of the YouTube-bound traffic from two autosurf networks.

[9]Since we milked Jingling at a higher rate, we use the number obtained from running the original bot through the Fiddler proxy, configured to only allow communication with the exchange.

7.2 Manual vs. Autosurf exchanges

A common property of the exchanges is that they facilitate dispersion of traffic; indeed the need for dispersion is likely the factor that has led to their popularity. Beyond this common point, we found little similarity between the manual and autosurf exchanges. Table 4 shows that the most popular destinations for traffic from the autosurf exchanges have ad-bearing properties, and ad-fraud of some kind dominates their traffic. By contrast, Table 3 shows that traffic from the manual exchanges is predominantly directed at sites that advertise affiliate programs, other manual traffic exchanges, or pages advertising goods sold by the participants themselves. In other words, pages visited in the manual exchanges appear directed at luring humans (the exchange participants) to upgrade, join other exchanges or affiliate programs, or purchase products that do not have mainstream appeal. Our inspection of manual exchange traffic turned up numerous examples of what appear to be implausible propositions and get-rich-quick schemes. For example, one site asks $9.99 for a pamphlet explaining how to "run your car on water." A number of destinations of the manual exchange traffic appear to be Ponzi or multi-level marketing schemes. For example, at http://imarketingfasttrack.com/ new entrants invest $25 and recruit four newcomers who do similarly; while participants are promised an easy $1,700, there is no mention of any product beyond participant recruitment. Another, http://www.theleadmagnet.com/, is a viral email lead-generation tool that offers commissions if members get others to purchase the tool.

7.3 Monetization

Monetization clearly poses a significant problem even if one can generate volumes of well-dispersed traffic. The direct cash-out payments offered by the exchanges are low: $0.30 per thousand manual clicks for EasyHits4U, and $0.02 per thousand automated clicks at eBesucher (though getting paid requires a photo ID, as noted in Section 4.1). Note that, based on Table 9, this amounts to 2 cents per machine per day, or $7.30 per machine per year.

For fraudsters, exploiting sites such as YouTube, which offer to pay for traffic delivered, represents an appealing approach. This has the benefit of low start-up costs: the fraudster simply sets up a channel with some content to which they direct the traffic, avoiding costs associated with registering a domain and hosting. Moreover, the content does not have to be of a quality sufficient to attract actual human viewers. If the fraud is discovered and the channel blocked, one can simply start over by setting up a new channel.

Figure 1 suggests that services such as YouTube successfully detect a good percentage of inorganic traffic, making it hard to linearly translate automatically generated views to money. Perhaps participants direct traffic to YouTube primarily for increasing view counts to gain popularity, rather than monetization.

7.4 Lack of sophistication

Some of our data supports the view that many traffic exchange participants lack technical sophistication and struggle to monetize their efforts, particularly on the manual exchanges. At first sight, the manual exchanges operate as web traffic brokers, offering free traffic/advertising in exchange for clicks. The exchange extracts money from participants via the exchange ratio (i.e., exchanging clicks at less than a 1–1 ratio so that the excess can be sold) or monthly fee. But this explanation does not seem satisfactory. A participant who seeks traffic for their product receives clicks only from others similarly motivated to sell their products. Since the average quality of clicks received is the same as those delivered, the average participant should lose in expectation once the exchange's cut is removed, unless outside money enters the exchange. It is also hard to believe that cash-out rates of \$0.30 per 1,000 clicks (about \$0.05 per hour of clicking) could be a long-term incentive for participants.

The emphasis on affiliate programs and related tools and services among the exchange ads points to a different explanation. These ads appear targeted at an unsophisticated user group lured by the prospect of easy money and a career in "Internet marketing". This description may well be representative of a large fraction of the actual manual exchange users. In this case, the sites advertising on the exchanges might be receiving quality traffic for their offers with acceptable conversion rates, even if they buy traffic from the exchange rather than clicking.

7.5 Revenue estimates

Table 6 gives the prices at which we bought clicks from various exchanges. The cost per thousand (CPM) ranges from \$0.06 at Otohits to \$10 at HitSafari. Obviously, the clicks are not all of the same quality, and some may suit far better than others for a particular purpose. For comparison, we compute the cost of generating clicks if one rented the cheapest instance of a virtual machine from a cloud hosting service. A micro-instance from Amazon AWS costs \$0.013 per hour [3]; setting such a machine to click twice per minute would cost \$$0.013 \times 1000/(2 \times 60) \approx \$\ 0.11$ per thousand. Of course these clicks would offer no IP address dispersion. However, these clicks can be dispersed by using the free service of one of the autosurf exchanges. Thus, we calculate that dispersed clicks can be produced at about \$0.11 per thousand. If one needed the customization that the premium services offer, then at 2.9€≈ \$3.5 per month the cost becomes \$ $(0.013 \times 24 \times 30 + 3.5) \times 1000/(30 \times 24 \times 60 \times 2) \approx \$\ 0.15$ per thousand (i.e., renting the instance for a month at the hourly rate plus the cost of premium exchange membership). Note that this is significantly lower than several of the auto exchanges charge; e.g., HitLeap charges \$0.44 and Jingling \$0.41 per thousand. The fact that dispersed customized clicks can be produced at significantly lower cost than the exchanges charge suggests an inefficient market. It could be that the exchanges price optimistically and sell very little at the advertised rates, or it could be that there are unsophisticated buyers for whom the task of setting up an AWS instance represents a barrier. Note that the manual exchanges ask more than $10\times$ more per click. A possible explanation is that proposed in Sections 7.2 and 7.4: it is the participants themselves rather than an ad network who are the targets, and sending Ponzi and get-rich-quick schemes at this population is more profitable on a per-click basis than sending inorganic traffic to YouTube. Our measurements (and the requirement to perform a CAPTCHA between clicks) suggest that a large fraction of traffic on the manual exchanges is in fact organic.

It is hard to pin down definitive stable numbers for how much one can expect to earn from inorganic traffic. Some URL shortening sites post their current payout rates: `adf.ly` pays \$3.96, `adfoc.us` pays \$5.50, and `sh.st` pays \$4.03 per thousand. YouTube does not post payout rates, and discussion forums seem to indicate it is a complex function of channel subscribers and number of "likes" that a video receives; between \$0.50 and \$4 per thousand appears to be a commonly mentioned range. (The need for "likes" and subscribers to help monetize views suggests a market for those commodities also).

If clicks can be produced at \$0.15 per thousand (or bought at the rates shown in Table 6) and sold for \$4 or so per thousand to URL shortening services, the opportunity is very profitable. At the two clicks per minute rate this would produce \$ $(4.03 - -0.15) \times 30 \times 24 \times 60 \times 2/1000 \approx \$\ 335.2$ profit per machine per month. It seems very likely however that URL shortening services detect (and refuse to credit) a large volume of inorganic traffic (e.g., in its TOU `adf.ly` forbids publishers from "advertising their adf.ly URL links directly on any form of traffic exchange/PTC website", and other services use similar wording). Equally, even the lower of YouTube payout rates would appear profitable if detection has negligible effect. Ethical considerations prevent us from sending automated traffic to any of these services to probe their detection rates. It is thus difficult to estimate the true return on inorganic traffic. We note, however, that even if only 5% of the inorganic traffic is credited ($\approx$ \$17 per machine per month), the ability to scale up by adding more machines would still allow healthy returns.

Table 9 gives our estimate of the number of impressions per day from each of the autosurf exchanges. Again, we caution that the fraction of exchange traffic that is detected (and thus not credited) is unknown. Nonetheless it is interesting to get rough ideas of the economic damage inorganic traffic might be causing. HitLeap, at 150M impressions per day, would cost \$66K per day to advertisers *if all exchange participants managed to monetize at \$0.44 per thousand impressions* (i.e., the rate the exchange charges for clicks as shown in Table 6). Similarly, the figures for eBesucher and Jingling would be \$3,230 and \$323K respectively. These can be considered somewhat loose upper bounds.

8. CONCLUSION

We examined several manual and automated traffic exchange services, which enable dispersion of traffic origins. We found that these exchanges range across three orders of magnitude in diversity of IP addresses and impressions served per day, and appear to differ significantly in the monetization strategies of participants. Our results shed light on a previously poorly understood part of the traffic and ad-fraud ecosystem.

Acknowledgments

Our thanks to Ben Edelman, Paul England, Mariusz Jakubowski, and Gloria Mainar-Ruiz for discussions on various aspects of this work.

This work was supported by the U.S. Army Research Office under MURI grant W911NF-09-1-0553, and by the National Science Foundation under grant CNS-1237265. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

9. REFERENCES

[1] 10khits. http://www.10khits.com. Online. Mar, 2015.

[2] 247autohits. http://www.247autohits.com. Online. Mar, 2015.

[3] Amazon ec2 pricing. http://aws.amazon.com/ec2/pricing/. Online. Mar, 2015.

[4] Appnexus. http://www.appnexus.com. Online. Mar, 2015.

[5] Black hat world. http://www.blackhatworld.com. Online. Mar, 2015.

[6] Clickbank. http://www.clickbank.com. Online. Mar, 2015.

[7] Common crawl. http://commoncrawl.org/the-data/. Online. Mar, 2015.

[8] Cyber criminals defraud display advertisers with tdss. http://goo.gl/e1dcsj. Mar, 2015.

[9] Discovered: Botnet costing display advertisers over six million dollars per month. http://www.spider.io/blog/2013/03/chameleon-botnet/. Online. Mar, 2015.

[10] Easyhits4u. http://www.easyhits4u.com. Online. Mar, 2015.

[11] Ebesucher. http://www.ebesucher.com. Online. Mar, 2015.

[12] Enhanceviews autowatcher. http://www.enhanceviews.com. Online. Mar, 2015.

[13] Frozen view count - youtube. https://support.google.com/youtube/troubleshooter/2991876?hl=en-GB/. Online. Mar, 2015.

[14] Hitleap. https://hitleap.com. Online. Mar, 2015.

[15] Hitsafari. http://www.hitsafari.com. Online. Mar, 2015.

[16] Jingling. http://service.spiritsoft.cn. Online. Mar, 2015.

[17] Limestone networks. https://www.limestonenetworks.com. Online. Mar, 2015.

[18] Mozilla public suffix list. http://wiki.mozilla.org/Public_Suffix_List. Online. Mar, 2015.

[19] Mustat. http://www.mustat.com. Online. Mar, 2015.

[20] Otohits. http://www.otohits.net. Online. Mar, 2015.

[21] Ovh - webhosting, cloud and dedicated servers. https://www.ovh.com/. Online. Mar, 2015.

[22] Ziddu. http://www.ziddu.com. Online. Mar, 2015.

[23] N. Christin. Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd international conference on World Wide Web*, pages 213–224. International World Wide Web Conferences Steering Committee, 2013.

[24] V. Dave, S. Guha, and Y. Zhang. Measuring and fingerprinting click-spam in ad networks. In *Proceedings of the Special Interest Group on Data Communication (SIGCOMM)*, August 2012.

[25] B. G. Edelman. Accountable? the problems and solutions of online ad optimization. In *IEEE Security & Privacy 12, no. 6 (November–December 2014): 102–107.*

[26] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: The underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security*, pages 27–37. ACM, 2010.

[27] H. Haddadi. Fighting online click-fraud using bluff ads. *SIGCOMM Comput. Commun. Rev.*, 40(2):21–25, Apr. 2010.

[28] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamalytics: An empirical analysis of spam marketing conversion. In *Proceedings of the 15th ACM conference on Computer and communications security*, pages 3–14. ACM, 2008.

[29] R. A. Lewis and J. M. Rao. On the near impossibility of measuring the returns to advertising. *Unpublished paper, Google, Inc. and Microsoft Research. http://justinmrao.com/lewis_rao_nearimpossibility.pdf*, 2013.

[30] D. McCoy, A. Pitsillidis, G. Jordan, N. Weaver, C. Kreibich, B. Krebs, G. M. Voelker, S. Savage, and K. Levchenko. Pharmaleaks: Understanding the business of online pharmaceutical affiliate programs. In *Proceedings of the 21st USENIX conference on Security symposium*. USENIX Association, 2012.

[31] B. Miller, P. Pearce, C. Grier, C. Kreibich, and V. Paxson. 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*, DIMVA, pages 164–183. Springer-Verlag, 2011.

[32] N. Nikiforakis, F. Maggi, G. Stringhini, M. Z. Rafique, W. Joosen, C. Kruegel, F. Piessens, G. Vigna, and S. Zanero. Stranger danger: Exploring the ecosystem of ad-based url shortening services. In *Proceedings of the 23rd international conference on World wide web*, pages 51–62. International World Wide Web Conferences Steering Committee, 2014.

[33] P. Pearce, V. Dave, C. Grier, K. Levchenko, S. Guha, D. McCoy, V. Paxson, S. Savage, and G. M. Voelker. Characterizing large-scale click fraud in zeroaccess. In *Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS)*. ACM – Association for Computing Machinery, November 2014.

[34] K. Springborn and P. Barford. Impression fraud in on-line advertising via pay-per-view networks. In *Proceedings of the 22nd USENIX Security Symposium*, pages 211–226, Washington, D.C., 2013. USENIX.

[35] B. Stone-Gross, R. Stevens, A. Zarras, R. Kemmerer, C. Kruegel, and G. Vigna. Understanding fraudulent activities in online ad exchanges. In *Proceedings of the eleventh ACM SIGCOMM Conference on Internet Measurement Conference*, pages 279–294, New York, NY, USA, 2011. ACM.

[36] TeamCymru. IP to ASN Mapping. http://tinyurl.com/5dtp78. Online. Mar, 2015.

[37] Q. Zhang, T. Ristenpart, S. Savage, and G. M. Voelker. Got traffic?: An evaluation of click traffic providers. In *Proceedings of the Joint WICOW/AIRWeb Workshop on Web Quality*, pages 19–26, New York, NY, USA, 2011. ACM.