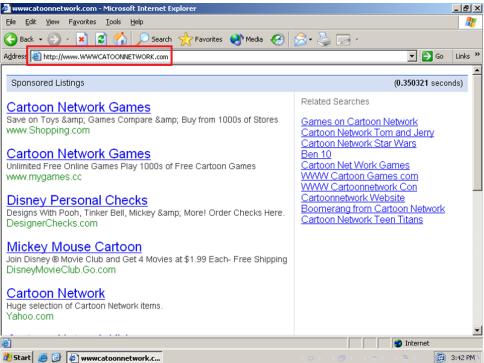
## Measuring the Perpetrators and Funders of Typosquatting

Tyler Moore and Benjamin Edelman

Center for Research on Computation and Society, Harvard University
Harvard Business School

Financial Crypto, Tenerife, January 27, 2010 http://www.benedelman.org/typosquatting/





#### Outline

- Measuring Typosquatting
  - Motivation and Industry Structure
  - Identifying Typosquatting Domains
  - Crawling Typosquatting Websites
- 2 How Typosquatting Domains are Used
  - Squatter Strategy 1: Pay-Per-Click Ads
  - Squatter Strategy 2: Redirection and Linked Domains
  - Do Pay-Per-Click Ads Promote Typosquatting?
- 3 Countering Typosquatting
  - Existing Efforts to Regulate Typosquatting
  - Identifying Servers that Distinctively Host Typo Domains
  - The Role and Responsibility of Ad Platforms



#### Outline

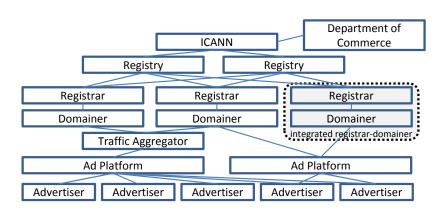
- Measuring Typosquatting
  - Motivation and Industry Structure
  - Identifying Typosquatting Domains
  - Crawling Typosquatting Websites
- 2 How Typosquatting Domains are Used
  - Squatter Strategy 1: Pay-Per-Click Ads
  - Squatter Strategy 2: Redirection and Linked Domains
  - Do Pay-Per-Click Ads Promote Typosquatting?
- Countering Typosquatting
  - Existing Efforts to Regulate Typosquatting
  - Identifying Servers that Distinctively Host Typo Domains
  - The Role and Responsibility of Ad Platforms



## Why care about typosquatting?

- Typosquatting confuses consumers, harms user experience
- Illegal under Anti-Cybersquatting Consumer Protection Act (1998)
  - Forbidden to register, traffic in or use domains confusingly similar to trademarks
- Undermines efficiency of online advertising market
  - Companies pay for traffic already intended for destination
  - Forces defensive advertising
- Squatters employ high-volume, low-success, automated approach similar to spam, phishing and other online crime

#### Structure and strategy of the domaining business



#### How to find typosquatting domains

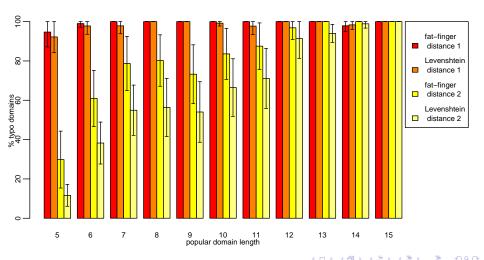
- Gather a list of popular domains
  - 3264 .com domains  $\geq$  5 characters long in Alexa top 6000
- Compute Damerau-Levenshtein distance between popular domains and all registered .com's
  - Minimum # of insertions, deletions, substitutions or transpositions needed to transform one string into another
  - faceboolk, facebok, faceboik, and faceboko each have a Damerau-Levenshtein distance of 1 from facebook
- Also compute fat-finger distance
  - Minimum # of insertions, deletions, substitutions or transpositions using letters adjacent on a QWERTY keyboard to transform one string into another

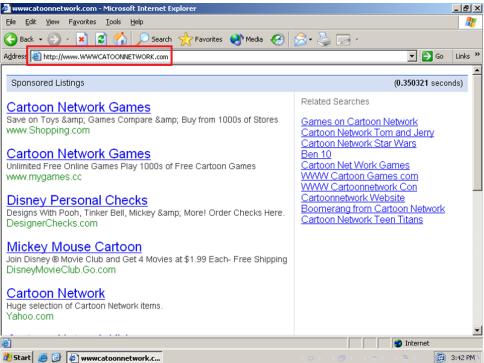


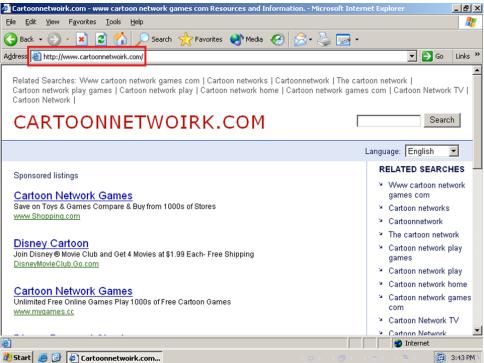
#### How to find typosquatting domains

- Enumerate all strings with a Damerau-Levenshtein (& fat-finger) distance  $\leq 2$  from each of the 3 264 popular domains
- Intersect this set with 81 million active .com domains
  - 1910738 candidate typo domains match 3 264 popular domains
  - Manually sampled 2 195 randomly selected candidate typos
    - 50 domains each for Lev. and fat-finger distance 1&2, for popular domains of length 5–15

## Typosquatting classification accuracy







#### 1278 cartoonnetwork.com typos, including...

cartoonnetsork com cartoonntewrk.com cargoonnetwork.com cartoonnetwkor.com cartoonetgork.com cartoomnetwoork com caretoonetwork.com cartoonnetwokr com cartoonnetworkcom.com carfoonnetwork com cartoopnetwork.com cagrtoonetwork.com cartoonntwoork.com cartoonetwoirk.com cartoonnnetwrk com cartoonhetwork.com cartoonnetook com catoonnettwork.com cartoonnetwork.com cartoconnetwrk com cartoonnexwork.com cartoonetworg.com crattonnetwork.com cartoonnerwort com cartoonnetwar com cattoonnetwok.com cartoolnnetwork com cartoonnedword.com carttoonnetwook com cartomnetwork.com

cartoinnetwork com cartoonnetlork.com carttoonnnetwork com catoonnnetwork.com cartoonnetgork.com carvoonetwork.com cartoonetwoork.com cartoonnetwokl com nartoonnetwork.com cartoonnotwork com cartoonnetwogk.com cartoonneework.com catoonneetwork com cartoonznetwork.com cartoommetwork com fcartoonnetwork.com cartoonnetkwork com cartooOnnetwork.com cartoonetwrk.com cartoonnethwork com caroconnetwork.com cartoonetworl.com cartoonnewtokr.com cartoonnerwork com cartoonnetwak com cartoonnetwwork.com cartoonetworkcom com cartoonnedwork.com cartoonnetwowrk com

cartolnnetwork com cartoonnetowok.com cartoonnwetwork com cartooconnetwork.com cartoonneetwort com cartoconetwork com cartoonnetwoer.com cartoonnetwoke com cartoonnstwork.com cartoonnnetwok com cartoonetwaork.com cartppnnetwork.com crattoonnetwork com cartoobnetwork.com cartoonnetwart com catoonnetwerk.com cartonnetwokr com cacrtoonnetwork.com cartoonnewark.com caetoconetwork com dartoonnetwork.com cartoonetwori.com carntoonnetwork.com cartoonnerworl com cartoonnekwork com cartoonnetgor.com casrtoonetwork com www.carttoonnetwork.com cartoonetwgork.com cartoonnetorg.com

cartoonnftwork com crtonnetwork.com cartoonetrwork com caconnetwork.com cartoonneetwork com cantoonnetwork com carttoonnetwerk.com cartoownnetwork com cartoounnetwork.com cartoonnnetwor com cartoonntewrok.com cartoonnetmwork.com cartoonnetweark com catoonnework.com www.cartonetwork.com artoonnetwor.com carltonnetwork com cartoonnetwoorkl.com cartoonnetwoirk.com cartoonknetwork com certoonnetwork.com cartoonetwork com caretoonnetwork.com cartoonnetfork com cartooknetwork com cartoonnetwowk.com cartoonnetswork com cartoonerwork.com crartoonnetwork com

cartoonnevwork.com cartoonnegwork.com cartoonnetwodrk com cartonbetwork.com catoonnetwrok.com cargoonnetworm.com chartoonnetwork.com cartoohetwork com cartoonework.com cartonnetwortk com cartoonetwoek.com cartoconework.com carttooonetwork com cartiinnetwork.com cartoonnttwork com cartoonnetwock.com cartoonetowrk com cartoonedtwork.com cartoknnetwork.com catcommetwork com cartoonetword.com cartoonetwort com cartoconnetwoork.com cartoonnetttwork com cartoonegwork.com www.cat.oonetwork.com cartoonnedwort com cattoonnetwark.com czrtoonnetwork com crt.connework.com

cartoonnetwrak.com

cratonnetwork com

#### Selected domains highly targeted by typosquatting

total for 3 264 domains	937 918	(710 872, 1 236 924)
canada.com	1 587	(854, 2158)
sprint.com	1 675	(1006, 2231)
games.com	1743	(1530, 2531)
hotels.com	1865	(1207, 2442)
freecreditreport.com	1 904	(1904, 1904)
myspace.com	1 960	(1457, 2440)
youtube.com	2 0 6 9	(1589, 2534)
google.com	2 5 3 7	(1728, 3252)
popular domain	typo domains	95% confidence interval

#### How the crawler works

- Built a crawler to determine how typosquatting domains are used
  - 937 918 sites too many to crawl, and contains too many false positives
  - So instead visit 284 914 domains where the Damerau-Levenshtein distance is 1 for popular domains 5–9 characters long, or the Damerau-Levenshtein distance  $\leq$  2 for popular domains  $\geq$  10 characters
- For each domain, pull down index page and headers, and visit
   3 randomly selected links up to depth 3
- Stop when a link or redirect to known advertising domain is encountered

#### What the crawler found

- Successfully classified 93 251 domains
  - 74 024 have pay-per-click ads
  - 19 227 use redirection or links to popular domains or competitors
- Blocked by 124 211 domains
  - CAPTCHAs, connection refused, etc.
  - Main reason for blocking: many domains are hosted on a few servers (up to 30K typo domains hosted on a single IP address)
- Unclassified 70 729 domains
  - Incorrect or incomplete javascript parsing, etc.

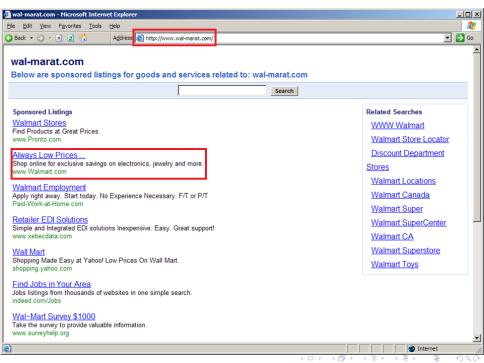
#### Outline

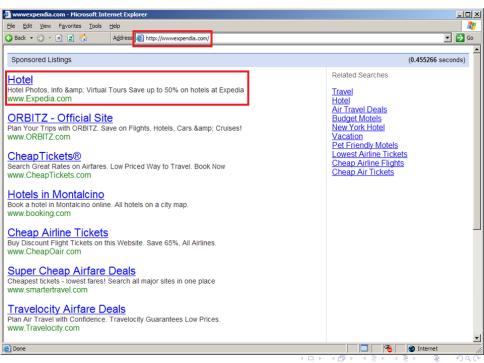
- Measuring Typosquatting
  - Motivation and Industry Structure
  - Identifying Typosquatting Domains
  - Crawling Typosquatting Websites
- 2 How Typosquatting Domains are Used
  - Squatter Strategy 1: Pay-Per-Click Ads
  - Squatter Strategy 2: Redirection and Linked Domains
  - Do Pay-Per-Click Ads Promote Typosquatting?
- Countering Typosquatting
  - Existing Efforts to Regulate Typosquatting
  - Identifying Servers that Distinctively Host Typo Domains
  - The Role and Responsibility of Ad Platforms



#### Pay-per-click ad breakdown

ad platform	typo domains	%
Google	53 364	59.3
Yahoo!/Overture	19 145	21.3
Ask.com	555	0.6
Miva	541	0.6
Enhance	297	0.3





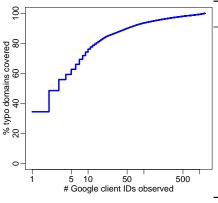
# Self-advertisement prevalence for 263 popular domains buying ads

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

## Inferring advertising partners from advertising links

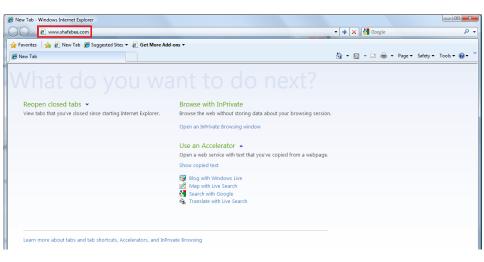
- Can often infer the advertising partner from a parameter included in advertising link
  - Yahoo ads obfuscate URL parameters, but Google does not
- http://domains.googlesyndication.com/ apps/domainpark/results.cgi?client=ca-dp-mborin&...

#### Few advertising client IDs match most typo domains

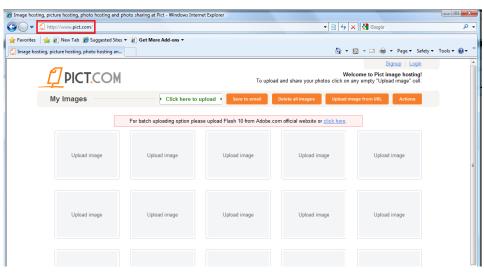


Google client ID	values	domains
ca-dp-highlands*	7	14724
ca-dp-godaddy*	37	7 949
ca-dp-sedo*	14	4 583
ca-dp-sphere*	7	3809
ca-dp-dopa*	5	1 402
ca-dp-namedrive*	12	489
ca-afdo-pub*	447	1 299
others	721	4 969
total	1 250	39 238

## A user mistypes sharebee.com (Alexa rank 1673)



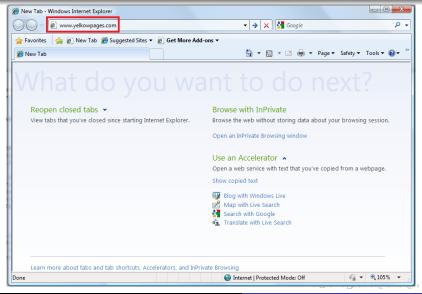
#### ...and is redirected to pict.com (Alexa rank 8581)



## 128 typo domains redirecting to pict.com

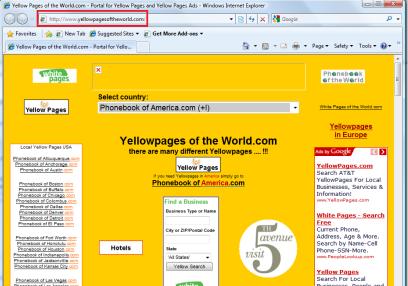
depositfiles.com typos:	deposktfiles.com	depositfimes.com	depositfipes.com	deposutfiles.com	depoeitfiles.com
	depositciles.com	depositfkles.com	deposigfiles.com	deplsitfiles.com	depositfilee.com
	depowitfiles.com	depositciles.com	deposktfiles.com	depositfimes.com	depoeitfiles.com
	depositfkles.com	depositfipes.com	deposigfiles.com	deposutfiles.com	deplsitfiles.com
	depositfilee.com	depositfilew.com	cepositfiles.com	depowitfiles.com	
picoodle.com typos:	picoodke.com	picoodpe.com	picoodme.com	piciodle.com	picpodle.com
	pkcoodle.com	pocoodle.com	picoosle.com	picoofle.com	pucoodle.com
	picooele.com	pkcoodle.com	picoosle.com	picoodke.com	pocoodle.com
	picoodme.com	piciodle.com	picoodpe.com	picpodle.com	picoofle.com
	pucoodle.com	picooele.com			
sharebee.com typos:	shaerbee.com	shafebee.com	shatebee.com	sharevee.com	sharegee.com
	shraebee.com	sharenee.com	sahrebee.com	shsrebee.com	shaerbee.com
	sharegee.com	shraebee.com	sharenee.com	shafebee.com	sahrebee.com
	shatebee.com	sharevee.com	shsrebee.com		
imagebam.com typos:	imahebam.com	inagebam.com	imagebsm.com	imagegam.com	ikagebam.com
	ijagebam.com	imagevam.com	imagenam.com	imahebam.com	ikagebam.com
	imagebsm.com	inagebam.com	ijagebam.com	imagevam.com	imagegam.com
	imagenam.com				
imagevenue.com typos:	imarevenue.com	kmagevenue.com	ijagevenue.com	imqgevenue.com	imagevehue.com
	imanevenue.com	ikagevenue.com	imarevenue.com	kmagevenue.com	ijagevenue.com
	imagevehue.com	imanevenue.com	imqgevenue.com	ikagevenue.com	
savefile.com typos:	savefkle.com	saverile.com	saevfile.com	savefike.com	savefole.com
	savefkle.com	saevfile.com	savedile.com	savefike.com	savefole.com
	saverile.com				
sendspace.com typos:	sehdspace.com	sendspqce.com	eendspace.com	sendslace.com	sehdspace.com
	sendspqce.com	eendspace.com	sendslace.com	sendwpace.com	
esnips.com typos:	esnipw.com	ewnips.com	esnils.com	ewnips.com	esnils.com
•	esnipw.com	•		•	
webshots.com typos:	webshlts.com	webehots.com	webshlts.com	webehots.com	
4shared.com typos:	4shafed.com	4shaged.com	4shaged.com	4shafed.com	■ト 4 ■ト ■ の
Tyler Moore: http://u		Ü	<u> </u>		oders of Typosquatting

## A user mistypes yellowpages.com (Alexa rank 762)



Squatter Strategy 1: Pay-Per-Click Ads Squatter Strategy 2: Redirection and Linked Domains Do Pay-Per-Click Ads Promote Typosquatting?

#### ... redirected to yellowpagesoftheworld.com (rank 884 116)



#### 156 domains redirecting to yellowpagesoftheworld.com

```
yellopasges.com
                     yeollwpages.com
                                          ydllowpages.com
                                                            yellowpanes.com
                                                                                yelkowpages.com
yellowpagyes.com
                     yellowpaegs.com
                                          yerllowpages.com
                                                            yepllowpages.com
                                                                                yellozpages.com
vellowpoages.com
                     yellowpes.com
                                          yellowwages.com
                                                            yelloqpages.com
                                                                                yellowpasges.com
wwwyellowpags.com
                     wwwellowpages.com
                                          yellowpazges.com
                                                            yellowpagfes.com
                                                                                yellowpagex.com
                                          yelloawpages.com
yellowpagee.com
                     yellowpagea.com
                                                            yellowespages.com
                                                                                yellowpagezs.com
jellowpages.com
                     wwwyallowpages.com
                                          ytellowpages.com
                                                            velliwpages.com
                                                                                yellowpajes.com
yellowpaghes.com
                     yellowpaves.com
                                          yewllowpages.com
                                                            yeppowpages.com
                                                                                yrellowpages.com
vellokwpages.com
                     vellowpafges.com
                                          yellow-ages.com
                                                            vellowapges.com
                                                                                yellowpagges.com
vfellowpages.com
                     vefllowpages.com
                                          vellowoages.com
                                                            yellpepages.com
                                                                                yyellowpage.com
                     vllopages.com
yellowpagescom.com
                                          yillowpages.com
                                                            yellowopages.com
                                                                                yellowpahges.com
yellowpagse.com
                     yelloapges.com
                                          vellowapages.com
                                                            vesllowpages.com
                                                                                yellowpagwes.com
                                          yuellowpages.com
yyellowpages.com
                     yellowspages.com
                                                            yellpwpages.com
                                                                                yellowage.com
wwwyellow-pages.com
                     yellowpgae.com
                                          ysellowpages.com
                                                            yellowpagexs.com
                                                                                yellowpawges.com
yelloapages.com
                     yellopasges.com
                                          veollwpages.com
                                                            yellpowpages.com
                                                                                vellowpagesss.com
yellowpayes.com
                     yellowpagw.com
                                          yellowpgaes.com
                                                            yellowpagfs.com
                                                                                www.yellowpagescom.com
yellowpwges.com
                     yellllowpages.com
                                          yellopwages.com
                                                            yellowpafes.com
                                                                                yellopwage.com
yellowpatges.com
                     ydellowpages.com
                                          yellowpagtes.com
                                                            yrllowpages.com
                                                                                yullowpages.com
yellowpag.com
                     yellowpae.com
                                          yellowpas.com
                                                            vellopwpages.com
                                                                                yellowlages.com
vwellowpages.com
                     yelleowpages.com
                                          ysllowpages.com
                                                            yellowwpages.com
                                                                                velloages.com
yellowpasge.com
                     yelloqwpages.com
                                          eyllowpage.com
                                                            yelwpages.com
                                                                                yellowpaeges.com
vellowpagews.com
                     ydllowpages.com
                                          vellowpagbes.com
                                                            vllowpage.com
                                                                                iellowpages.com
ywllowpages.com
                     hellowpages.com
                                          yelloppages.com
                                                            yellowpagres.com
                                                                                wwwyllowpages.com
pellowpages.com
                     yellowpages.com
                                          vellowpagesd.com
                                                            yellowpwages.com
                                                                                yellowpaeg.com
yellowpaes.com
                     yellowppages.com
                                          vlowpages.com
                                                            vellowpeages.com
                                                                                yhellowpages.com
yellowpagds.com
                     yellowages.com
                                          yellowmpages.com
                                                            yellowpagws.com
                                                                                yellowpagwa.com
yellowpge.com
                     yellowpageds.com
                                          yellowpagss.com
                                                            yellowpahes.com
                                                                                yellowpsges.com
yeklowpages.com
                     yellowpaiges.com
                                          yello3pages.com
                                                            yellowpanges.com
                                                                                yellkowpages.com
                                                            yeyllowpages.com
veellowpages.com
                     yekllowpages.com
                                          yellowpagge.com
                                                                                yeowpages.com
vellowpagdes.com
                     yellowpabges.com
                                          uellowpages.com
                                                            velelowpages.com
                                                                                vlleowpages.com
yellowpagves.com
                     yellowpavges.com
                                          yellowpagrs.com
                                                            yellowges.com
                                                                                llowpages.com
vellwpage.com
                     vellowpsages.com
                                          vwllowpage.com
                                                            velloewpages.com
                                                                                vekkowpages.com
```

#### Automatically identifying typos that link to competitors

- Explore the beneficiary domains linked to by typo domains
- Websites such as pict.com are picky when it comes to registering typos
- So the heuristic has two conditions
  - Beneficiary domains must be linked by at least 75 typo domains
  - 2 Typo domains must impersonate at most 40 popular domains

#### Other example domains linked to by popular domains

```
bet365.com: 367 typo domains
sportsbook.com: saportsbook, sxportsbook, sportszbook & 325 more
betclic.com: betclico, betclicm, betclicj & 7 more
```

smashinggames.com: 376 typo domains

```
dailyfreegames.com: dailyfreegsmes, dwilyfreegames, dailygreegames & 35 more freeworldgroup.com: freeworldhroup, freeworldgroup, freeeorldgroup & 33 more modthesims2.com: modthrsims2, mpdthesims2, modrhesims2 & 27 more
```

fulltiltpoker.com: fulltilt6poker, fuylltiltpoker, fulltiltpoke4r & 5 more

movietheatertickets.biz: 85 typo domains

```
movietickets.com: movietikits, mpvietickets, muvietickets & 19 more rottentomatoes.com: rottentomaos, rottentomatoes, rotentomatoe & 10 more fandango.com: fandango, fandango, faneango & 9 more
```

## total: 75 beneficiary domains on 4879 typo domains impersonating 668 competing popular domains

#### Other redirections and links from typo domains

- Companies also defensively register mispellings
  - Found 4133 typo domains redirect to corresponding popular domain and use same name server as popular domain
- Affiliate marketing
  - Another way to monetize traffic intended for popular domain
  - Squatter receives commission from popular domain for providing a 'lead'
  - Suspect affiliate marketing is used in 10185 domains

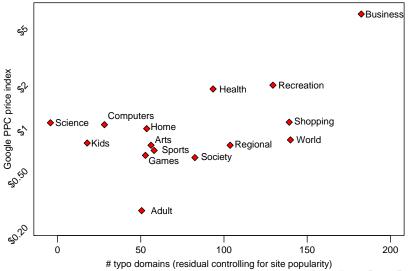
#### Why are some domains typosquatted more than others?

- Hypothesized that hard-to-spell sites are disproportionately targeted
  - Found no effect for various proxies of difficulty, including # of double letters, i/e or e/i presence
- Did find significant difference across website categories, as assigned by Alexa
  - Might pay-per-click ad prices help explain the difference?

## Google PPC price index

- Extracted META keywords for every popular domain in each category
- Fetched the minimum and maximum PPC price estimates for each category's top 10 keywords using Google Traffic Estimator
- Define Google PPC price index as average of
  - median of minimum PPC prices for keywords in the category
  - @ median of maximum PPC prices for keywords in the category

#### Websites with higher PPC payouts face more typosquatting



#### Outline

- Measuring Typosquatting
  - Motivation and Industry Structure
  - Identifying Typosquatting Domains
  - Crawling Typosquatting Websites
- 2 How Typosquatting Domains are Used
  - Squatter Strategy 1: Pay-Per-Click Ads
  - Squatter Strategy 2: Redirection and Linked Domains
  - Do Pay-Per-Click Ads Promote Typosquatting?
- 3 Countering Typosquatting
  - Existing Efforts to Regulate Typosquatting
  - Identifying Servers that Distinctively Host Typo Domains
  - The Role and Responsibility of Ad Platforms



#### Existing efforts to regulate typosquatting

- Uniform Domain-Name Dispute-Resolution Policy (UDRP)
  - Reduced cost and sped up removal of infringing domains
  - Invoked 45 000 times between 1999-2009
- ACPA allows trademark holders to sue in rem
  - Can sue the domain itself, even if the registrant is nowhere to be found
  - Puts registrars and registries on the hook instead
- ACPA also allows for statutory damages (\$1K-\$100K per domain)
  - Intent was to deter would-be typosquatter with threat of fines

#### Challenges to existing regulatory efforts

- A few lawsuits have been successful
  - Verizon (\$33 million judgment), Microsoft, Neiman Marcus
  - Yet even these marks are still typosquatted: Verizon (767),
     Microsoft (437), Neiman Marcus (65)
- 45 000 UDRP requests covers less than 5% of the typosquatting domains identified
- Conclusion: individual trademark holders, pursuing individual squatters, have not been effective in stopping typosquatting
- Should instead influence those who distinctly benefit from typosquatting: domain aggregators and ad platforms

#### Name servers with the most typosquatted domains

Name servers > 100 000 domains		Name servers > 1 000 domains			
name server	% typo	typos	name server	% typo	typos
dnsnameserver.org	4.75	19 217	moniker.com	61.65	910
trellian.com	4.47	11 962	ipmanagerinc.net	55.63	787
hitfarm.com	3.76	17 073	citizenhawk.net	31.88	1766
dsredirection.com	3.60	59845	dexner.com	18.85	375
linkz.com	2.98	3765	aphost.com	17.96	4 244
fastpark.net	2.77	7715	freeredirection.net	17.94	1 438
above.com	2.77	16691	ehostinginc.com	17.89	181
sedoparking.com	2.51	35 216	nnw.net	17.10	250
parked.com	2.48	13 993	onlinednsservice.net	15.09	2844
trafficz.com	2.38	5 337	topdns.com	15.01	889
bottom 5	÷	:	plus 96 name servers above 5% typo domains		domains
ipowerweb.net	0.32	569			
ipowerdns.com	0.30	522			
123-reg.co.uk	0.26	860			
abac.com	0.14	248			
vpweb.com	0.12	127			

#### Role and responsibility of ad platforms

- 80% of crawled typo domains show pay-per-click ads from ad platforms operated by top search engines
- Ad platforms are well-placed to stop typosquatting
  - Select partners more carefully
  - Sever ties to partners found to engage in typosquatting
  - Deduct penalties from payments to partners found to be typosquatting
  - . . .
- Ad platforms can prevent typosquatting with less effort than other parties
  - Already developed good typo detection for core business ('Did you mean?')
  - Feasible to stop syndicating ads on mispellings of top trademarks, thereby eliminating the squatters' main revenue source

#### Conclusion

- Typosquatting remains widespread nearly 1 million domains impersonating top 3264 .com sites – despite more than a decade of public and private efforts
- The financial payout from advertisers and ad networks outweighs any individual efforts from trademark holders
- High concentration at a few large domainers and ad platforms means intermediaries could reduce typosquatting if so inclined
- See online appendix for more examples
   http://www.benedelman.org/typosquatting/
- For more . . .
  - Web: http://people.seas.harvard.edu/~tmoore/
  - Email: tmoore@seas.harvard.edu

