Dirty Jobs: The Role of Freelance Labor in Web Service Abuse

Marti Motoyama Damon McCoy Kirill Levchenko Stefan Savage Geoffrey M. Voelker

Department of Computer Science and Engineering
University of California, San Diego

Abstract

Modern Web services inevitably engender abuse, as attackers find ways to exploit a service and its user base. However, while defending against such abuse is generally considered a technical endeavor, we argue that there is an increasing role played by human labor markets. Using over seven years of data from the popular crowd-sourcing site Freelancer.com, as well data from our own active job solicitations, we characterize the labor market involved in service abuse. We identify the largest classes of abuse work, including account creation, social networking link generation and search engine optimization support, and characterize how pricing and demand have evolved in supporting this activity.

1 Introduction

Today's online Web services—search engines, social networks, and the like—create value for their users by helping them find and interact with content generated by other users. While these services typically rely on advertising for their revenue, their open access and reliance on user-generated content create powerful opportunities for abusers to fabricate secondary, extremely cheap advertising channels as well. The result is well-known: Web-mail spam, polluted search results, "friend" requests from fake persons and so on. These activities are broadly termed *service abuse*: they exploit some feature of a public service for an attacker's financial gain at the expense of the service provider.

Each Web service provider aims to prevent such activities and preserve the value of their advertising enterprise. To that end, most Web sites include extensive contracts declaring limits on the way their services may be used. However, implicit threats of legal action rarely deter attackers, and so the provider must rely on a broad range of defenses and countermeasures to enforce their terms of service. While the technical details of this "arms race" are themselves interesting, they are ultimately just symptoms of this larger struggle over controlling who may monetize access to a site's users.

Thus, in this paper we do not focus deeply on the underlying technical attacks themselves, but rather explore the human labor markets in which these capabilities are provided. Though not widely appreciated, today there are vibrant markets for such abuse-oriented services and in a matter of minutes, one can buy a thousand phone-verified Gmail accounts for $300 or a thousand Facebook "friends" for $26. Much of this activity occurs on freelance work sites in which buyers "crowdsource" work by posting jobs they need done, and globally distributed workers bid on projects they are willing to take on.[1]

There are multiple advantages in this approach. First, many anti-abuse countermeasures are designed to detect or deter mechanistic automation and can be bypassed through the use of low-cost human labor. Perhaps the best known example of this phenomenon is found in CAPTCHAs, human-solvable puzzles designed to be challenging for automated solvers. While these puzzles are specifically designed to prevent computer-based service abuse, we have previously documented how a robust CAPTCHA-solving marketing has emerged by aggregating large amounts of cheap human labor instead [12].

A second advantage is that the crowdsourcing medium allows innovative attackers to quickly explore different schemes for evading anti-abuse defenses (due to the agility of a large contract labor pool). Finally, once a new attack scheme becomes sufficiently popular to commoditize, competitive pressures naturally drive workers to develop the most efficient means of satisfying the demand. Indeed, eventually the most popular activities (e.g., CAPTCHA-solving or phone verified accounts) can support their own branded retail services outside the scrum of the spot labor market.

In this paper, we characterize the abuse-related labor on Freelancer.com, one of the largest and most popular freelancer sites. Using almost seven years of historic data, and a range of our own contemporary work solicitations, we examine four classes of jobs:

- Account registration and verification,
- SEO content and link generation,
- Ad posting and bulk mailing,
- Social network linking

Each of these represents a kind of service abuse, incorporating manual labor to bypass existing controls, and each is ultimately a building block in some larger, economically-driven, advertising enterprise.

[1] To be clear, while the majority of such work is legitimate—anything from corporate logo design to software development—a large minority serves the online service abuse ecosystem.

The rest of this paper is organized as follows. Section 2 describes crowdsourcing and the Freelancer.com service in particular. In Section 3 we explain our methodology, the data we have gathered and the different categories of jobs in our study. Section 4 explains, as case studies, several components of the abuse value chain that have become semi-commoditized, followed by a characterization of the Freelancer labor market in Section 5. Section 6 places these abuse activities into a larger, interrelated context and Section 7 summarizes our findings.

2 Background

Outsourcing has long been a cost-cutting strategy in developed economies—pushing out key business processes to exploit the efficiencies or lower labor costs of third-party service providers. A more recent innovation is "crowdsourcing", further unbundling labor from any structured organization and leveraging the broad connectivity provided by the Internet. In this model, individuals participate in the labor force as free agents, responding to open calls for work on a piecework basis. In many cases, crowdsourcing is built on free labor (e.g., for many contributors to open-source projects, or in von Ahn's seminal ESP game [3]). However, fee-based crowdsourcing sites quickly emerged, the most famous being Amazon's Mechanical Turk service. Using such services, employers post requests for service at a particular price, while laborers in turn can "solve" the subset of requests that appeal to them.

However, crowdsourcing also presents a number of concerns. First, as an employment vehicle, crowdsourcing is controversial, since critics claim its pure free-market approach to labor has the potential to be highly exploitative, particularly of those in developing countries; one recent analysis estimates that the average hourly wage on Mechanical Turk is $5/hour [8]. Moreover, even on the employer side of the equation, crowdsourcing can be problematic since—absent any strong reputation mechanism—there may be little incentive for workers to provide quality work-products. Consequently, third-party services, such as *crowdflower*, have emerged that trade cost for data quality by replicating work requests and voting among them [1].

However, a less appreciated negative impact of this ecosystem is how anonymous access to cheap aggregated labor impacts the security of existing of Internet services. Indeed, as we show in this paper, the crowdsourced market for Web service abuse labor is thriving.

Much of this activity takes place on "freelancing" sites, in which employers post jobs and select individual workers based on their bids and bilateral negotiation. There are a large number of such sites with the most popular being Freelancer, Elance, RentACoder, Guru, and oDesk. In this paper we specifically examine the activity at Freelancer.com, one of the oldest and largest sites, claiming roughly two million employers and workers [6] from 234 different geographic regions and with close to nine hundred thousand projects posted on the site since 2004. We specifically chose Freelancer because the site offers an open API for querying information about past jobs and users. We have also gathered smaller amounts of data from most of the other large freelancer sites (i.e., via scraping) but since the activity is extremely similar across sites we chose to focus on the one for which our data was comprehensive.

Visitors to Freelancer must register and select a handle by which they are visible to other users. The only due diligence concerning a user's identity is a requirement to have a valid email address. The site does offer "skills tests" for a fee, by which individual users may demonstrate proficiency in various skills and earn "badges" visible on their profile. There is no discrimination between employers and workers and any user can participate on either or both sides of the labor market.

To post a project on the site, the project poster, or buyer, must pay a $5 fee, which is refunded once a worker is selected. Buyers may choose to pay an additional fee of $14 to have their jobs "featured", meaning that they are listed towards the top of the job listings. Workers independently scan these jobs listings to find projects matching their particular skill sets and then place bids (a combination of structured fields, such as dollar amounts, and freeform text). Buyers then select the workers who are most appropriate for their tasks.

Once workers are chosen, Freelancer charges either $3 or 3% of the total project cost to the buyers, depending on whichever amount is higher (Freelancer acts as the middleman in the transaction, using online payment methods such as PayPal, Moneybookers and Webmoney). However, some less scrupulous buyers are reputed to simply cancel their orders and settle with workers out-of-band. Finally, while job postings are effectively "broadcast", there are a range of such posts that identify themselves as private by specifically identifying the workers they are interested in employing.

3 Data Overview

In this section, we describe our methodology for collecting data on Freelancer job activity and categorizing the jobs into various kinds of "dirty" tasks.[2]

3.1 Data Collection Methodology

Freelancer.com exports an API for programmatically querying for information regarding projects and users. Using this API, we implemented a crawler to collect both

[2] While space prevents a detailed description of the oversight and ethical considerations here, our protocols were reviewed by our Human Research Protections Program and we consulted with key brand holders in advance of any active purchasing activity.

Activity	Count	
Projects	842,199	
Projects w/ Selected Workers	388,733	(46%)
Project Bids	12,656,978	
Active Users	815,709	
Buyers Only	179,908	(22.1%)
Workers Only	590,806	(72.4%)
Buyer & Workers	44,995	(5.5%)

Table 1: Summary of Freelancer activity between February 5, 2004 and April 6th, 2011.

contemporary and historical information about Freelancer activity. We ran the crawler from December 16, 2010 through April 6, 2011 to minimize load on the site. For historical data, we observed that Freelancer uses monotonically increasing IDs for both projects and users. To crawl all projects over time, we iterated through the entire available project ID space, which at the time ranged from 1–1,015,634. As a result, the job postings in our data set represent all of the jobs that were viewable through the API. We derived the set of user IDs based upon the set of projects, including any user associated with a project as buyer, bidder, or worker.[3]

For all crawled projects, we extracted the project details and the corresponding project bids, as well as the buyer, bidders, and selected workers who were awarded the projects (if any). For all users we encountered, we downloaded their public account metadata and feedback comments.

3.2 Data Summary

Starting with the earliest project posted on February 5, 2004 at 12:28 EST, we collected data through April 6th, 2011, capturing over seven years of activity. Table 1 summarizes this data set. During this time, 842,199 jobs were posted to Freelancer[4] and 815,709 users were active on the site. Roughly 46% of the posted jobs report a worker selected for the job. This number represents a lower bound on the number of job transactions; a buyer and a worker will sometimes use Freelancer to rendezvous, but will negotiate the transaction through private messaging and, thus, never report a selected worker. Among all users associated with at least one project, 22.1% were buyers only, 72.4% were bidders/workers only, and 5.5% served as both.

[3]Unlike projects, we did not exhaustively collect information for all two million users by crawling the user ID space since the majority of users do not appear to be active on the site.

[4]Note the discrepancy of 173,435 jobs between the maximum ID and the number of postings we obtained through the API. When crawling these IDs, Freelancer's API returned an error indicating that the ID was invalid. We assume that invalid IDs are jobs that never existed or have been deleted—which, according to complaints, happens for only a select number of jobs that egregiously violate Freelancer rules.

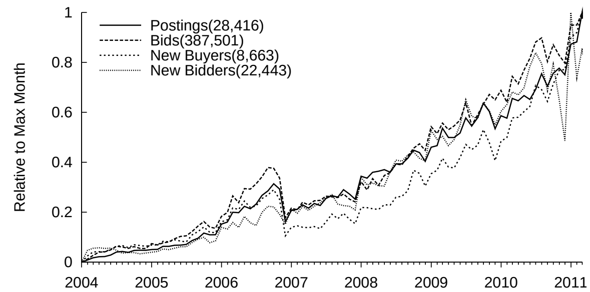

Figure 1: Growth in Freelancer activity over time. Numbers in parentheses in the legend denote the largest activity in a month.

Activity on Freelancer has grown steadily over time. Figure 1 shows the number of jobs offered and the number of bids made per month, as well as the number of new buyers and bidders per month. To overlay and compare the curves, we normalized them to their maximum monthly value as listed in Table 1. The curves all show a drop in activity in December 2006 (the reason for which we have not been able to determine). After this point, Freelancer experiences strong linear growth in buyers and bidders and their associated posting and bidding activity. Freelancer's job market is healthy and growing. Work posted by a steadily increasing number of buyers (5,000 new buyers a month on average in 2010) has been satisfied by an equally steadily increasing supply of bidders (15,000 new bidders/month).

3.3 Categorizing Jobs

Our first step in understanding Freelancer activity is to categorize the types of jobs found on the site. We use a two-step process for this categorization. We first manually browse sampled projects to identify a meaningful list of job categories. We then use a combination of keyword matching and supervised learning to identify jobs from the entire Freelancer corpus that fall into the categories.

From browsing random job postings, gauging the interest level in various tasks from observed bidding activity, and incorporating awareness of the larger underground cybercrime ecosystem, we identified 22 types of jobs falling into six categories. To establish a baseline of the prevalence of these types of jobs, we manually inspected a random sample of 2,000 jobs, tagging each job with a category.

Table 2 summarizes the list of categories and the distribution of jobs that fall into each category from our random sample. Note that a job may be tagged under multiple categories; for example, social bookmarking jobs for search engine optimization (SEO) usually also require account creation. Legitimate projects comprise 65.4% of these jobs and primarily involve Web-related programming and content creation tasks. We include private jobs, corresponding to projects targeted to one specific user, in the legitimate job class, since we typically do not know

Category	Job Type	Description	Count	%
Legitimate [§A.1]	Web Design/Coding	Create, modify, or design a Web site	769	38.5
	Multimedia Related	Complete multimedia-related task (e.g., Flash)	265	13.2
	Private Jobs	Jobs designated for a particular worker	138	6.9
	Desktop/Mobile Applications	Create a desktop or mobile application	100	5.0
	Legitimate Miscellaneous	Miscellaneous jobs	177	8.8
Accounts [§A.2]	Account Registrations	Create accounts with no defined requirements	22	1.1
	Human CAPTCHA Solving	Requests for human CAPTCHA solving	19	0.9
	Verified Accounts	Create verified accounts (e.g. phone)	14	0.7
SEO [§A.3]	SEO Content Generation	Requests for SEO content (e.g., articles, blogs)	195	9.8
	Link Building (Grey Hat)	Get backlinks using grey hat methods	53	2.6
	Link Building (White Hat)	Get backlinks using no grey/black hat methods	20	1.0
	SEO Miscellaneous	Nonspecific SEO-related job postings	61	3.0
Spamming [§A.4]	Ad Posting	Post content for human consumption	25	1.2
	Bulk Mailing	Send bulk emails	8	0.4
OSN Linking [§A.5]	Create Social Networking Links	Get friends/subscribers/fans/followers/etc.	33	1.7
Misc [§A.6]	Abuse Tools	Tools used for abuse (e.g., CAPTCHA OCR)	41	2.1
	Clicks/CPA/Leads/Signups	Get clicks, emails, zip codes, signups, etc.	32	1.6
	Manual Data Extraction	Manually visit websites and scrape content	21	1.1
	Gather Email/Contact Lists	Research contact details for targeted people	17	0.9
	Academic Fraud	Write essays, code homework assignments, etc.	10	0.5
	Reviews/Astroturfing	Create positive reviews	1	0.1
	Other Malicious	Miscellaneous jobs with malicious intentions	35	1.8

Table 2: Distribution of 2,000 random, manually-labeled projects into job categories. Referenced sections of the appendix include examples of jobs in the corresponding category.

the job details; private postings, however, will sometimes contain enough data to determine their intent. In our manually labeled corpus, we were unable to determine the intent of 5.4% of the jobs. The remaining 29.2% of the jobs correspond to various kinds of "dirty" jobs, ranging from delivering phone-verified Craigslist accounts in bulk to a wide variety of search-engine optimization (SEO) tasks.

We then focused on identifying jobs in the entire Freelancer corpus that fall into "dirty" categories. Since we could not manually classify all jobs, we used keyword matching to generate training sets and supervised learning to train classifiers for each category. We then applied the classifiers to each job to determine the dirty category it falls into, if any.

To find positive examples for each classifier, we used keywords associated with the job type to conservatively identify jobs that fall into each category. For example, to locate jobs about CAPTCHA solving, we searched job postings for the terms "CAPTCHA" and "type" or "solve". For negative examples, we randomly chose jobs from the other orthogonal job types. For features, we computed the well-known *tf-idf* score (term frequency-inverse document frequency) of each word present in the title, description, and keywords associated with jobs in the training sets. We then used *svm-light* [9] to train classifiers specific to each category.

Table 3 shows the results of applying these classifiers to all Freelancer jobs. We focus on just those dirty job categories that had at least 1,000 jobs. Although the classifiers are not perfect (e.g., some jobs placed in the "link building" categories might be better placed in the more generic "SEO" category), they sufficiently capture the set of jobs in each category and greatly increase the number of jobs we can confidently analyze. Note that we did not attempt to be complete in the categorization of the postings: there are likely jobs that should be in a category that we have missed. However, such jobs are also likely not well-marketed to workers, since they most likely lack the typical keywords and phrases commonly used in jobs under those categories.

We focus on the jobs comprising these categories in the analyses we perform in the subsequent sections.

3.4 Posting Job Listings

Pricing information is a crucial aspect of our study, since it represents the economic value of an abusive activity to attackers. Both job descriptions and bids contain pricing info, often at odds with each other. To determine which source of pricing info to use, we performed an experiment where we posted jobs on Freelancer and solicited bids. In the process, bidders posted public bids and, in some cases, sent private messages to our user account.

Class	Job Type	Count	%
Accounts	Account Registrations	6,249	0.7
	Human CAPTCHA Solving	4,959	0.6
	Verified Accounts	3,120	0.4
SEO	SEO Content Generation	72,912	8.7
	Link Building (Grey Hat)	16,403	1.9
	Link Building (White Hat)	10,935	1.3
Spamming	Ad Posting	11,190	1.3
	Bulk Mailing	3,062	0.4
OSN Linking	OSN Linking	11,068	1.3

Table 3: Freelancer jobs categorized using the classifiers.

These private messages occasionally reveal the external Web store fronts operated by Freelancer workers, in addition to the tools, services, and methods they use to complete each type of job. We posted 15 job listings representative of the categories for which we have classifiers. We also randomly posted half of the jobs as a "featured" listing to determine whether this increased the quantity of bids we received (which it did).

Table 4 summarizes the results of our job posting experiments. Of the 228 total bids we received, 47 were commensurate with market rates for these projects. Most of the remaining bids, however, were simply minimum bids used as "place holders". The actual bid amount was either included in a private message to our buyer account, or the bidder provided an email address to negotiate a price outside of the Freelancer site to avoid the Freelancer fee.

Because many prices in the public bids severely underestimate market prices, we use the prices in job descriptions by buyers in our studies in Section 4. Even so, we note that the pricing data has some inherent biases. They are advertised prices and not necessarily the final prices that may have been negotiated with selected workers. Further, we use prices that were systematically extracted from the job descriptions. Even with hundreds of hand-crafted regular expressions, we were only able to extract pricing data from about 10% of the jobs. Job descriptions are notoriously unstructured, ungrammatical, and unconventional. These biases notwithstanding, the pricing data is still useful for comparing the relative value of jobs, as well as trends over time.

Class	Job Type	Bids	Cost
Accounts [§B.1]	Craigslist PVA	10 (4)	$4.25
	Gmail Accounts	6 (5)	$0.07
	Hotmail Accounts*	21 (12)	$0.007
	Facebook Accounts*	24 (10)	$0.07
SEO [§B.2]	Blog Backlinks*	10 (5)	$0.30
	Linking (White Hat)*	17 (8)	$0.81
	Forum Backlinks	12 (9)	$0.50
	Social Bookmarks*	44 (21)	$0.13
	Bulk Article Writing	29 (23)	$3.00
Spamming [§B.3]	Bulk Mailing	10 (5)	0.075¢
	Craigslist Posting	10 (3)	$0.60
OSN Linking [§B.4]	Facebook Friends*	11 (4)	$0.026
	Facebook Fans	5 (5)	$0.039
	MySpace Friends	2 (2)	$0.037
	Twitter Followers*	7 (6)	$0.02

Table 4: Results from posting job listings to Freelancer. A "*" indicates the post was featured, the number within the "()" is the number of bids that included prices. All prices in the cost column are for the smallest unit of service (i.e., per one account, backlink, email, post, and 500-word article).

4 Case Studies

This section features case studies of the four groups of abuse-related Freelancer jobs summarized in Table 2.

4.1 Accounts

Accounts on Web services are the basic building blocks of an abuse workflow. Because they are the main mechanism for access control and policy enforcement (e.g., limits on number of messages per day), circumventing these limits requires creating additional accounts, often at scale. Thus account creation has become the primary battlefield in abuse prevention.

Accounts primarily enable a wide variety of spamming and scamming. For Web mail services like Gmail and Yahoo, spammers use accounts to send email spam, taking advantage of the reputation of the online service to improve their conversion rate. For online social networks like Facebook and Twitter, spammers use accounts to spam friends and followers (Section 4.2), taking advantage of relationships to improve conversion. For classified services like Craigslist, spammers use accounts to create highly-targeted lists, post high-ranking advertisements for a variety of scams, recruit money laundering and package handling mules, advertise stolen goods, etc. Further, accounts on some services easily enable paired accounts on related services (e.g., creating a YouTube account from a Gmail account), further extending the opportunities for spamming.

4.1.1 Account Creation Insights

In the context of another research effort, we obtained approval from a major Web mail provider to purchase fraudulently-created accounts on their service. We purchased 500 such accounts from a retail site selling accounts, gave them to the provider, and in return received registration metadata for the supplied email accounts, including account creation times and the IP addresses used to register the accounts. We later discovered that the supplier we contacted was a very active member of Freelancer.com; this worker is responsible for account set $\mathtt{IN}_1$ in Table 5.

The supplier had bid on 2,114 projects, had been cho-

Name	Rating	Tested	Valid (%)	Age (Days)
$\mathtt{IN}_1$*	9.8	500	100.0	0.4
$\mathtt{UK}_1$	9.9	3,500	99.9	25.7
$\mathtt{BD}_1$	10	6,999	99.6	24.7
$\mathtt{IN}_2$	9.8	5,015	99.6	9.7
$\mathtt{PK}_1$	10	4,999	99.4	78.6
$\mathtt{PK}_2$	9.8	4,000	95.4	82.6
$\mathtt{PK}_3$	9.9	4,013	77.3	414.7
$\mathtt{IN}_3$	9.9	6,200	76.2	30.7
$\mathtt{CA}_1$**	9.6	508	15.7	21.7

Table 5: Summary of the results from purchasing email accounts. The names of the account sets embed the worker countries: IN is India, UK is the United Kingdom, BD is Bangladesh, PK is Pakistan, and CA is Canada. The rating column refers to the average rating of the selected worker. *Notes:* *We purchased $\mathtt{IN}_1$ in 2010, the rest in 2011. **The worker responsible for $\mathtt{CA}_1$ repeatedly copied and pasted 508 accounts to meet the 5k requirement.

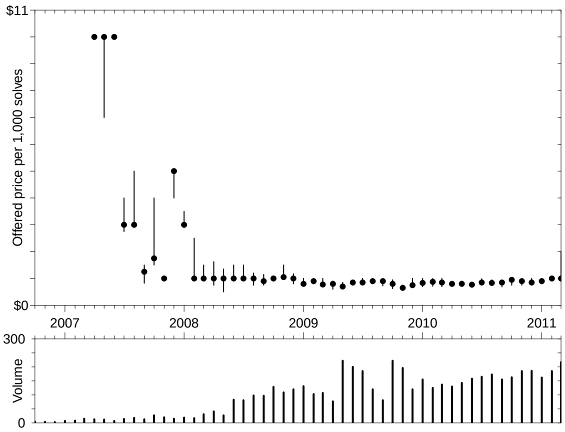

Figure 2: Median monthly prices offered by buyers for 1,000 CAPTCHA solves (top) and the monthly volume of CAPTCHA solving posts (bottom), both as functions of time. The solid vertical price bars show 25% to 75% price quartiles.

sen as a selected worker on 147 projects, and served as a buyer on 84 projects. Interestingly, the supplier acted as a buyer for 25 jobs that involved the creation of other Web mail account types. The supplier contracted out this task at a rate of $10–20 per 1,000 accounts, and yet the supplier charged $20 per 100 accounts on the retail Web site, an order of magnitude more.

The accounts we purchased were created an average of only 2.8 seconds apart, suggesting the use of either automated software or multiple human account creation teams in parallel.[5] Such automation would be one way to earn money bidding on account jobs for this particular worker. Further, 81% of the IP addresses used to register the accounts were on the Spamhaus blacklist, suggesting the use of IP addresses from compromised hosts to defeat IP-based rate limiting of account creation.

4.1.2 Experience Purchasing Accounts

In 2011, we commissioned a job to purchase additional email accounts for the same Web mail provider in quantities ranging from 3,500–7,000. We selected nine different workers, of which eight ultimately produced accounts, listed in Table 5 after $\mathtt{IN}_1$. Once given the accounts and the corresponding passwords, we logged into the accounts and downloaded the newest and oldest inbox pages (assuming the account was valid). Table 5 shows the results of the purchasing and account checking. Of the eight email sets, seven consisted of largely valid accounts, with over 75% of the tested email accounts yielding a successful login. $\mathtt{IN}_3$ was particularly interesting; the worker previously used the email addresses to create Facebook and Craigslist logins and posts, then resold the accounts to us. Also, four of the account batches are relatively old (as determined by the date of their oldest emails), with the median age of the accounts between two months and over one year. These ages indicate that workers are likely sitting upon a stockpile of email accounts. Lastly, the worker ratings do not seem to reflect the quality of the accounts, as demonstrated by the high ratings (out of 10) achieved by those workers responsible for the $\mathtt{PK}_3$, $\mathtt{ID}_3$, and most notably, $\mathtt{CA}_1$ account sets.

4.1.3 CAPTCHA Solving

To keep the barrier to participation extremely low, creating an account at an online service today requires little more than solving a CAPTCHA. CAPTCHAs are designed to be hard to solve algorithmically, and thus create an obstacle to automating service abuse. In response to their widespread deployment, human-based CAPTCHA-solving services emerged in abuse ecosystem. Such services depend on cheap human labor to provide a simple programmatic interface for solving CAPTCHAs to an otherwise completely automated abuse processes chain. In a previous study [12], we described a robust retail CAPTCHA-solving industry capable of solving a million CAPTCHAs a day at $1 per 1,000 solved. Thus today, CAPTCHAs are neither more nor less than a small economic impediment to the abuser, forming the first step in the account value chain.

By their nature, CAPTCHAs are ideally suited to the Freelancer outsourcing paradigm, and indeed the Freelancer marketplace has played a key role in the evolution of CAPTCHA solving. Figure 2 shows the history of prices offered for CAPTCHA solving as well the demand (in number of job offers per month) since 2007. We see a rise in demand starting from their first appearance, and a corresponding drop in prices to the $1 per 1,000 price

[5] We know that an effective automated CAPTCHA solver existed at this time for this Web mail provider, so automation is the likely suspect.

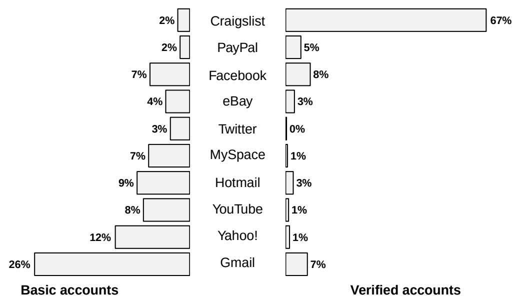

Figure 3: Sites targeted in account registration jobs.

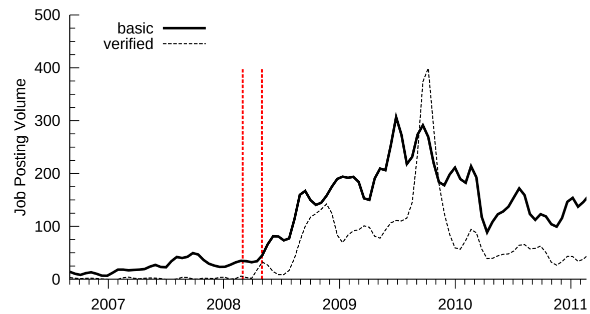

Figure 4: Demand for account registration jobs over time. The dashed vertical lines indicate approximate dates when Craigslist introduced phone verification for erotic services ads (March 2008) and other services (May 2008).

seen today, corroborating our previous findings [12].

4.1.4 Account Verification

Because creating a basic account—even one requiring solving a CAPTCHA—is so cheap, to curb online abuse services must necessarily take advantage of some limited resource available to a user. To increase the limits placed on a basic account, a user must sometimes undergo *account verification*, which takes a variety of forms (e.g. phone numbers, credit cards, etc.). Verification increases the user's standing within the service, giving the account holder greater access to the service and thereby increasing the value of the account. For this reason, verification is a step in the value chain of many abuse processes.

The most popular type of verified account uses phone verification. Beyond the steps for creating a basic account, phone-verified accounts (PVAs) require a working phone number as an additional validation factor in account authorization. Services will either call or message a code to the number, and the user must submit the number back to the service to complete authorization. For some services phone verification is mandatory (e.g., for posting advertisements in certain forums on Craigslist, creating multiple accounts in Gmail from the same IP address), and for other services, phone verification adds convenience (e.g., avoids CAPTCHAs with Facebook). Services typically require the phone number to be associated with a landline or mobile phone since, unlike VoIP phone numbers, it is much more difficult to scale the abuse of such numbers. Phone verification is effective: immediately after Gmail introduced phone verification to limit account abuse, for instance, prices for Gmail accounts on underground forums skyrocketed to 10 times other Web-mail accounts [2]. However, even more so than CAPTCHAs, PVAs add further delay and inconvenience to users and is the primary reason why services do not use phone verification uniformly.

4.1.5 Web Services Targeted

Figure 3 shows the distribution of services targeted in job postings for basic and verified account registrations. For ease of comparison, it shows the top 10 targeted services for both kinds of accounts, combined. For a job targeting multiple services, we count it in the total for each service mentioned. Job postings target accounts in every major category of Internet service: Web mail, social networks, as well financial and marketplace services. However the distribution of specific services differ markedly between the two types of account registration jobs, reflecting how services vary in their deployment of additional verification mechanisms (if any). Basic accounts are useful for many purposes, including obtaining accounts up for other Internet services (Facebook, Craigslist, etc.), and Gmail is by far the most popular. When it comes to verified accounts, on the other hand, Craigslist is the dominant target, most certainly because Craigslist sections targeted by spammers all require PVAs.

We posted a job soliciting bids for "CraigsList Phone Verified Accounts PVA" on Freelancer.com. Of the 10 bids we received, 4 contained prices: \$3, \$4, \$4.50, and \$6. These prices are consistent with the currently observed buyer offers for Craigslist PVAs. The pricing of PVAs tells us in monetary terms the value of phone verification as a security mechanism. For Craigslist, PVAs have made account abuse extremely expensive. In contrast, retail services sell Gmail PVAs for around 25¢, a 10–20 fold price difference compared to Craigslist.

4.1.6 Trends

Demand for accounts through Freelancer grew dramatically starting mid-2008. Figure 4 shows the number of account creation jobs posted over time. Demand for basic accounts steadily increased through mid-2008, then dramatically increased until it peaked in mid-2009.

Demand for verified accounts rose greatly when Craigslist introduced phone verification for the erotic services section of their site in early March 2008 [4]. Demand grew steadily until about October 2009, and then dropped. We extracted prices from the Craigslist postings, and observed that Craigslist PVAs first rose to \$4 by the end of 2008 and then settled around \$2. In October of 2009, prices spiked to more than $5, then hovered between $2 and $3 through 2010.

For both types of accounts—basic and verified—demand dropped during 2010. We do not know the cause; however we suspect this may be due to stricter policing on behalf of Freelancer.com; our own price solicitation for Craigslist posting was canceled by the site.

4.2 OSN Linking

Online social networking links can be abused in two ways: (1) as a communication channel to market to real users, which is a finished product ready to directly monetize; (2) as an intermediate product to increase the reputation—and thus influence—of accounts by adding social links to other fake accounts. Previous work has shown that online social networking spam has a higher click-through rate than traditional email-based spam [7]. Thus, OSN platforms have emerged as a lucrative marketing venue where spammers are exploiting the trust relationships that exist in social networks to improve their conversion rates. However, it is difficult for a spammer to contact users on a social networking site until they have established a *social link* with real users. These social links take many different forms, depending on the targeted social networking site, such as convincing a user to friend the spammer, follow a spammer's Twitter feed, become a fan of the spammer's page, or subscribe to the spammer's YouTube channel. Building social links to real users is analogous to gathering email addresses that will later by monetized with email spamming. Once this social link is established, the spammer has a communication channel that is both highly reliable and not subject to aggressive filtering.

Adding fake social links is a relatively inexpensive method for increasing the reputation of an account, which in turn presumably improves the success rate of establishing links to real users. This method is effective because people are more willing to establish or accept social links that are more popular in terms of the number of previously-established social links or other endorsements. If the account has many social links and, more importantly, if mutual social links exist, the likelihood increases that the targeted real user will establish or accept a social link with the spammer.

In this section we survey the Freelancer.com market for buying both real and fake bulk social links.

4.2.1 Characterization

There are two main categories of social networking links requested in jobs. The first are friendship relationships (e.g., MySpace and Facebook friends), where an active invitation is offered and, if accepted, targeted messages can then be delivered to a user's private inbox. The second are subscription relationships (e.g., Facebook fans, Twitter followers, YouTube subscribers) where, if a user can be induced to follow a spammer's account, messages will appear in a user's feed; depending on the site, the relationship also grants the ability to send private messages to the user. A closely related goal is to use social links to increase the perceived popularity of an object. Examples of this type of task are increasing the view count of YouTube videos, or digging links on Digg. We group all these jobs into the category of social network links and they all follow the form of increasing the reputation of an account/object or establishing a marketing channel to real users.

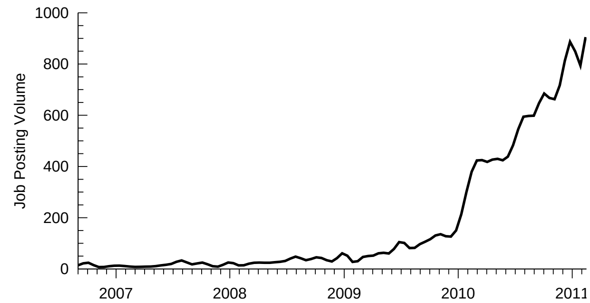

Figure 5: Number of job postings for social networking links.

Jobs for bulk social link building range from a few hundred to hundreds of thousands of links. Typically jobs interested in acquiring fake social links will request a relatively small number of links spread out over a large number of accounts (e.g., add 500 friends to 50 accounts). The requests for social links to real users often specify a target demographic for the links, thereby exploiting the same targeted marketing potential of using information included in a profile that legitimate advertisers on these sites also use to improve ad targeting. For example, a job might require that most social links be to male accounts in the US over the age of 18. The most targeted geographic demographics are high-income English speaking countries including the US (46%), UK (13.2%), Canada (9.5%) and Australia (6.2%). Also, based on keyword searches, females are specifically targeted in 8% of jobs and males in 3% of the jobs.

4.2.2 Trends

Figure 5 shows the demand over time for job postings for social networking links. Overall demand for social links has skyrocketed since the early part of 2010, suggesting that spammers have only recently realized the potential for monetizing social links. The social networking sites with the largest English-speaking user bases (Facebook, MySpace, Twitter, and YouTube) are targeted by 97% of the job postings for social links. Over 50% of social link jobs included words such as "real" and "active" indicating that they were seeking to buy a more finished type of social link that could be directly spammed. This percentage is a lower bound, however, as it is unclear how many

Name	Rating	Links	Top Countries (%) US	IN	BD	PH
BD_2	9.8	1,034	26.2	13.8	5.9	7.7
BD_3	9.8	1,081	43.3	7.4	32.5	4.4
BD_4	8.4	1,063	74.5	0.3	25.2	—
BD_5	10	1,071	—	—	100	—
BD_6	10	1,145	60.0	8.7	8.4	5.3
BD_7*	9.8	555	30.6	10.4	10.6	8.4
IN_4	9.9	1,095	64.3	25.1	10.5	—
MY_1	9.8	1,110	99.1	—	—	0.1
PK_4	–	1,015	24.7	9.2	5.9	7.0
RO_1	10	1,058	31.8	11.0	8.8	8.4

Table 6: Summary of the social links purchased to pages for our custom Web sites. The names of the sets correspond to the selected workers' home countries, while the rating column refers to his or her average rating. The worker responsible for BD_7 did not complete the job in a timely manner. Country codes: BD – Bangladesh, IN – India, RO – Romania, MY – Malaysia, PK – Pakistan.

postings did not include these types of words but were actually seeking real social links.

Overall the median offered price in posts were \$0.01 per social link, and median bids were between \$0.02–0.03 per a social link. These prices were similar across all of the social networking sites. This low price point raises the interesting question of whether proposed defenses that mitigate Sybil attacks via analysis of social link structure [14, 15] might be vulnerable to adversaries that are willing to simply hire humans to create real social links.

4.2.3 Experiences Purchasing Social Links

In preparation for purchasing social links, we instantiated several Web sites on the topic of cosmetics consulting [16] and created separate "pages" about each site on a popular social networking service. We then commissioned a job to obtain one thousand social links for these pages. The posted job explicitly targeted users from the US, Canada, and the UK. We assigned the task to 10 different workers, each given a different Web site to target.

Table 6 shows the results of this task. The name of the sets correspond to the selected workers' home countries, and the links column is the maximum reported daily number of social links. Most of the workers delivered the required number of social links in a timely manner (except for the BD_7 set); the quality of the social links, however, was quite poor. Most of the workers did not deliver social links from users that met our specifications, particularly in regards to user countries. Also, several of the workers added social links at a rapid pace, with some jobs being completed in as few as two days. Next, we observed substantial overlap between the users linked to our target pages, shown in Figure 6. As many as 50% of the

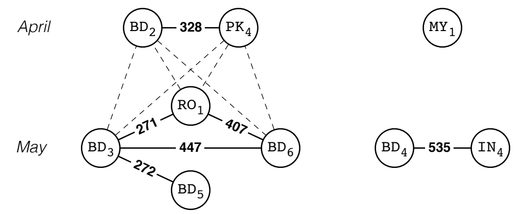

Figure 6: The number of user accounts common to each pair of workers hired to create social links. Labeled solid lines indicate at least 100 user accounts (out of 1,000 requested) in common, dashed lines indicate at least 10 but fewer than 100 user accounts in common. Work performed by MY_1, PK_4, and BD_2 was done in April, while the remaining jobs were done roughly a month later.

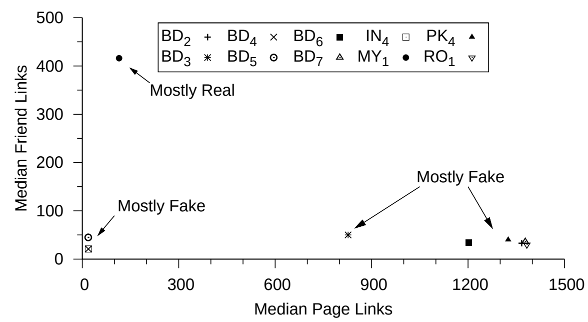

Figure 7: Median number of friends vs. median number of page social links for the sets of users linked to our websites.

users (between IN_4 and BD_4, for example) overlapped. This overall suggests that the workers are all manipulating the same set of users to produce these social links, or even perhaps subcontracting out the task to the same groups of workers. Only one worker, responsible for MY_1, had no overlap with any of the other sites. Again, the selected worker ratings do not reflect the quality of the delivered products; we posit that buyers who hire these workers find it difficult to evaluate social link quality.

Next, we extracted the profiles for the OSN users who were linked to our target Web sites, and looked at the number of friends and page links listed on their profiles. Figure 7 shows a scatterplot of the median number of friends versus the median number of page links for these OSN users. Several clusters emerge in the graph. Within each user batch, we manually visited the profiles of those users; only one worker, MY_1, appears to have delivered social links from legitimate users. The rest used predominately fake accounts, many of which had few friends and a large number (>1,000) of page social links.

4.3 Spamming

In our study, we consider spamming to be the dissemination of an advertiser's message to users by means other than established advertising networks. Spamming provides the buyer with a direct marketing channel to his targets, and as such, represents one of the most finished commodities in the advertising value chain.[6]

In our survey and classifier-based labeling (Tables 2 and 3), the class of spamming jobs is comprised of ad posting and bulk mailing.[7] Because Craigslist is the main target of ad posting jobs (82%), we treat it separately. We begin by first analyzing the pricing data for bulk mailing.

4.3.1 Bulk Mailing

Bulk mailing is simply traditional email spam and represents 0.3–0.4% of all jobs posted on Freelancer.com. In most cases, the buyers supply their own mailing lists, although some—generally targeting larger volumes—expect bidders to supply their own address lists.

We extracted pricing data from the job descriptions of 236 postings. We averaged these prices and discovered that buyers on Freelancer.com were willing to pay approximately \$5.62 to send 1,000 emails, with a median price of \$1.00. The extracted prices varied wildly; thus, we manually scanned another 100 random postings. Again, we observed a wide range of prices, from one buyer willing to pay only \$0.06/1,000 emails, to another buyer willing to pay \$5.00/1,000 emails.

A final point of comparison is our own posting for bulk mailing services. We posted a job that involved sending bulk emails to three million individuals and received 10 responses. Of the 10 responses, five included a price, and these prices ranged from \$0.30 to \$2 per 1,000 messages (with a median of \$0.75/1,000 emails).

4.3.2 Craigslist Ad Posting

Posting an ad on Craigslist is typically free, but Craigslist takes special measures to restrict the number of ads posted by a single individual (e.g., IP rate limiting, CAPTCHAs, etc.). In the context of our study, when Freelancer.com buyers create jobs to "spam" Craigslist, their goal is to obtain *repeated* ad postings from workers, usually on a daily basis. This is done to keep a buyer's ads at the top of the search results. Our classifier identified 11,190 job postings of this type, 9,096 (81%) of which contained the service name "Craigslist" or a variation thereof (in total comprising 1.1% of all jobs on Freelancer.com).[8]

Figure 8 shows the prices offered by buyers for a single Craigslist posting (top) and the average number of job posts per day pertaining to Craigslist ad posting (bottom). The solid circles indicate monthly median prices, and the solid bars show the 25% to 75% quartiles of the prices. In early March 2008, Craigslist added a phone verification requirement for posting in the erotic services section [4], and later extended the requirement to posting in other parts of the site some time in early May 2008 (both dates indicated with dashed vertical lines in the graph).

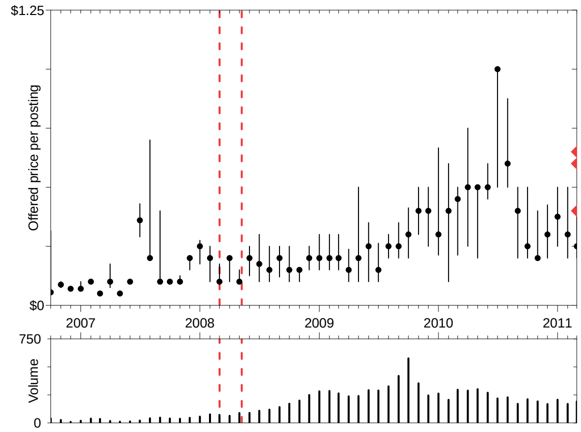

Figure 8: Median monthly prices offered by buyers for each Craigslist ad posted (top), and the monthly number of posts (bottom), both as a function of time. The solid vertical price bars show 25% to 75% price quartiles. The dashed vertical lines indicate approximate dates when Craigslist introduced phone verification for erotic services ads (March 2008) and other services (May 2008). The three bids received in response to our solicitation are indicated with a triangle on the right edge.

Figure 8 illustrates that the demand for *posting* to Craigslist started growing gradually after the policy changes, and the prices offered by buyers stayed essentially unchanged until mid-2009. Recall that in mid-2009, demand for phone verified accounts (which are dominated by Craigslist) appears to drop dramatically (Figure 4), having increased rapidly over the past year. Note, however, that the demand for Craigslist ad *posting* continues to rise during that same time period, nearly quadrupling in price within a year.

To further compare pricing data, we posted a job description on Freelancer soliciting bids for "Experienced Craigs List Posters." We received 10 responses, with three bids of \$0.40, \$0.60, and \$0.65 per ad; these prices are shown for comparison purposes in the top graph of Figure 8 as solid triangles on the right edge. These prices are roughly in accordance with the buyer offers.

4.4 Search Engine Optimization

Search engine optimization (SEO) represents the second major advertising channel along with spam. SEO is a multi-billion dollar industry for improving the ranking of sites and pages returned in search results on popular search engines. Improving the ranking of pages in

[6] The most finished commodity is actual site traffic; however, traffic of reasonable quality (with respect to conversion rate) usually requires site-specific targeting and additional advertiser-provided material ("creatives").

[7] While we found several other kinds of spam-like jobs (e.g., bulk SMS), they did not represent a significant fraction of all jobs, and are not part of our study.

[8] Classified ad sites BackPage and Kijiji represented 6.6% and 5.5% of jobs classified as ad posting; we chose to focus on Craigslist because it dominated this job category.

search results increases traffic to that page. "White hat" SEO improves the search rank of pages while obeying the guidelines provided by search engine companies like Google that prevent abuse of the indexing and ranking algorithms. "Black hat" SEO abuses the indexing and ranking algorithms, sacrificing the relevance of a page with the sole goal of attracting traffic via search results.

There are three kinds of black hat SEO offerings on Freelancer.com, spanning the spectrum from least to most "finished": content generation, link building, and search placement.

Content generation increases the number of sites that contain indexable content together with links to a target page. This goal is achieved either by having writers generate unique content for sites, often by rewriting existing material, or by using a semi-automated technique known as *spinning*. Spinning often uses structured templates together with a variety of word, phrase, and sentence "dictionaries" to generate many variants of effectively the same content, and is analogous to the template-based techniques used to generate polymorphic spam that can defeat spam filters [11].

Link building is a more focused type of SEO job whose goal is to place links on pages with existing content, emphasizing placement on pages with high rank as defined by search engines. Rather than generating and distributing content across many sites as a basis for improving the ranking of a target page, link building bootstraps on existing highly-ranked pages.

The most finished kind of SEO job is search placement. The buyer does not care *how* the desired search placement is achieved, only that they place in the top search results on Google. Such jobs were relatively rare on Freelancer, and we only survey content generation and link building jobs in further detail.

4.4.1 Content Generation

A popular form of abusive SEO is to post "articles" to various sites and forums. These articles contain keywords and links intended to increase the search engine PageRank of a page in search results returned from queries that use the same keywords. With proper accounts (Section 4.1), the posting step can be automated. However, defenses implemented by search engines can detect automatically-generated article content. Such defenses have thereby created a demand for human workers to generate sufficiently realistic articles that defeat the countermeasures. Indeed, such article writing jobs represent the most popular abusive job category by far, accounting for over 10% of all Freelancer jobs (Table 2).

Article job descriptions request batches of 10–50 articles at a time in grammatically correct English on a particular topic, seek articles typically 250–500 words in length, and often have a variety of requirements

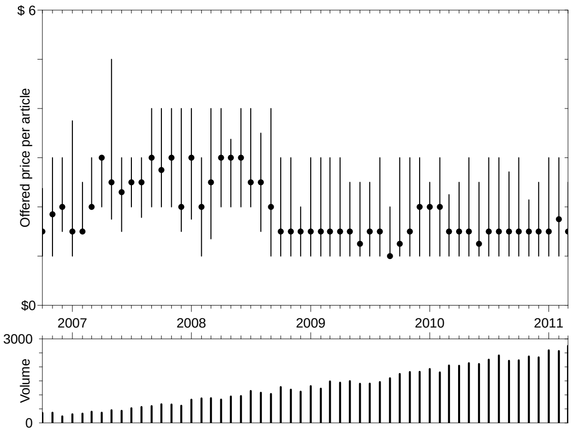

Figure 9: Median monthly prices offered by buyers for each article posted (top) and monthly average number of buyer posts per day (bottom), as a function of time. Vertical price bars show 25% to 75% price quartiles.

that reflect perceived countermeasures implemented by the search engines. A frequent requirement is sufficient "originality" (albeit often of simply rewritten text) to pass CopyScape, a popular plagiarism detection tool; such originality counters the capability of search engines to detect and discount similar content. Other such requirements request rewritten text beyond straightforward manipulation of existing content (simple synonym substitution, transposing sentences, etc.).

Figure 9 shows buyer demand and offered prices over time for article content generation jobs. Growth in demand for articles has been strong, with the number of jobs offered increasing linearly, with a peak of nearly 3,000 article jobs posted in August 2010. This substantial growth in demand strongly suggests that article writing is indeed an effective form of SEO abuse. Yet prices for articles have been relatively stable over the past four years, with buyers offering $2–4/article.

4.4.2 Experiences Purchasing Articles

To evaluate the quality of the articles written by Freelancer workers, we solicited and employed ten workers to write six original articles on the topic of skin care products. We required each article to contain at least 400 words, have a keyword density of at least 2%,[9] and pass the CopyScape [5] plagiarism detection system. Table 7 shows the results of this assignment. Workers are identified by their two-letter country and a digit. In addition to the three criteria above, we also computed the articles' Flesch–Kincaid Grade Level [10]—a measure of text

[9] "Keyword density" is the frequency of occurrence of a set of keywords provided by the bidder to be included in the text. Keyword density thresholds ensure that search engines index a Web page with respect to the specified keywords. In our experiment, we provided workers with keywords such as "dry skin moisturizer" and "exfoliating scrub".

		Failed articles			
ID	Rating	Len	KD	CS	FKGL
IN_5	9.50	–	6	–	8.8 ± 1.0
PH_1	9.75	4	5	–	7.7 ± 0.9
BD_8	–	–	4	–	8.1 ± 0.7
KW_1	9.62	–	3	–	10.0 ± 0.3
IN_6	9.62	–	2	–	7.2 ± 0.8
UK_2	10	–	1	2	9.0 ± 0.5
US_1	10	–	1	–	8.6 ± 0.2
BD_9	9.81	–	1	–	9.3 ± 0.5
AU_1	–	–	1	–	11.0 ± 1.0
KE_1	10	–	–	–	9.6 ± 1.0

Table 7: Quality of articles written by workers on the topic of skin care products. Columns *Len*, *KD*, and *CS* show how many of each worker's six articles failed the length, keyword density, and CopyScape plagiarism detection requirements. The *FKGL* column shows the Flesch–Kincaid Grade Level [10] range of each worker's text after excluding their lowest and highest scoring articles. Country codes: PH – Philippines, IN – India, BD – Bangladesh, KW – Kuwait, UK – United Kingdom, US – United States, AU – Australia, KE – Kenya.

readability based on word and sentence length, roughly indicating the school grade level required to comprehend the text. The FKGL column shows the score range of the work produced by each worker after excluding their lowest and highest scoring articles.

Quality of the work produced by the ten workers varied considerably. More than half of the articles produced by workers IN_5, PH_1, and BD_8 did not meet our 2% keyword density requirement; in addition, PH_1 failed to produce articles of the required length (400 words). On the other hand, half of the workers produced articles satisfying our criteria in at least five out of six cases. Unfortunately, two of the articles produced by UK_2 did not pass the CopyScape plagiarism detection tool, and as such, would likely not be indexed by search engines.

Articles written by the workers were understandable and on topic. The Flesch–Kincaid Grade Level of the articles reveals a notable level of English composition. For comparison, five Wikipedia articles on the same topic had scores in the range 12.1 ± 0.5, while six articles from Cosmopolitan—a popular women's magazine in the US—about skin care fell in the 7.9 ± 0.8 range. Thus, at least with respect to SEO, our results show Freelancer to be a useful source of inexpensive content that would be difficult to distinguish mechanically from work produced by more highly-paid specialist writers.

4.4.3 Link Building

Google reports a PageRank (PR) metric for every page, accessible via the Google Toolbar. The PR ranges from 0–10, with new and least popular pages having a PR of 0 and the highest ranked pages having a PR of 10. This PageRank is a combination of the number of sites that link to the page—so-called backlinks—and the PageRanks of the pages with the backlinks. Not surprisingly, another common SEO abuse is to increase the number of sites that backlink to a page, and to have those backlinks on sites with high PageRank.

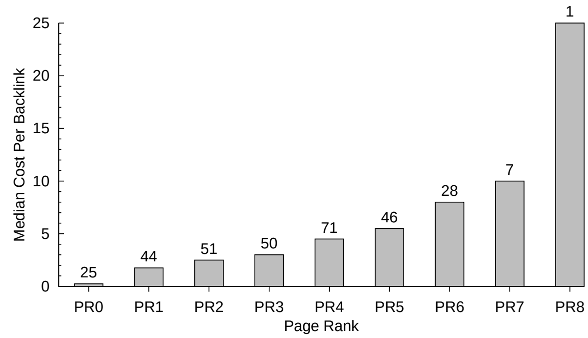

Figure 10: Average price buyers offered for backlinks on pages with a given PageRank (PR). Higher PRs correspond to more popular (and valuable) pages. The number above the bar corresponds to the number of jobs requesting backlinks of that PR.

Hiring people to perform this kind of SEO task is another frequent kind of abusive job on Freelancer, accounting for over 3% of all jobs. We placed such link-building tasks into two categories, "white hat" and "grey hat". White hat link building jobs have requirements that specifically try to avoid search engine countermeasures, such as no link farms, no blacklisted sites, no redirects or JavaScript links, links on sites with generic top-level domains, and so on. Jobs also specify the PageRank of the pages on which the backlinks will be placed, and that the buyer will validate the links created according to all of their criteria. Grey hat link building is much more indiscriminate, such as spamming blogs with links embedded in comments.

How much do people value backlinks as an SEO technique? The job postings quantify this value in economic terms. For the "white hat" link building jobs for which we could automatically extract pricing data, Figure 10 shows that the median price per backlink buyers offered is directly correlated with the PageRank (PR) of the page containing the backlink. One buyer offered over $25 per backlink on pages with PR8, while buyers offered nearly $5 per backlink on PR4 pages, the most popularly-requested PR.

Next, we look more closely at buyers who posted "grey hat" link building jobs, or ones that allow for such questionable SEO methods as blog commenting, forum posting, etc. For these Freelancer job postings, buyers oftentimes directly specify the URL that they are interested in using greyhat techniques on. We extracted over two thousand URLs that were present in the body of the grey-

Domain Name	Num. Sites	Num. Inlinks
Blogspot	316	10,028
Wordpress	213	2,402
Yahoo	147	1,187
ArticlesBase	143	747
Folkd	108	302
ArticleSnatch	107	491
Google	97	184
Squidoo	88	154
Diigo	88	277
ArticleAlley	88	471

Table 8: Summary of top 10 targeted domain names for greyhat link purchasing.

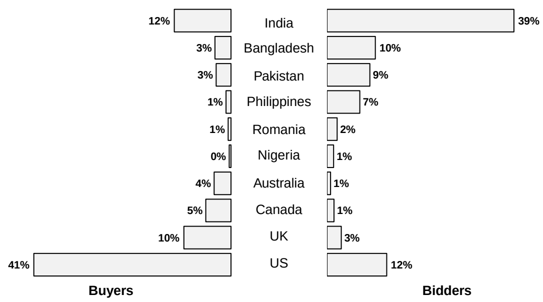

Figure 11: Distributions of countries for buyers and bidders.

hat link building posts. Using Yahoo Site Explorer [13], we checked the first 1,000 inlinks (restricted by the API) pointing to each URL. Then, we filtered URLs with more than 1,000 inlinks remaining (i.e., not retrievable via the Yahoo API), yielding 813 sites. Table 8 shows the top domain names for the inlinks. As expected, Blogspot and Wordpress are highly targeted for link spamming. Yahoo Answers and Groups, as well as Google Knol and Google Sites, are also targeted.

5 User Analysis

We end our investigation of Freelancer activity by surveying the geographic demographics and job specialization of Freelancer users.

5.1 Country of Origin

There are clear demographic differences between buyers and bidders. Figure 11 shows the distribution of countries of origin for all buyers and bidders of the abuse-related jobs categorized in Table 2. (The distribution for selected workers closely follows the overall bidder distribution.) We extract the country of origin for users from their profile information. We note that this information is self-reported and nothing prevents users from being dishonest; further, we have seen instances where buyers post jobs specifically avoiding bidders from India, for instance, providing a potential motive for dishonesty. Numbers for such countries are therefore a lower bound.

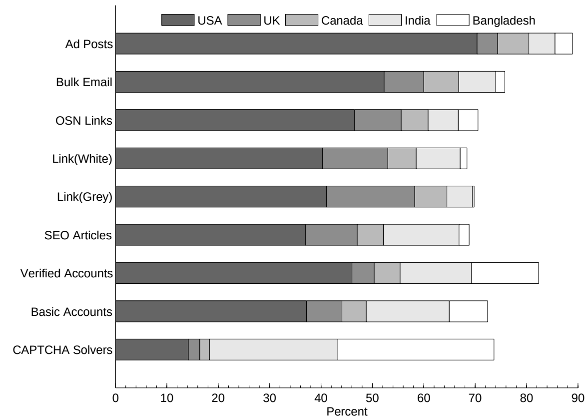

Figure 12: Top five countries of buyers posting abusive jobs.

The largest group of buyers is from the United States, and other English-speaking countries feature prominently (UK, Canada, Australia, even India). In contrast, the largest group of bidders is from India, followed by neighboring Pakistan and Bangladesh—countries with a large cheap labor force, substantial Internet penetration, and where English is an official language or has widespread fluency.

The country of origin demographics for each category reveals yet more detail. Figures 12 and 13 show the top five countries of buyers and bidders, respectively, for each abusive job category in Table 2. Buyers for advertisement posting (generally targeting Craigslist, Section 4.3.2) are primarily from the United States, whereas, somewhat surprisingly, buyers for human CAPTCHA solvers are primarily from Bangladesh and India—these are buyers looking to form teams of solvers. Bidders from India and Bangladesh dominate white hat and social networking link building jobs, respectively. Bidders from the only Western country (US) in the top five target article generation, creating PVAs, and advertisement posting.

5.2 Specialization

Aside from some uniform basic fundamental requirements, such as understanding English and having access to and basic knowledge of the Internet, the abuse jobs posted on Freelancer essentially require unskilled labor. As a result, Freelancers need not necessarily specialize—focus solely on a particular job category—in the tasks that they undertake.

As one metric of whether specialization occurs or not, we examined whether buyers and bidders participated in more than one category of job (for those buyers and bidders who engaged in more than one job). Indeed, bidders clearly do not specialize. For all but one category, on average fewer than 5% of the jobs that bidders bid on are within the same category; the exception is article content

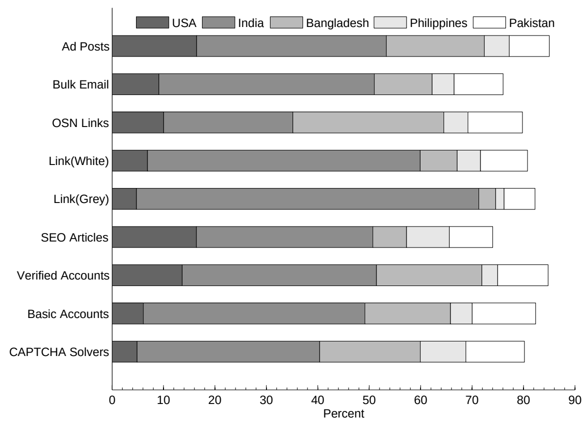

Figure 13: Top five countries of bidders on abusive jobs.

generation, where nearly 15% of bids per bidder are on other article jobs. Moreover, not only are most bids on other job categories, but the majority of bids are on jobs that did not even fall into an abuse category in Table 2. In other words, for bidders who bid on at least one abuse job, 70–80% of their other bids were for a non-abuse job.

Buyers follow a similar pattern as bidders, but are slightly more focused: 10% of a buyer's jobs, on average, are for jobs in the same category, while 60–70% of a buyer's jobs were for a non-abuse job. Article content generation again is the one exception, with 30% of a buyer's jobs requesting articles.

6 Discussion

Figure 14 illustrates how the various markets described in this study fit together in the Web abuse chain. At the lowest level, workers need access to Web proxies (due to account registration limits placed on IP addresses), CAPTCHA solvers/OCR packages, and phone numbers. Utilizing these components, abusers can create Web-based email accounts, the primary building blocks for service abuse. The email accounts can be used to register accounts for a number of Web services, including Craigslist, Facebook, Twitter, Digg, etc.

The abusers can then implement various monetization schemes with the accounts, most of them involving "spamming". The most direct form of spamming utilizes the Web email accounts to send spam. Craigslist PVAs allow abusers to post repeated, daily advertisements, making a retailer's product consistently appear near the top of the search results. Abusers can use social networking accounts in several ways, the most direct involving the creation of social links (fan, friend, follower, etc.) for marketing purposes.

The relationship between this ecosystem and SEO is subtle: the accounts on social networking sites can also be used for SEO purposes. For example, abusers may spam blogs with comments that link to a Web page to ob-

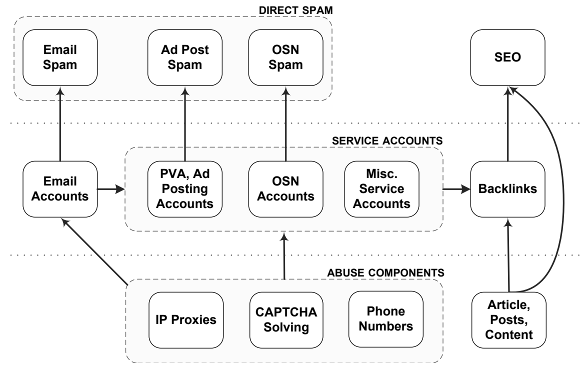

Figure 14: How the various elements of the market fit together

tain more backlinks for the site. Abusers may also submit links to social bookmarking sites, or utilize forum accounts to create posts containing links (most often in the signature field). Many of these SEO jobs require content, either in the form of articles, or actual content to include in blog comments or forum posts. Lastly, abusers can also directly purchase backlinks on sites.

7 Conclusion

This paper demonstrates how web service abuse can be augmented by the use of low-cost freelance labor. Seven years of historical data have allowed us to collect information on abuse-related work on freelancer.com, one of the largest online websites offering piecework labor outsourcing. Potential employers offered jobs such as link building on social network sites, mass email account creation, and tasks related to search engine optimization. In addition, we found that the demand for freelancers to fill these jobs is being matched by an increase in the number of freelancers around the world who will compete for the work.

Freelancer.com, and other sites that offer freelance jobs and employment are prime sources of new types of service abuse. The willingness of many freelancers to take part in these schemes allow those who offer the jobs to quickly ascertain new schemes and their success rate; if they are judged to be profitable, the jobs quickly become a staple income for the willing freelancer and thus, the employer. Services developed by experts to ensure the security of websites, such as CAPTCHA technology, are now targeted by employers who hire freelancers to break encoding and circumvent the site's security measures. These trends point to the need for anti-abuse fortifications that will defend against attackers who have a workforce of virtually unlimited knowledge at an inexpensive price. [10]

[10] The conclusion of this paper is an example of *article rewriting*: modifying text to pass plagiarism detection systems like CopyScape, commonly as a means of producing high-quality SEO content. The original text, given to the freelancer, is given below:

Acknowledgments

We would like to thank the anonymous reviewers for their feedback, Qing Zhang for the cosmetic Web sites, and Do-kyum Kim and Lawrence Saul for helpful discussions on job classification. This work was supported in part by National Science Foundation grants NSF-0433668 and NSF-0831138, by the Office of Naval Research MURI grant N000140911081, and by generous research, operational and in-kind support from Yahoo, Microsoft, Google, and the UCSD Center for Networked Systems (CNS). McCoy was supported by a CCC-CRA-NSF Computing Innovation Fellowship.

References

[1] Crowdflower. http://crowdflower.com/.

[2] Data entry assistant. http://www.dataentryassistant.com/.

[3] Esp game. http://www.espgame.org/gwap/.

[4] J. Buckmaster. Phone verification in erotic services. http://blog.craigslist.org/2008/03/phone-verification-in-erotic-services, March 2008.

[5] Copyscape.com. http://www.copyscape.com/.

[6] Freelancer.com. http://www.freelancer.com/info/about.php.

[7] 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*, CCS '10, pages 27–37, New York, NY, USA, 2010. ACM.

[8] P. G. Ipeirotis. Analyzing the Amazon Mechanical Turk Marketplace. *XRDS: Crossroads*, 17:16–21, Dec. 2010.

[9] T. Joachims. *Making large-scale support vector machine learning practical*, pages 169–184. MIT Press, Cambridge, MA, USA, 1999.

[10] J. P. Kincaid, R. P. Fishburne, R. L. Rogers, and B. S. Chissom. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. Naval Technical Training Command Research Branch Report 8–75, February 1975.

[11] C. Kreibich, C. Kanich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamcraft: An Inside Look at Spam Campaign Orchestration. In *Proceedings of the USENIX Workshop on Large-scale Exploits and Emergent Threats (LEET)*, Boston, MA, Apr. 2009.

[12] M. Motoyama, K. Levchenko, C. Kanich, D. McCoy, G. M. Voelker, and S. Savage. Re: CAPTCHAs — Understanding CAPTCHA-Solving from an Economic Context. In *Proceedings of the USENIX Security Symposium*, Washington, D.C., Aug. 2010.

[13] YahooSiteExplorerAPI. http://developer.yahoo.com/search/boss/boss_guide/site_explorer.html.

[14] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao. SybilLimit: A Near-Optimal Social Network Defense against Sybil Attacks. In *IEEE Symposium on Security and Privacy*, pages 3–17, 2008.

[15] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. SybilGuard: Defending Against Sybil Attacks via Social Networks. In *Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM '06, pages 267–278, New York, NY, USA, 2006. ACM.

[16] Q. Zhang, T. Ristenpart, S. Savage, , and G. M. Voelker. Got Traffic? An Evaluation of Click Traffic Providers. In *Proceedings of the WICOM/AIRWeb Workshop on Web Quality*, 2011.

A Interesting Jobs

This appendix includes representative real jobs posted to Freelancer from all the job groups. These examples provide context and help to clarify the various legitimate and dirty job categories.

A.1 Legitimate

Private. project has already be awarded to <...>. thanks

Legitimate Miscellaneous. I have a simple document for translation from Dutch to English. Those who are available for immediate start and freelancers only apply.

A.2 Accounts

Human CAPTCHA Solving. PixProfit.com is the portal for data-typist. We're looking for individuals or team of data-entering workers. We'll pay from $1 for 1000 correctly typed images.

Phone Verified Accounts. We are looking for a reliable provider of new CL Phone Verified Accounts(PVA).Will be buying up to 1000-2000/month. Willing to pay no more than $2.00/PVA Or best offer.

A.3 SEO

SEO Content Generation.
I need 20 articles written about penis enlargement and 40 articles written about male enhancement. The total is 60 articles with the following requirements. Your writing must be your own original work (no article spinning). Length 500-600 words per article. Written in excellent english with perfect grammar. Keyword density of 2%.

SEO Spinning Article. I am looking for native content providers to provide me articles with spinner syntax.

In this paper we document how low-cost freelance labor enables Web service abuse. Using historical data spanning over seven years, we survey the market for such abuse-related work on Freelancer.com, a popular online market for piecework labor outsourcing. We found a broad range of such activities, including mass account creation, SEO-related tasks, and social network link building. Moreover, we witnessed a steadily increasing demand for such services matched by a highly competitive world-wide labor force.

Freelance labor markets like Freelancer.com serve as an incubator and catalyst for new kinds of service abuse. Such a general labor pool allows nascent abuse schemes to be prototyped and evaluated quickly, and, if ultimately profitable, leads naturally to the efficient commoditization of the requisite services. Mature services, such as CAPTCHA solving, eventually evolve into standalone services capable of meeting growing market demand [12]. Modern anti-abuse defenses must, in the end, contend with sophisticated attackers having a versatile and inexpensive labor force at their disposal.

Something like this : {Deciding||Determining} in what {type||kind||sort} of credit card to {apply||go for||lend oneself||put on||employ} for {depends||counts||reckons} on your {past||previous||recent||former} credit {history||account||report||theme}. Providers without prior spinning knowledge, Please don't bid. I will pay 1.5 USD per spun article to start with only through Paypal.

Link Building/Grey Hat. I am looking to outsource large numbers of blog commenting. Quality blog commenter needed. Can provide 1000 comments per week upwards. This will be for a trial of 100-200 comments per week.

Link Building/White Hat. 100 Gambling Links from related PR 4 or higher pages. All on different sites and servers Requirements: No link farms, link-exchange programs, No black hat links or Tricks.

SEO Miscellaneous. keyword : trader joes
website : will mention via message
SE : google.com
i wan't my website rank 1 in google.com. If interested pls send detail what is your skill to get this website top on google.com

A.4 Spamming

Human Oriented Postings. I need per day 2K Classified Ad Posting for my site I willings to pay for it $100. Per ad $0.05

A.5 OSN Linking

Create Social Networking Links. I am lonely I want to give my facebook account details to someone and have them populate it with 5000 English speaking friends help me please.

A.6 Miscellaneous

Abuse Tools. The first tool necessary is Micro Niche Finder. You will need this to do keyword research, and select keywords based on our requirements. The tool will also allow you to see which keywords have .com, .org, or .net domains available. Once the available domains have been determined, we will review your picks, and purchase them after approval. Once purchased, you will need to create articles for each page, and install the necessary wordpress theme and plugins. Once this is complete, you will need to run SE Nuke or Evo II for each site, at least 4 times per month.

Academic Fraud. For this project, you will put together several techniques and concepts learned in CS <deleted > and some new techniques to make an application that searches a large database of people which we will call a Personal Information Manager (PIM), even though it only contains a few fields, and even fewer advanced functions. This project creates a simple program that allows people to enter names or email addresses and check whether they are found in the PIM.

Account Creation Tools. Hey all! I'm in need of US telephone numbers with call forwarding for CL PVA creation. Please quote your rate. Bids lower or equal to $1 will be given higher priority.

Other Malicious. Hello, I have a small sized EXE file of 40KB and I need someone who can build a script who will DOWNLOAD AND EXECUTE the EXE file AUTOMATICALLY. What I mean by automatically? By entering a single URL in the browser.
Here is a PERFECT example: http://www.<deleted>.com. In the example above the EXE file is EXECUTED even when you click on CANCEL in the javascript prompt screen.

B Interesting Bids

This appendix includes representative real bids received from Freelancer workers from some abuse job groups. These bids shed light into the various tools and techniques used by workers to circumvent Web security mechanisms. Also, the bids provide some insight into worker demographics.

B.1 Accounts

Account Creation. 1 Account create on 1 ip, Cookies/Cache is cleared after every account automatically. All accounts are created using real human names. We have the ability to provide accounts as per your required format. ——— We created those account with this requirement as below: 1) All Gmail accounts created with unique US IP Addresses 2) All Gmail accounts created separate/unique passwords 3) All accounts created a prefix with names &/or words. Preferably no numbers 4) All accounts to have random First and Last names assigned. 5) All passwords have minimum of 8 characters and preferably alpha-numeric

B.2 SEO

SEO Content Generation. Hi!I am <deleted>.I am currently a stay at home mom with 9 month old daughter so I currently have free time throughout the day. I can write quality articles/blogs, academic research papers and LSI/SEO written content of any nature.These articles are put through Copyscape premium dupe test before submission. Also find attached a sample News article I did for a local News paper.I assure you that your articles will be written in the most professional manner possible. I charge $1 per 100 word.I look forward to working with you.Take care

B.3 Spamming

Create Social Networking Links. Techniques:(100% white hat) 1. Following people manually: Twitter let us follow 500 people in a day and maximum 2000 follow using one account. So i found a nice technique by which i am able to make 1000 follower. That is
#First follow huge people manually up to 500 using an account similar to your account and after following 500 i will receive a massage, "You have cross the hourly limit . You cant follow now". Then i will use another account to follow targeted follower up to 500...

B.4 OSN Linking

Human Oriented Postings. I am experienced with the CL posting .Now i am working use for daily posting (RDSL With AT@T Line ,CLAD Soft, Ip rental,Proxy,AOL,US hide IP, line with Logmein soft Or,Team Viewer & go to my PC), We have so much experience a team for all adds posting site such as craigslist, backpage, kijiji, gumtree, olx, oddle and all classified site) also have all requirements which need your project done.