

Measuring the Perpetrators and Funders of Typosquatting

Tyler Moore¹ and Benjamin Edelman²

¹ Harvard School of Engineering and Applied Sciences, tmoore@seas.harvard.edu

² Harvard Business School, bedelman@hbs.edu

Abstract. We describe a method for identifying “typosquatting”, the intentional registration of misspellings of popular website addresses. We estimate that at least 938 000 typosquatting domains target the top 3 264 .com sites, and we crawl more than 285 000 of these domains to analyze their revenue sources. We find that 80% are supported by pay-per-click ads, often advertising the correctly spelled domain and its competitors. Another 20% include static redirection to other sites. We present an automated technique that uncovered 75 otherwise legitimate websites which benefited from direct links from thousands of misspellings of *competing* websites. Using regression analysis, we find that websites in categories with higher pay-per-click ad prices face more typosquatting registrations, indicating that ad platforms such as Google AdWords exacerbate typosquatting. However, our investigations also confirm the feasibility of significantly reducing typosquatting. We find that typosquatting is highly concentrated: Of typo domains showing Google ads, 63% use one of five advertising IDs, and some large name servers host typosquatting domains as much as four times as often as the web as a whole.

1 Introduction

At the dawn of commercial Internet activity, aggressive website registrants discovered that they could profit by registering domain names matching others’ company names, product names, and trademarks – “cybersquatting,” as the practice came to be known. Initially, cybersquatting promoted competitors, as in Princeton Review’s 1994 registration of kaplan.com to divert Internet traffic intended for a competing test preparation service. Once domain names started requiring annual renewals, squatters raced to grab domain names when the prior registrant failed to renew [3]. By 1999, squatters began “typosquatting” – intentionally registering misspellings of popular websites in anticipation that users mistype those domains and reach squatters’ sites [5].

Cybersquatters have employed several strategies to profit from their registrations. After grabbing particularly valuable domains, some squatters sought small ransoms from the organizations that most wanted those domains. In one notorious case [3], a squatter redirected thousands of expired domains to adult websites, making it all the less palatable to leave the domains with the squatter, and all the more tempting to pay to get the domains back. Separately, following

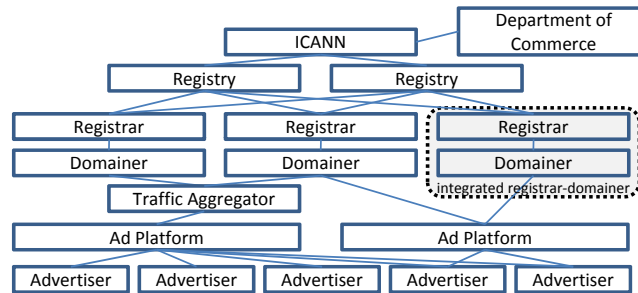


Fig. 1. Diagram of relationships.

in Princeton Review’s footsteps, other squatters redirected squatting domains to direct competitors of the sites users expected. Through such redirects, a squatter could either profit directly (if the squatter also operated the destination site) or indirectly (through marketing commissions paid by the destination site). Finally, a growing share of squatters found profits through advertising – typically, showing pay-per-click ads through the web’s top ad networks.

As the Internet matured, cybersquatting domain registrations came to be viewed as both disruptive and improper. In 1999, countermeasures to squatting were introduced, including an arbitration procedure (the UDRP) and a new federal law (the ACPA), both discussed in Section 6.1. The subsequent decade featured more than 45 000 UDRP disputes and more than \$40 million of ACPA damage awards. Yet our paper shows that cybersquatting and especially typosquatting remain widespread.

In this article, we explore modern typosquatting. Our methodological contribution is the development of a software system that effectively identifies typosquatting using telltale patterns in domain registrations and configurations. Our substantive contribution is a characterization of the typosquatting problem, including estimating its size, assessing who is responsible, and identifying factors that put some brands, marks, and domains at heightened risk of typosquatting.¹

2 Structure and Strategy of the Domaining Business

Most large domain registrants present themselves as “domain parkers” or *domainers*. Figure 1 outlines the relationship between domainers, advertisers and Internet authorities. Domainers submit registration requests to *registrars* (e.g., GoDaddy), which provide domain names after confirming availability with *registries* (e.g., VeriSign), which in turn are authorized by the Internet Corporation for Assigned Names and Numbers (ICANN) which coordinates certain Internet identifiers under contract with the United States Department of Commerce. At each step, money changes hands: domainers pay registrars which pay registries, and both registrars and registries pay fees to ICANN.

¹ <http://www.benedelman.org/typosquat/> details data collected for the paper.

As Figure 1 depicts, the domaining business can feature numerous variations. For example, a large domainer may elect to become a registrar – simplifying relationships and eliminating an intermediary. In 2006 litigation, Neiman Marcus alleged exactly that in a 155-page complaint claiming that Dotster (a large domain registrar) hoarded typosquatting domains for its own benefit, failed to disclose domain ownership via Whois records, and improperly extracted fees from companies wanting domains Dotster had collected [9]. There is also variation in the structure of domainers’ relationships with *ad platforms* (e.g., Google AdSense): Large domainers typically work directly with ad platforms, while smaller domainers typically work through traffic aggregators (e.g., `information.com`) that combine traffic from many domainers.

The domain parking business is premised on users arriving at parking sites. But *why* do users go to parking sites? Some users seem to type in domain names randomly, rather than using search engines to find the materials that meet their requirements – requesting a site like `discounthoteldeals.com` when seeking “discount hotel deals.” Such users might end up at domain parking sites matching the generic keywords that embody their requests. But there is another way for users to arrive at parked domains: misspelling the address of a more popular site. This practice, typosquatting, is the focus of our paper.

3 Measuring Typosquatting

3.1 Identifying Typosquatting Domains

The first step in studying typosquatting is to identify which domains are similar enough to be deemed typos. We start by gathering a list of *popular domains* that could tempt a squatter to register many *typo domains*. For this paper, we decided to study the 3 264 `.com` domains at least 5 characters long appearing in the most popular 6 000 domains according to Alexa’s June 29, 2009 ranking. We focus on the most popular sites because, all else equal, popular sites are more likely to be targeted for typosquatting: The more users seek to visit a domain, the more users are likely to mistype the domain’s address. (Section 5 empirically examines the factors affecting typosquatting prevalence.) We excluded very short domains (4 characters or less) because even one-letter variations may reflect intentional requests for other short domains, rather than typos of a popular domain. Finally, we only consider `.com` domains due to `.com`’s ubiquity and because the zone file listing all `.com` domains is publicly available.

Next, we generated a list of plausible misspellings of the 3 264 popular domains. To identify plausible misspellings, we rely on the Damerau-Levenshtein distance [2,6]: the minimum number of insertions, deletions, substitutions or transpositions required to transform one string into another. For example, `faceboolk`, `facebok`, `faceboik`, and `faceboko` each have a Damerau-Levenshtein distance of 1 from `facebook`. Damerau found that 80% of spelling errors are caused by one such operation. We also created a list of typos with `www` and `com` appended to the start and end of the strings, respectively. These appendages help recognize frequent mistakes arising out of omitting a ‘.’ when typing a URL.

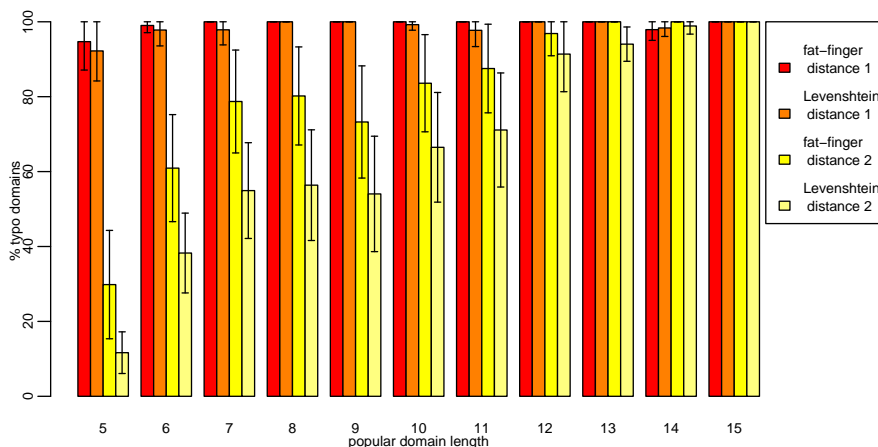


Fig. 2. Typosquatting classification accuracy.

We also devised a new measure of string distance useful for keyboard misspellings, called *fat-finger distance*: the minimum number of insertions, deletions, substitutions or transpositions *using letters adjacent on a QWERTY keyboard* to transform one string into another. For example, `facebojk` has a fat-finger distance of one from `facebook`, since ‘j’ is next to ‘k’ on a standard keyboard.

To identify typosquatting domains, we enumerated all strings with a Damerau-Levenshtein (and fat-finger) distance of up to 2 from each of the 3 264 popular domains. This captured all plausible one- and two-letter typos of popular domains. We next intersected this set with the nearly 81 million registered `.com` domains (according to the `.com` zone file). This process yielded 1 910 738 registered `.com` domains as *candidate* typo domains of the 3 264 popular domains.

We manually checked a sample of 2 195 domains randomly selected from the list of 1.9 million candidates. To form this sample, we selected candidate domains targeting popular domains of length 5–15, allowing variations of Levenshtein distances of 1 and 2 and with fat-finger distances of 1 and 2. Figure 2 plots the results with 95% confidence intervals. (In particular, we have 95% confidence that the true number of `.com` domains up to a Levenshtein-Damerau distance of two from the popular domain lie between the numbers posted in the table. Of course, our analysis omits typo domains of distance ≥ 3 and also typos in other top-level domains such as `.cm`.) No matter the length of the popular domain, typo domains within Levenshtein or fat-finger distance 1 of popular domains were overwhelmingly confirmed as typos. When we consider typos of distance 2 from popular domains, false positives become more frequent. However, domains within fat-finger distance 2 of popular domains are more likely to be typos than domains within only Levenshtein distance of 2. Furthermore, for increasingly lengthy popular domains, it is increasingly likely that domains with a fat-finger or Levenshtein distance of 2 are in fact typo domains.

With false positive estimates from our manual checks, we estimated the number of `.com` typo domains targeting the popular domains we identified. To do

popular domain	candidate typo domains	point estimate typo domains	95% confidence interval
google.com	5 731	2 537	(1 728, 3 252)
youtube.com	3 616	2 069	(1 589, 2 534)
myspace.com	3 482	1 960	(1 457, 2 440)
freecreditreport.com	1 904	1 904	(1 904, 1 904)
hotels.com	4 465	1 865	(1 207, 2 442)
total for 3 264 domains	1 910 738	937 918	(710 872, 1 236 924)

Table 1. Selected domains highly targeted by typosquatting.

so, we added up the number of candidate typos matching each popular domain; then we adjusted each candidate’s weight based on our confidence in its accuracy in light of the typographical distance between the typo domain and the popular domain. By this methodology, we estimate that approximately 938 000 typo domains target variations of the 3 264 popular domains we studied. On average, each popular domain is targeted by 281 typo domains, but some domains attract more typosquatting than others. Table 1 lists the domains that are most targeted by typosquatting. Topping the list is `google.com`, for which we found an estimated 2 537 typo domains. That said, as we show in Section 4.1, Google also supports typo domains by providing both technical assistance and advertisement payments.

3.2 Crawling Typosquatting Websites

We cannot easily visit all 938 000 typo domains without also visiting many sites that are not typos. Because we wish to learn more about only typosquatting websites, we decided to crawl only a subset of the typo domains where we know the vast majority are in fact typos. To that end, we developed a crawler to visit the 284 914 typo domains where the Damerau-Levenshtein distance between typo and popular domains is at most one for popular domains between five and nine characters in length, and a distance of up to two for popular domains at least ten characters long. Consequently, the totals discussed in the subsequent sections should be interpreted as a sample of the larger typosquatting population.

The crawler explored each typo site and its links to determine how a given site is being used, and the crawler recorded all HTML and headers that it received. The results of our crawl are presented in Sections 4.1 and 4.2.

We designed the crawler to avoid burdening websites or advertisers. The crawler follows three randomly-selected links on each page, up to (at most) a depth of three links. Since indiscriminately following pay-per-click links and redirects could yield unwarranted cost to advertisers and unearned revenues to squatters, the crawler only invokes a link after comparing that link to a list of known ad servers. If a link references a known advertising domain, the site is marked as containing advertisements, the link is recorded, and the crawler proceeds no further at that site. The same logic is used whenever the crawler

classification	typo domains	%
pay-per-click ads	74 024	79.4
Google	53 364	57.2
Yahoo!/Overture	19 145	20.5
Ask.com	555	0.6
Miva	541	0.6
Enhance	297	0.3
domain redirection/link	19 227	20.6
self-registration	4 133	4.4
affiliate marketing	10 215	11.0
redirect or link to competing site	4 879	5.2
blocked	124 211	–
unclassified	70 729	–

Table 2. How typo domains are used.

encounters a HTTP redirect. The crawler assures that at most one of its threads visits a single server (at a single IP address) at a time.

Although our crawler could not retrieve and classify all the typo domains it identified, we believe the classified domains provide appropriate insight into the usage of the other domains. For one, 131 637 of the 194 940 blocked or unclassified domains share the same IP address and name server with domains where we confirmed the presence of PPC ads.

Our crawler found that 124 211 domains blocked its efforts at inspection. Some servers host tens of thousands of typo domains. Despite our crawler’s best efforts, such servers may notice after a machine from a single IP address tries to visit several links on thousands of domains. In manual tests, we confirmed that a few of these domains were truly unreachable. But often these “blocked” domains reside on IP addresses that had previously responded as expected, and often these “blocked” domains loaded as expected when tested from another IP address. We therefore conclude that at least some typosquatting hosts recognized our crawler’s examinations and took steps to prevent our analysis.

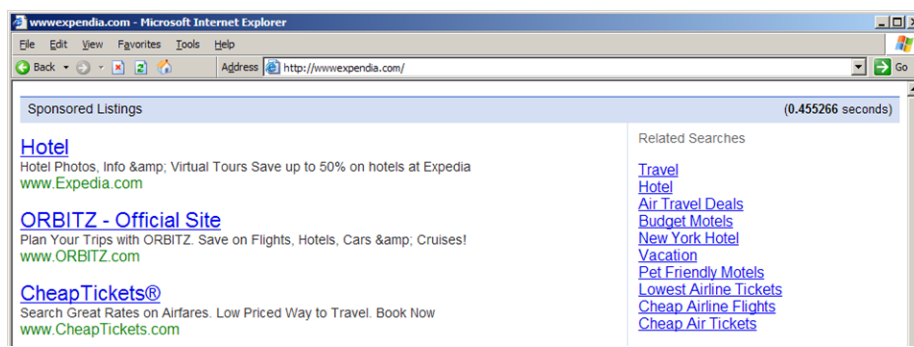
In addition to sites that affirmatively blocked our analysis, our crawler was unable to classify a further 70 729 domains. Many of these domains included JavaScript links, which our crawler could not reliably follow.

4 How Typosquatting Domains are Used

From crawling typosquatting pages, we confirmed two main uses for traffic diverted to typo domains: placing pay-per-click ads and redirecting to other (often competing) domains. We discuss our findings for each strategy in turn.

4.1 Squatter Strategy 1: Pay-Per-Click Ads

By far the most common use for typo domains is displaying pay-per-click advertisements. Of the typo domains our crawler could classify, Table 2 reports that 80% – over 74 000 – included pay-per-click ads. Most of these websites –



self-advertising rate	popular domains	examples
$\geq 75\%$	18	papajohns (90%), saksfifthavenue (88%)
$50\% \leq \dots < 75\%$	58	expedia (50%), t-mobile (70%)
$25\% \leq \dots < 50\%$	106	wellsfargo (43%), businessweek (48%)
$< 25\%$	81	findlaw (21%), tigerdirect (22%)
overall: 36%	263	

Fig. 3. wwwexpedia.com shows ads for expedia.com and competitors (top); self-advertisement prevalence for 263 popular domains buying ads (bottom).

at least 53364 – partnered with Google to sell ad space to advertisers, select which ads to display, track clicks, and collect payments, among other functions. Google’s prevalence in part reflects Google’s large market share in pay-per-click advertising, and Google further benefits from its development of an advertising service dedicated to placing ads onto parked domains.² Next-largest after Google is Yahoo; we found Yahoo ads on at least 19145 typo domains. We detected three additional PPC ad providers also being used, but with dramatically lower prevalence, as detailed in Table 2.

Figure 3 (top) shows PPC ads on wwwexpedia.com. The top advertisement promotes expedia.com, the same domain misspelled in the user’s request. The Expedia ad appears because Expedia pays Google to advertise on websites with “relevant” content, and Google’s algorithm select wwwexpedia.com as a suitable place for those ads. Consequently, Expedia pays Google whenever a user misspells Expedia and clicks the sponsored link to Expedia [4]. Meanwhile, immediately below Expedia are advertisements for competitors Orbitz and CheapTickets. Had Expedia chosen not to pay Google to place ads on parked domains, Google would have shown links only to competing sites.

We found self-advertisements on typo domains targeting 263 popular websites that bought ads (Figure 3 (bottom)). Sometimes, nearly all typo domains included ads to the popular site (e.g., 90% for typos of papajohns.com). For others, self-advertising occurred less often (e.g., 22% for tigerdirect.com). Overall, we saw ads corresponding to the popular domain on 36% of typo domains.

² See Google AdSense for Domains, <http://www.google.com/domainpark>.

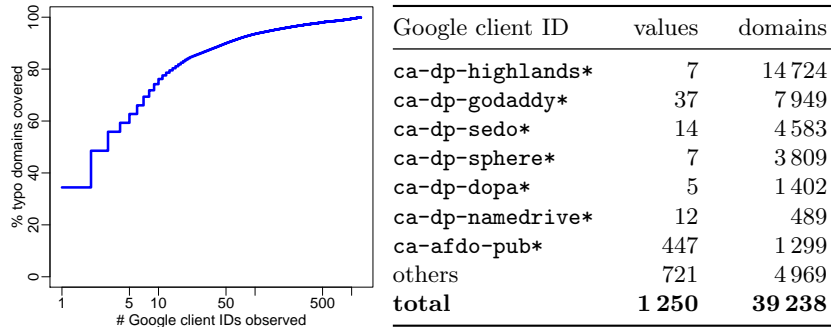


Fig. 4. Advertising client IDs matching typo domains.

While stopping so many typo domains may seem like a Sisyphean task, we found considerable concentration upon closer inspection of PPC ad links. An ad provider needs to know who to pay for a given advertisement placement, so a partner ID is passed as a parameter in an ad’s click URL. For example, in the link <http://domains.google syndication.com/apps/domainpark/results.cgi?client=ca-dp-mborin&...> on `cartoonntewrok.com`, the `client` parameter is set to `ca-dp-mborin` – indicating Google will pay the corresponding partner if a user clicks that ad link. On other domains, these parameters appear in HTML entity encoding, in redirects, and/or in JavaScript variables.

We found partner ID codes for 74% of typo domains showing Google ads. While 1 250 different codes were found, some turned up disproportionately often. Figure 4 (left) plots the cumulative distribution of typo domains by partner ID (note the logarithmic x-axis). The top 5 partner IDs cover 63% of the Google typo domains we explored, and the top 10 cover 76%. The most frequent partner ID was `ca-dp-highlands19_3ph_xml`, appearing in ad links on 13 542 typo domains. The table in Figure 4 lists specific partner ID we observed particularly frequently on typo domains. Large domainers and traffic aggregators often have recognizable Google IDs, such as `ca-dp-godaddy` (GoDaddy), `ca-dp-namedrive` (NameDrive), `ca-dp-sedo` (Sedo) and `ca-dp-namesphere` (NameSphere).

Unfortunately, we could not identify partners from Yahoo ads, since the ads use a single parameter `xargs` presenting a lengthy obfuscated string apparently combining ad destination, partner, and more. While we cannot determine which Yahoo partner receives credit for a given placement, we can still demonstrate high concentration among Yahoo partners. For example, one typosquatter passed Yahoo PPC links as a parameter within redirect URLs with the distinctive parameter `provider` set to 1200. This same pattern is found on 10 446 typo domains, nearly all using the same name server and IP address.

4.2 Squatter Strategy 2: Redirection and Linked Domains

Rather than showing pay-per-click ads, other typo domains redirect or link to predetermined destination domains. We saw three practices in this vein: (i) self-registrations/defensive registrations, (ii) affiliate marketing and (iii) redirect or link to competing site.

Self-registrations / defensive registrations. In some instances, a company will “self-register” misspellings of its key domains. Often, requests for these typo domains redirect a user to the company’s main site, where the user likely intended to go. Through self-registrations, a company can avoid unwarranted marketing expense, such as paying for ads on typo sites, as shown in Figure 3. Self-registrations also help users reach their intended destinations without extra clicks or delays. We found 4 133 typo domains that match this profile, in that they share the same name servers as the popular domains of which they are variants.

Affiliate marketing. Through merchants’ affiliate marketing programs, some typosquatters send users to the sites users intended to visit – but charge the merchants a fee for providing these referrals.

In affiliate marketing, advertisers pay for referrals on a performance basis: Send a user to Dell and Dell will pay a commission of 2% or more. Because affiliate merchants generally only pay when a user makes a purchase, many merchants fail to supervise their affiliates’ specific promotional methods. Few affiliate merchants affirmatively allow typosquatting, and most disallow it when it comes to their attention. But to date, few merchants have uncovered affiliates engaged in typosquatting. (In a rare exception, Lands’ End sued several squatters who registered typosquatting domains and redirected resulting traffic to Lands’ End affiliate links [7].)

We saw 10 215 typosquatting domains that linked or redirected to the corresponding popular site, where the name server used by the squatting site and the popular domain differed. We checked all redirections from the misspelled domain name, looking for redirections to the popular domain via an affiliate marketing network. We confirmed 2,697 domains redirected to affiliate marketing networks, including 905 typo domains promoting Commission Junction merchants, 652 promoting LinkShare merchants, and 290 promoting Performics (Google Affiliate Network) merchants. Another 4 629 redirected to the legitimate domain, either as a result of defensive registration or for directly managing affiliates (e.g., `bookihng.com` redirects to `booking.com/?aid=311266;label=11-booking-promo`).

Redirects or links to competing site. When users attempt to visit a popular site, some typosquatters instead forward the users to a competing site – often in the same industry, but a notch less popular. For example, `pict.com` is a relatively little-known document sharing site (Alexa rank 8 581 as of Aug 27, 2009). But `pict.com` is redirected to by typos of 128 competing, more popular sharing sites – 24 typos of `depositfiles.com` (Alexa rank 167), 22 typos of `picoodle.com` (Alexa rank 5 040), 18 typos of `sharebee.com` (Alexa rank 1 673), and more. These redirects take users directly to `pict.com` with no link codes or partner IDs of any kind – suggesting that `pict.com` itself registered these domains and that, in any event, `pict.com` is probably not paying partners for this traffic. Similarly, we found 156 typo domains that are variations of `yellowpages.com`, which all redirect to the website `yellowpagesoftheworld.com`.

We developed a simple heuristic to identify typo domains linking to competing domains. First, we group typo domains that all link to the same *beneficiary domain* (e.g., `pict.com`, a domain benefiting from this group of typo domains).

yellowpagesoftheworld.com: 158 typo domains	
yellowpages.com: yellopasges, yeollwpages, yelkowpages & 153 more	
whitepages.com: whigtetpages & whitepagecom	
bet365.com: 367 typo domains	
sportsbook.com: saportsbook, sxportsbook, sportszbook & 325 more	
betcllic.com: betcllico, betcllicm, betcllicj & 7 more	
fulltiltpoker.com: fulltilt6poker, fuylltiltpoker, fulltiltpoke4r & 5 more	
pict.com: 128 typo domains	
depositfiles.com: dopskfiles, depositfimes, depositciles & 21 more	
picoodle.com: picoodke, picoodme, piciodle & 19 more	
sharebee.com: shaerbee, shafebee, shatebee & 15 more	
movietheatertickets.biz: 85 typo domains	
movietickets.com: movietikits, mpvietickets, muvietickets & 19 more	
rottentomatoes.com: rottentomaos, rottentmoatoes, rotentomatoe & 10 more	
fandango.com: fansango, fandnango, faneango & 9 more	
<hr/>	
total: 75 beneficiary domains on 4 879 typo domains targeting	
668 competing popular domains	

Table 3. Example domains linked to by typo variations of competing domains.

Next, we consider only those beneficiary domains that are linked by typo domains targeting a *small number* of popular domains. By focusing on beneficiary domains receiving traffic from typos on a small number of popular domains, we identify beneficiary domains that are targeting typosquatting on specific popular sites (typically, in the same sector), rather than aggregating typo traffic more generally. Through testing, we adjusted the parameters, and we elected to focus on beneficiary domains linked by at least 75 typo domains that target no more than 40 popular domains. Using this criteria, we identified 75 beneficiary domains that are linked from 4 879 distinct typo domains, which collectively target 668 competing popular domains.

Table 3 lists selected beneficiary domains identified using our heuristic. Notably, every beneficiary domain is linked by typos in the same category: typos of popular casino websites link to `bet365.com`, popular movie sites link to `movietheatertickets.biz`, and so on. This trend is consistent for all beneficiary domains. It is not always clear whether the beneficiary domain directly registered and configured the typo domains; affiliate marketing and similar relationships can motivate partners to register typo domains.

5 Do Pay-Per-Click Ads Promote Typosquatting?

Table 1 reveals that some popular domains are targeted by typosquatting far more than others. Why? We initially hypothesized that typosquatting disproportionately afflicts domains that are difficult to spell. To check, we regressed number of typos on popular domain spelling difficulty (with controls for various measures of popularity of the popular domain). We found no effect of spelling difficulty as measured by number of double letters or presence of adjacent i/e

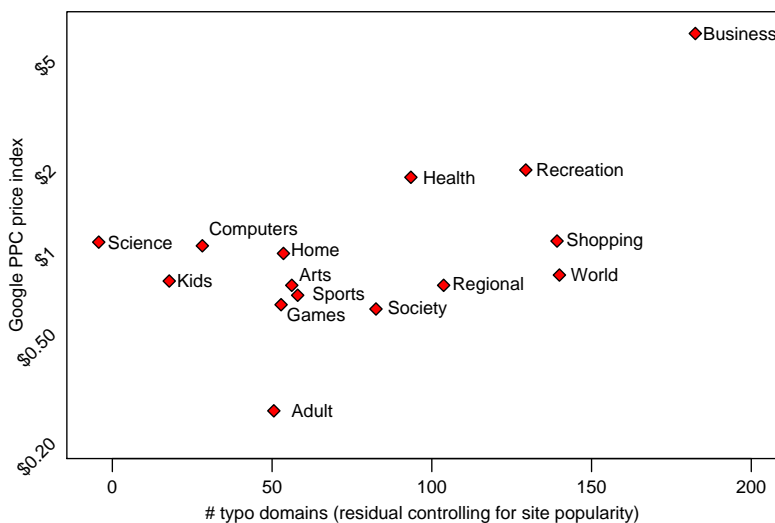


Fig. 5. Scatter plot comparing typosquatting incidence to amount paid out by pay-per-click ads.

tuples – perhaps reflecting that these top popular domains have limited variation in spelling difficulty (as measured by these proxies).

However, we do find significant differences across website categories. To assign popular domains to categories, we used Alexa’s listings of the top 500 websites for 15 different categories (e.g., Kids and Teens, Business, News). 1 075 of the popular domains we studied also appeared in one or more of Alexa’s top 500 categories. In a regression controlling for each popular site’s popularity and number of category listings, we included a fixed effect for each category, and we noted the coefficient associated with each category variable. These coefficients form the x coordinates in Figure 5. For example, the average popular domain Alexa places in “Shopping” is targeted by 143 more typo domains than the average popular domain Alexa places in “Science.”

Because most typo domains are funded by pay-per-click ads, we examined patterns in PPC pricing across Alexa categories. For each popular domain Alexa listed in each category, we extracted META keywords, and we identified the ten most frequent keywords in each category. Using the Google Traffic Estimator, we obtained minimum and maximum PPC price estimates for each frequent keyword. We formed a *Google PPC price index* for each category, given by the average of 1) the median of the minimum PPC prices for keywords in that category, and 2) the median of the maximum PPC prices in that category.

Combining Alexa’s categorizations with our PPC price index yields the result shown in Figure 5. Notice the positive association: In categories with higher PPC prices, parkers registered more typosquatting domains. We interpret this relationship as evidence that high PPC prices spur typosquatting registrations in the corresponding categories.

6 Countering Typosquatting

6.1 Existing Efforts to Regulate Typosquatting

The rise of typosquatting in the 1990's prompted a series of regulations intended to put a check on abusive domain registrations. Initially, domain registrations were challenged primarily under trademark law, common law, and the arbitration procedures specified under domain registration agreements. For example, in arbitration arising out of Princeton Review's 1994 registration of `kaplan.com`, a panel held that Princeton Review had obtained the domain in bad faith with the objective of confusing consumers and harming Kaplan's reputation, and the panel ordered that the domain be transferred to Kaplan pursuant to the registration agreement Princeton Review had accepted upon registering the domain. In the subsequent *MTV Networks v. Curry* [8], a federal court noted similarities between domain names and mnemonic telephone numbers, suggesting that existing trademark law could apply to domain names.

After half a decade chasing cybersquatters, repeat plaintiffs offered three major complaints. First, it was increasingly burdensome to pursue many infringing domains; tens of thousands of dollars of attorney time to resolve each dispute compares unfavorably to tens of dollars for squatters to register new domains. In response, ICANN developed the Uniform Domain-Name Dispute-Resolution Policy (UDRP). For a relatively small filing fee of \$1 300 to \$4 000, complainants could seek electronic adjudication of an allegedly-infringing domains.

Second, plaintiffs faced cybersquatters who failed to disclose their true names and addresses, making a traditional lawsuit hard. The 1998 Anti-cybersquatting Consumer Protection Act (ACPA) (15 USC §1125(d)) offered an alternative, allowing a plaintiff to sue a domain *in rem* – suing the domain itself, rather than the domain's registrant. Domains were found to be at the location of the relevant registrar, registry, or other domain name authority. A plaintiff could petition a court in that jurisdiction for transfer or cancellation of a disputed domain.

Finally, plaintiffs worried that cybersquatters faced skewed incentives that invited infringements. Previously, after registering an infringing domain, a typosquatter could profit from its use until a court or arbitrator ordered the domain transferred or canceled. Cybersquatters therefore faced little real downside – at most, the forfeiture of the initial registration fee and litigation costs (minimal if the cybersquatter ignored litigation). The ACPA added the threat of significant statutory damages – \$1 000 to \$100 000 per domain name (15 USC §1117(d)). The threat of such damages were to deter would-be cybersquatters.

Private plaintiffs pursued these new mechanisms to put a check on cybersquatting. Between 1999 and August 2009, complainants invoked the UDRP arbitration procedure more than 45 000 times, reclaiming domains in over 85% of disputes [1]. Meanwhile, some companies pursued ACPA claims in court. Neiman Marcus filed lawsuits against typosquatters including Dotster, Name.com, and Spot Domains. Verizon sued Chinese registrar OnlineNIC, which ignored the proceedings and suffered a \$33 million default judgment for 633 typo domains of Verizon marks [10]. Microsoft sued OnlineNIC, Maltuzi, and others, and sent

<i>Name servers > 100 000 domains</i>			<i>Name servers > 1 000 domains</i>		
name server	% typo	typos	name server	% typo	typos
<code>dnsnameserver.org</code>	4.75	19 217	<code>moniker.com</code>	61.65	910
<code>trellian.com</code>	4.47	11 962	<code>ipmanagerinc.net</code>	55.63	787
<code>hitfarm.com</code>	3.76	17 073	<code>citizenhawk.net</code>	31.88	1 766
<code>dsredirection.com</code>	3.60	59 845	<code>dexner.com</code>	18.85	375
<code>linkz.com</code>	2.98	3 765	<code>aphost.com</code>	17.96	4 244
<code>fastpark.net</code>	2.77	7 715	<code>freeredirection.net</code>	17.94	1 438
<code>above.com</code>	2.77	16 691	<code>ehostinginc.com</code>	17.89	181
<code>sedoparking.com</code>	2.51	35 216	<code>nw.net</code>	17.10	250
<code>parked.com</code>	2.48	13 993	<code>onlinednsservice.net</code>	15.09	2 844
bottom 5	:	:	plus 97 name servers above 5% typo domains		
<code>ipowerweb.net</code>	0.32	569			
<code>ipowerdns.com</code>	0.30	522			
<code>123-reg.co.uk</code>	0.26	860			
<code>abac.com</code>	0.14	248			
<code>vpweb.com</code>	0.12	127			

Table 4. Name servers with the most typosquatted domains.

hundreds of subpoenas to identify typosquatters. Meanwhile, as early as 2005, Microsoft Research documented 8 923 typo domains (Internet-wide, not just for Microsoft marks) and noted how many typo domains showed PPC ads [12].

Despite thousands of complaints against typosquatting, the problem remains. 45 000 UDRP complaints represents less than 5% of the currently active typosquatting sites we found. Remarkably, even vigilant companies remain highly targeted. Months after its widely-reported judgment, Verizon still suffers at least 767 typo domains on its `verizonwireless.com` and `verizon.com` domains. For Neiman Marcus, we still see 65 typo domains, and for Microsoft 437. It seems the current approach of individual trademark holders pursuing individual squatters has not been effective in preventing or discouraging typosquatting by others. Therefore, we next consider methods to influence companies that distinctively benefit from typosquatting: domain aggregators and advertising platforms.

6.2 Identifying Servers that Distinctively Host Typo Domains

Large domainers typically host their domains on a single set of name servers. By comparing the incidence of typo domains across name servers, we assess which name servers host disproportionately many typo domains.

Our analysis found 937 918 typo domains out of 80 988 864 `.com` domains; consequently, any name server with over 1.16% typo domains is above average. Table 4 (left) shows the incidence of typosquatting at large name servers. Many large parking companies identified in Section 4.1 have disproportionately many typo domains: 2.5% of domains resolved by `sedoparking.com` are typos, over twice the rate on the web as a whole. On smaller name servers, typo domains can be even more frequent. Table 4 (right) considers name servers hosting at least 1 000 names. Topping the list is `moniker.com`, with 62% typo domains.

At the same time, other name servers feature disproportionately infrequent typo domains. For example, the bottom of Table 4 (left) shows large name servers with typo domains as infrequent as 0.12%, one tenth the Internet-wide average.

6.3 The Role and Responsibility of Ad Platforms

We pause for an important disclosure: One of the authors (Edelman) is co-counsel in litigation seeking to hold Google liable for using typosquatting domains to display advertising [11]. However, we now write not as lawyers but as engineer and economist seeking to address typosquatting in the most efficient way possible.

As shown in Section 3.1 and Table 2, of the typo domains we successfully crawled, nearly 80% showed pay-per-click advertisements that came from the ad platforms operated by the web’s top search engines, principally Google and (to a significantly lesser extent) Yahoo. Because ad platforms are the primary or sole source of revenue for these typo domains, we believe ad platforms are well-positioned to substantially reduce typosquatting. Among other responses, ad platforms could select partners more carefully, select only partners with a demonstrated record of avoiding typosquatting, and/or sever ties to partners who are found to engage in typosquatting. Furthermore, ad platforms could require that new partners showing ads on many domains post a bond that is forfeited upon typosquatting, or deduct penalties from payments to any partners found to engage in typosquatting. To the best of our knowledge, ad platforms have taken none of these steps.

Ad platforms typically claim that a website or trademark owner targeted by typosquatting should address its complaint directly to the typosquatter, not to the ad platform that pays the typosquatter. For example, Google’s AdSense for Domains complaint page argues that “Google is not in any way involved with the selection or registration of these domain names, and is not in a position to arbitrate trademark disputes between the registrants, our partners, and trademark owners. Accordingly, we encourage trademark owners to resolve their disputes directly with the registrants or registrars.”³ By stepping out of disputes between sites and typosquatters, ad platforms’ preferred approach simplifies disputes (to entail two parties rather than three) and, of course, limits ad platforms’ potential liability.

Despite the simplification resulting from ad platforms’ preferred approach, we see multiple problems with ad platforms disclaiming all responsibility for the typosquatting they fund. For one, our analysis confirms that payments from ad platforms are the sole force behind most typosquatting registrations. Furthermore, ad platforms are least-cost avoiders – able to prevent typosquatting with less effort than any other party. In particular, thanks to the semantic analysis capabilities and spelling correction skills search engines gained through their principal businesses, ad platforms are well equipped to identify typosquatting registrations. (Consider Google’s well-known and strikingly accurate “Did you mean?” function.) Indeed, search engines already receive information about the specific domains users visit (necessary to target ads accordingly). It would be straightforward to compare these requests to a list of top trademarks, and disallow parking ads from appearing on sites that are misspellings of popular domains.

The dynamics of the typosquatting business give ad platforms a particularly powerful opportunity to undermine typosquatting. Suppose a site owner pursued

³ <http://adwords.google.com/support/aw/bin/answer.py?answer=50003&topic=26>

a few large typosquatters. The associated typo domains would tend to scatter to numerous smaller typosquatters who could not be identified, located, or pursued cost-effectively (as has already happened to Microsoft, Verizon and others). In contrast, ad platforms enjoy unique positions of authority, buttressed by their relationships with advertisers. Consequently, ad platforms can authoritatively undermine typosquatting, in a way that no individual site owner can.

7 Conclusions

We are struck by the scale of the problem of typosquatting – at least many hundreds of thousands of typo domains, and probably millions – despite substantial public and private efforts to discourage such registrations. Yet with such strong economics supporting typosquatting – payments from Google and others – perhaps it is no surprise that typosquatting is as prevalent as ever.

We suspect typosquatting will continue so long as advertisers and ad networks continue to fuel and fund these practices. But let no one suggest identifying typo domains is impossible: The overwhelming majority of typos are easy to recognize, by hand or using straightforward automation. At the same time, with typo domains highly concentrated at a few large domainers and ad platforms, intermediaries could significantly discourage the registration and use of typo domains if they were so inclined.

References

1. Brand Owners Could Have Prevented \$220 Million In Domain Name Recovery By Spending \$1 Million. Corporation Service Company. August 24, 2009.
2. F.J. Damerau: A Technique for Computer Detection and Correction of Spelling Errors. *Communications of the ACM* **7**(3) 171-176, 1964.
3. B. Edelman: “Domains Reregistered for Distribution of Unrelated Content: A Case Study of “Tina’s Free Live Webcam,” 2002. <http://cyber.law.harvard.edu/people/edelman/renewals/>.
4. B. Edelman: “How Google and Its Partners Inflate Measured Conversion Rates and Increase Advertisers’ Costs,” 2009. <http://www.benedelman.org/news/051309-1.html>.
5. B. Edelman: “Large-Scale Registration of Domains with Typographical Errors,” 2003. <http://cyber.law.harvard.edu/people/edelman/typo-domains/>.
6. V.I. Levenshtein: Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics Doklady*, 1966.
7. *Lands’ End, Inc. v. Eric Remy, et al.*, W.D.Wis. 2006.
8. *MTV Networks v. Curry*, 867 F.Supp. 202. SDNY 1994.
9. *The Neiman Marcus Group Inc, et al., v. Dotster Inc, et al.*, W.D.Wa. 2006.
10. M. Perez: Verizon wins \$33 Million In Cybersquatting Case. *Information Week*, December 30, 2008.
11. *Vulcan Golf, LLC, et al., v. Google, Inc., et al.*, N.D.Ill. Case No 1:2007cv03371.
12. Y. Wang, D. Beck, J. Wang, C. Verbowski, and B. Daniels: “Strider Typo-Patrol: Discovery and Analysis of Systematic Typo-Squatting,” 2nd USENIX Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI), July 2006.