

# Fashion Crimes: Trending-Term Exploitation on the Web

Tyler Moore<sup>1</sup>, Nektarios Leontiadis<sup>2</sup>, Nicolas Christin<sup>2</sup>

Computer Science Department, Wellesley College<sup>1</sup>  
CyLab, Carnegie Mellon University<sup>2</sup>

ACM Conference on Computer & Communications Security  
Chicago, Illinois  
October 18, 2011

# Outline

- 1 How trending search terms are abused
  - Monetizing traffic: malware or ads?
  - Research objectives
- 2 Measuring trending-term abuse
  - Data collection methodology
  - Incidence of abuse
  - How search-term characteristics affect abuse prevalence
- 3 Economics of trending-term exploitation
  - Estimating the exposed population
  - Revenue analysis: ad abuse
  - Revenue analysis: malware
- 4 What happens when search engines intervene?
  - Measuring the effect of Google's intervention
  - Cautionary tale on crackdowns

# Outline

- 1 How trending search terms are abused
  - Monetizing traffic: malware or ads?
  - Research objectives
- 2 Measuring trending-term abuse
  - Data collection methodology
  - Incidence of abuse
  - How search-term characteristics affect abuse prevalence
- 3 Economics of trending-term exploitation
  - Estimating the exposed population
  - Revenue analysis: ad abuse
  - Revenue analysis: malware
- 4 What happens when search engines intervene?
  - Measuring the effect of Google's intervention
  - Cautionary tale on crackdowns

# Search terms can be highly dynamic

Google trends

Search Trends

Tip: Use commas to compare multiple search terms.

## Hot Searches (USA)

Aug 10, 2011 - [change date](#)

Updated 40 minutes ago

iGoogle <sup>New!</sup>  
Gadget

 Site  
Feed

- |                                     |  |  |   |
|-------------------------------------|--|--|---|
| 1. <a href="#">pga_championship</a> | 6. <a href="#">raksha_bandhan</a>          | 11. <a href="#">jenelle_evans</a>            | 16. <a href="#">beach_volleyball</a>    |
| 2. <a href="#">kendall_jenner</a>   | 7. <a href="#">samsung_galaxy_tab_10_1</a> | 12. <a href="#">sdn</a>                      | 17. <a href="#">wisconsin_news</a>      |
| 3. <a href="#">anon</a>             | 8. <a href="#">jennifer_hudson</a>         | 13. <a href="#">search</a>                   | 18. <a href="#">krav_maga</a>           |
| 4. <a href="#">snow_leopard</a>     | 9. <a href="#">jennifer_grey</a>           | 14. <a href="#">crush</a>                    | 19. <a href="#">utc</a>                 |
| 5. <a href="#">guy_fawkes</a>       | 10. <a href="#">brownie</a>                | 15. <a href="#">sons_of_anarchy_season_4</a> | 20. <a href="#">frances_bean_cobain</a> |

Google Trends provides insights into broad search patterns. Please keep in mind that several approximations are used when computing these results.

©2008 Google - [Discuss](#) - [Terms of Use](#) - [Privacy Policy](#) - [Help](#)

However, not all of the search results are relevant!

How trending search terms are abused  
Measuring trending-term abuse  
Economics of trending-term exploitation  
What happens when search engines intervene?

Monetizing traffic: malware or ads?  
Research objectives

# At best you may encounter ad-filled sites

The screenshot shows a web browser window displaying the eWorldPost website. The page features a navigation menu with categories like HOME, SCI & TECH, WORLD, ENTERTAINMENT, etc. The main content area includes several advertisements and search results. A red circle highlights the text "After a long wait from" in the search results section. Another red circle highlights the text "fans, Christopher Nolan" in the footer area. A third red circle highlights the text "Ads by Google" in the footer area. A large advertisement on the right side of the page reads "Newark to Oslo Daily Nonstop".

Stay Connected / Friday, December 17, 2010

# eWorldPost

HOME SCI & TECH WORLD ENTERTAINMENT DOMAINING WEBSITE REVIEWS WEBMASTERING  
BUSINESS SPORTS POLITICS HEALTH TRAVEL GENERAL SOFTWARE RELIGION ARTS

ABOUT US RECOMMEND WEBSITE REVIEW SUBMIT A PRESS RELEASE

Dark Knight Rises, the next fiesta for Batman fans, by Nolan!

Share

**GroupVIL COUPONS** [www.GroupVil.com/Boston](http://www.GroupVil.com/Boston)  
1 ridiculously huge coupon a day. It's like doing Boston at 90% off!

**Take an IMAX Thrill Ride** [www.mos.org](http://www.mos.org)  
Ride Wild Rollercoasters - Only At The Museum of Science. Buy Tix Now!

**Batman & Robin** [xfinitytv.fancast.com](http://xfinitytv.fancast.com)  
The Adventures of Batman & Robin. Catch Up On Your Favorite Shows.

**Free Batman Games Online**  
Find more sources/options for Free Batman Games Online  
[www.webcrawler.com](http://www.webcrawler.com)

**How to Draw Batman**  
Access Step-by-Step Instructions Learn How to Draw Batman  
[howtotutorials.net](http://howtotutorials.net)

**Robbie the Reindeer Fan?**  
See The Illusionist - Nominated For 5 Annie Awards Incl. Best Picture!  
[SonyClassics.com/TheIllusionist](http://SonyClassics.com/TheIllusionist)

**Batman Costumes & Access**   
Also 15,000+ other quality costumes  
Guaranteed low prices.  
[Costumes4Less.com](http://Costumes4Less.com)

Ads by Google

**LATEST NEWS**

Ac.Milan Vs Sampdoria 1-1 Serie A Highlights- Sampdoria Manage To Halt The Leaders  
The leader of the Serie A Italian football, AC Milan drew 1-1 with Sampdoria in Stadio Comunale Luigi Ferraris over in...

Check out our great deals

After a long wait from

fans, Christopher Nolan

Ads by Google

# At worst you may encounter malware

The screenshot shows a web browser window with a Resident Shield alert dialog box open. The alert is titled "Multiple threat detection" and contains a table with the following data:

File	Infection	Result
C:\Documents and Settings\Administrator\Lo...	Virus identified JS/Poyme 0W	Infected
C:\Documents and Settings\Administrator\Lo...	Virus identified JS/Poyme 0W	Infected

Below the table, there are options to "Remove threat as Power User" (unchecked) and buttons for "Remove selected infections", "Remove all unheated infections", and "Close". At the bottom of the dialog, it shows "Process Name: C:\PROGRAM~1\INTERN~1\explore.exe", "Process ID: 11264", and "Detected on open." with a link to "More information about this threat..."

The background shows the Cleveland.com website with a "SUN" logo and a "Bryant & Stratton College" sponsored by section. There are also advertisements for "SME CLEVELAND" and "SME - U Morning".

## Research goals

- 1 Measure the prevalence of abuse in trending terms' search results relative to other terms
- 2 Identify whether certain types of search terms are more susceptible to abuse and why
- 3 Construct an economic model of revenue from malware and ads to understand the behavior of profit-minded adversaries
- 4 Measure the impact of a search-engine crackdown on low-quality, "made for Adsense" (MFA) sites

# Why worry about dodgy advertising?

- Legal crackdowns on the underground economy might tempt criminals to shift to more reliable income sources
- Online advertising is a logical target
  - Ad platforms **lack the incentive** to detect fraud, since detection directly reduces profit
  - Advertisers struggle to monitor for abuse due to **lack of transparency**
  - Criminals already profit from online advertising: botnets carry out click-fraud, spyware games affiliate-marketing programs



## Related work

- Empirical investigations of the underground economy
  - Underground fora (Franklin et al. CCS 2006, Caballero et al. USENIX Security 2011)
  - Email spam (Kanich et al. CCS 2008, Levchenko et al. S&P 2011)
  - Phishing (Moore and Clayton eCrime 2007)
  - Online social networks (Grier et al. CCS 2010)
- Empirical investigations of web-based scams
  - Social engineering (Christin et al. CCS 2010)
  - Drive-by downloads (Provos et al. USENIX Security 2007)
  - Web spam to promote fake antivirus (Rajab et al. LEET 2010, Cova et al. RAID 2010, Stone-Gross et al. WEIS 2011)
  - Web spam to promote ads (Wang et al. WWW 2007, Moore and Edelman FC 2010)
- Empirical investigations of trending abuse
  - Uncovering trending abuse tactics (John et al. USENIX Security 2011, Lu et al. CCS 2011)
  - Cloaking measurement (Wang et al. CCS 2011)

# Outline

- 1 How trending search terms are abused
  - Monetizing traffic: malware or ads?
  - Research objectives
- 2 Measuring trending-term abuse
  - Data collection methodology
  - Incidence of abuse
  - How search-term characteristics affect abuse prevalence
- 3 Economics of trending-term exploitation
  - Estimating the exposed population
  - Revenue analysis: ad abuse
  - Revenue analysis: malware
- 4 What happens when search engines intervene?
  - Measuring the effect of Google's intervention
  - Cautionary tale on crackdowns

# Data collection methodology

- 1 Construct a set of trending and control queries
  - *Trending set*: collect 20 Google Hot Trends hourly, and consider a term hot if it has appeared in last 72 hours
  - *Control set*: 495 persistently popular terms (most popular terms in 2010 for 27 categories according to Google)
- 2 Issue queries across multiple search engines
  - Gather top results from Google, Yahoo, Twitter every 4 hours
  - Over 60 million search results and tweets collected
- 3 Classify the search results as malicious or benign
  - *Malware*: Check each URL against Google's Safe Browsing API
  - *MFA*: Supervised machine-learning algorithm classifies websites appearing in results of more than 20 different trending terms

# Total incidence of malware and MFA

	Terms			Results		
	Total	Infected	%	Total	Infected	%
<b>Malware</b>						
<i>Web Search</i>						
Trending set	6 946	1 232	18	9.8M	7 889	.08
Control set	495	123	25	16.8M	7 332	.04
<i>Twitter</i>						
Trending set	1 950	46	2.4	466K	137	.03
Control set	495	53	11	1M	139	.01
<b>MFA sites</b>						
<i>Web Search</i>						
Trending set	19 792	15 181	76.7	32.3M	954K	3.0
<i>Twitter</i>						
Trending set	1 950	1 833	94	466K	32 152	6.9

# More control terms include at least 1 infected result

	Terms			Results		
	Total	Infected	%	Total	Infected	%
<b>Malware</b>						
<i>Web Search</i>						
Trending set	6 946	1 232	18	9.8M	7 889	.08
Control set	495	123	25	16.8M	7 332	.04
<i>Twitter</i>						
Trending set	1 950	46	2.4	466K	137	.03
Control set	495	53	11	1M	139	.01
<b>MFA sites</b>						
<i>Web Search</i>						
Trending set	19 792	15 181	76.7	32.3M	954K	3.0
<i>Twitter</i>						
Trending set	1 950	1 833	94	466K	32 152	6.9

## More trending results are infected

	Terms			Results		
	Total	Infected	%	Total	Infected	%
<b>Malware</b>						
<i>Web Search</i>						
Trending set	6 946	1 232	18	9.8M	7 889	.08
Control set	495	123	25	16.8M	7 332	.04
<i>Twitter</i>						
Trending set	1 950	46	2.4	466K	137	.03
Control set	495	53	11	1M	139	.01
<b>MFA sites</b>						
<i>Web Search</i>						
Trending set	19 792	15 181	76.7	32.3M	954K	3.0
<i>Twitter</i>						
Trending set	1 950	1 833	94	466K	32 152	6.9

## MFA sites much more pervasive than malware

	Terms			Results		
	Total	Infected	%	Total	Infected	%
<b>Malware</b>						
<i>Web Search</i>						
Trending set	6 946	1 232	18	9.8M	7 889	.08
Control set	495	123	25	16.8M	7 332	.04
<i>Twitter</i>						
Trending set	1 950	46	2.4	466K	137	.03
Control set	495	53	11	1M	139	.01
<b>MFA sites</b>						
<i>Web Search</i>						
Trending set	19 792	15 181	76.7	32.3M	954K	3.0
<i>Twitter</i>						
Trending set	1 950	1 833	94	466K	32 152	6.9

## Malware incidence at any given point in time

- By viewing the exposure to malware through search at a single point in time, we find that the risk can be substantial
- Trending terms are more likely to be infected
- Trending terms are more likely to be undetected

	Terms		Results
	#	%	#
<i>Trending terms – web search</i>			
detected	12.8	4.4	14.8
top 10	2.9	1.0	3.2
undetected	6.2	2.1	7.6
top 10	1.2	0.4	1.5
<i>Control terms – web search</i>			
detected	9.5	1.9	14.1
top 10	3.1	0.6	3.9
undetected	1.0	0.2	1.0
top 10	0.1	0.0	0.1

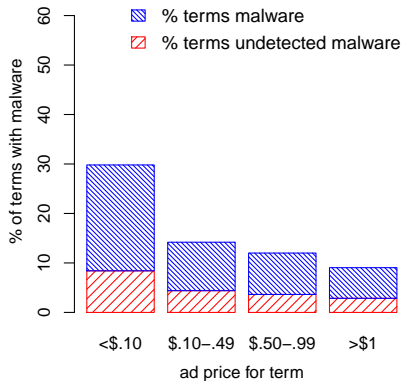
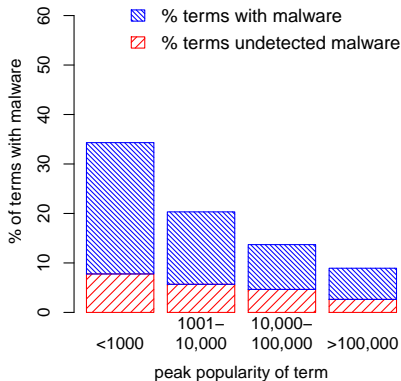


## Does abuse incidence vary by term category?

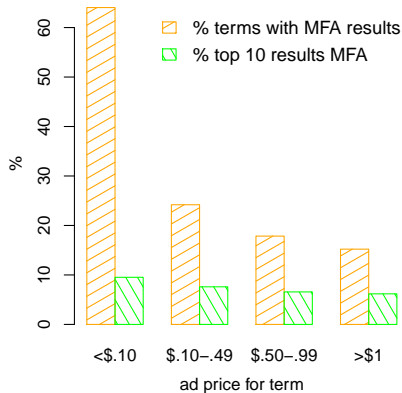
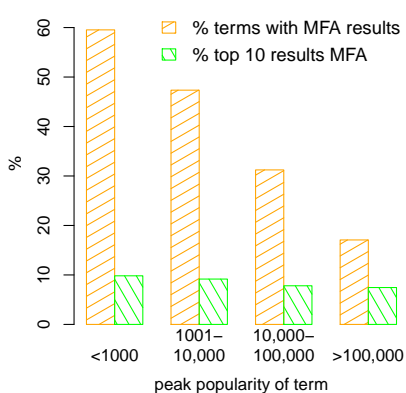
- News, Sports and Local account for over half of all terms
- We found no statistically significant difference in malware incidence across categories
- Some categories were more disposed to MFA (Food&Drink, Reference, Science, Shopping)
- Some categories were less disposed to MFA (e.g., Entertainment, Local, Health)

Category name	Malware		MFA		coef.
	%	CPC	% terms	% top 10	
Arts & Humanities	2.7	\$0.44	20.1	40.6	6.8
Automotive	1.3	\$0.67	16.0	29.2	5.2
Beauty & Personal Care	0.8	\$0.76	19.6	32.5	6.9
Business	0.4	\$0.87	7.4	32.9	6.9
Computers & Electronics	2.4	\$0.61	14.5	31.7	5.9
Entertainment	30.6	\$0.34	18.6	41.0	6.4
Finance & Insurance	1.4	\$1.26	20.2	30.4	5.6
Food & Drink	2.9	\$0.43	17.1	49.5	7.9
Games	2.3	\$0.32	13.4	30.0	5.6
Health	2.5	\$0.85	14.1	27.6	5.9
Home & Garden	0.5	\$0.76	7.1	29.7	7.2
Industries	1.6	\$0.50	26.1	38.6	6.6
Internet	0.7	\$0.49	7.7	43.7	6.0
Lifestyles	4.5	\$0.33	25.4	45.8	6.5
Local	11.0	\$0.51	21.8	39.2	6.9
News & Current Events	3.6	\$0.39	19.7	45.0	7.0
Photo & Video	0.2	\$0.59	0.0	21.9	6.4
Real Estate	0.2	\$1.02	6.2	34.2	6.5
Recreation	1.0	\$0.43	13.7	43.5	6.5
Reference	1.4	\$0.43	14.5	55.4	8.7
Science	1.4	\$0.40	16.0	44.9	9.1
Shopping	3.2	\$0.56	11.6	43.7	8.8
Social Networks	0.5	\$0.19	27.8	59.1	6.4
Society	5.1	\$0.62	15.2	33.7	5.6
Sports	15.4	\$0.38	20.7	44.9	6.9
Telecommunications	0.8	\$0.91	10.9	36.4	4.6
Travel	1.7	\$0.88	10.1	29.3	6.4
<b>Average (category)</b>	<b>3.7</b>	<b>\$0.59</b>	<b>18.4</b>	<b>38.3</b>	<b>6.6</b>

# How a term's popularity and ad price affect malware



# How a term's popularity and ad price affect MFA incidence



# Outline

- 1 How trending search terms are abused
  - Monetizing traffic: malware or ads?
  - Research objectives
- 2 Measuring trending-term abuse
  - Data collection methodology
  - Incidence of abuse
  - How search-term characteristics affect abuse prevalence
- 3 Economics of trending-term exploitation
  - Estimating the exposed population
  - Revenue analysis: ad abuse
  - Revenue analysis: malware
- 4 What happens when search engines intervene?
  - Measuring the effect of Google's intervention
  - Cautionary tale on crackdowns

## Estimated visits to MFA and malware sites

- $V$ : # Visits to a website  $w$  from searching for  $s$  for time period  $t$

$$V(w, s, t) = C(\text{Rank}(w, s)) \cdot \text{Pop}(s) \cdot \frac{4}{30 \times 24} \times t$$

- Click probability
- Website  $w$  position for term  $s$
- Monthly peak popularity of term  $s$

## Estimated visits to MFA and malware sites

On 24 Sep 2010 5:00, a search for “dream act 2010 status” (72 600 searches per month), the following URL appears as the third result in Google:

<http://www.eworldpost.com/dream-act-2010-status-17168.html>

$$V(w, s, t) = C(\text{Rank}(w, s)) \cdot \text{Pop}(s) \cdot \frac{4}{30 \times 24} \times t$$

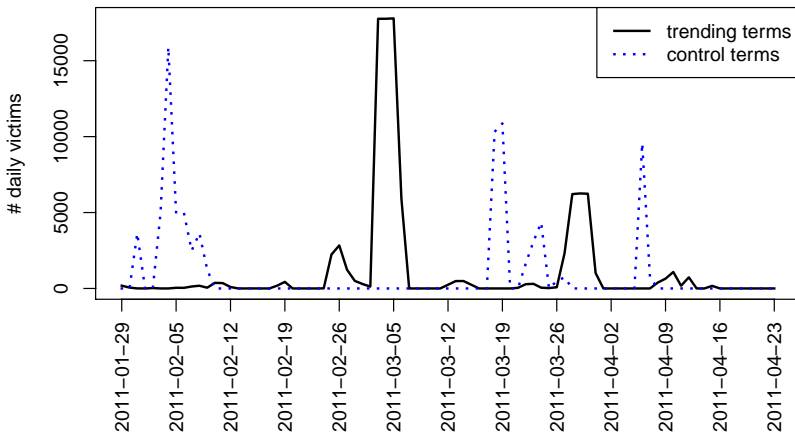
$$V(\text{eworldpost.com}, \text{“dream act 2010 status”}, 1) = C(3) \cdot 72\,600 \cdot \frac{4}{30 \times 24} \times 1$$

$$V(\text{eworldpost.com}, \text{“dream act 2010 status”}, 1) = 44 \text{ visits}$$

## Estimated visits to MFA and malware sites

	Total	# Visitors	
		Period	Monthly Rate
<b>MFA</b>	39 274 200	275 days	<b>4 284 458</b>
<b>Malware (trending set)</b>			
detected	454 198	88 days	154 840
undetected	143 662	88 days	<b>48 975</b>
<b>Malware (control set)</b>			
detected	12 825 332	88 days	4 372 272
undetected	83615	88 days	<b>28 505</b>

# Estimated daily victims to malware in search results





# MFA Revenue Analysis

- Add up all MFA results
- Pay-per-click ads
- Visits per result

$$\sum_{w \in \text{MFA}(s)} \sum_s \left( V(w, s, t) \times \left( p_{\text{PPC}} \cdot p_{\text{clk}} \cdot r_{\text{PPC}} + p_{\text{banner}} \cdot r_{\text{banner}} + p_{\text{aff}} \cdot p'_{\text{clk}} \cdot r_{\text{aff}} \right) \right)$$

- Banner ads
- Pay-per-conversion ads

## MFA revenue for one month

- Pay-per-click ads
  - Visits
- $$4\,300\,000 \times \left( 0.94 \times 0.02 \times \$0.97 + 0.53 \times \$0.00529 + 0.33 \times 0.02 \times \$0.265 \right)$$
- Banner ads
  - Pay-per-conversion ads
  - Monthly MFA revenue: **\$98 000**

# Malware revenue analysis

$$\sum_{w \in \text{mal}(s)} \sum_s \left( V(w, s, t) \cdot p_{\text{exp}} \cdot p_{\text{pay}} \cdot r_{\text{AV}} \right)$$

- Add up all malware results
- Expected # visits per result
- Expected revenue from fake antivirus

## Malware revenue for one month

$$49\,000 \cdot 1 \cdot 0.0216 \cdot \$58$$

- Visits
- Expected revenue from fake antivirus  
(Source: Stone-Gross et al. WEIS 2011)
- Monthly malware revenue: **\$61 000**

# Outline

- 1 How trending search terms are abused
  - Monetizing traffic: malware or ads?
  - Research objectives
- 2 Measuring trending-term abuse
  - Data collection methodology
  - Incidence of abuse
  - How search-term characteristics affect abuse prevalence
- 3 Economics of trending-term exploitation
  - Estimating the exposed population
  - Revenue analysis: ad abuse
  - Revenue analysis: malware
- 4 What happens when search engines intervene?
  - Measuring the effect of Google's intervention
  - Cautionary tale on crackdowns



## The Dirty Little Secrets of Search



Photo illustration by The New York Times

 By DAVID SEGAL  
 Published: February 12, 2011

 PRETEND for a moment that you are [Google's](#) search engine.

## Add to Portfolio

- [+ Yahoo! Inc](#)
- [+ Time Warner Inc](#)
- [+ Penney, J C, Co](#)
- [+ Google Inc](#)
- [+ Bed Bath & Beyond Inc](#)
- [+ AT&T Inc](#)
- [+ Samsonite](#)
- [+ eBay Inc](#)
- [+ Macy's Inc](#)

Someone types the word "dresses" and hits enter. What will be the very first result?

There are, of course, a lot of possibilities. Macy's comes to mind. Maybe a specialty chain, like J. Crew or the Gap. Perhaps a Wikipedia entry on the history of hemlines.

O.K., how about the word "bedding"? Red Bath & Beyond

 Log in to see what your friends are sharing on nytimes.com. [Privacy Policy](#) | [What's This?](#)


## What's Popular Now

 Rare Strong  
 Earthquake Hits  
 Colorado

 Nick Ashford, of  
 Motown Writing  
 Duo, Dies at 70


## MOST POPULAR - BUSINESS DAY

[E-MAILED](#) | [BLOGGED](#) | [VIEWED](#)

1. Hip Implant Complaints Surge, Even as the Dangers Are Studied
2. The Media Equation: Comic's PAC Is More Than a Gag
3. On Food Safety, a Long List but Little Money
4. Advertising: A Magazine Bets That Readers Play Tag
5. Your Money: Your Voice Mail May Be Even Less Secure Than You Thought
6. The Scramble for Access to Libya's Oil Wealth Begins
7. Antibacterial Chemical Raises Safety Issues
8. Local TV Newscasts Expanding
9. Attorney General of N.Y. Is Said to Face Pressure on Bank Foreclosure Deal
10. Media Decoder: Bill Moyers Returns to Public Television, but Not PBS

[Go to Complete List >](#)

RECOMMEND

TWITTER

 COMMENTS  
(156)

E-MAIL

PRINT

REPRINTS

SHARE



Insights from Googlers into our products, technology, and the Google culture.

## Finding more high-quality sites in search

2/24/2011 06:50:00 PM

Our goal is simple: to give people the most relevant answers to their queries as quickly as possible. This requires constant tuning of our algorithms, as new content—both good and bad—comes online all the time.

Many of the changes we make are so subtle that very few people notice them. But in the last day or so we launched a pretty big algorithmic improvement to our ranking—a change that noticeably impacts 11.8% of our queries—and we wanted to let people know what's going on. This update is designed to reduce rankings for low-quality sites—sites which are low-value add for users, copy content from other websites or sites that are just not very useful. At the same time, it will provide better rankings for high-quality sites—sites with original content and information such as research, in-depth reports, thoughtful analysis and so on.

We can't make a major improvement without affecting rankings for many sites. It has to be that some sites will go up and some will go down. Google depends on the high-quality content created by wonderful websites around the world, and we do have a responsibility to encourage a healthy web ecosystem. Therefore, it is important for high-quality sites to be rewarded, and that's exactly what this change does.

It's worth noting that this update does not rely on the feedback we've received from the [Personal Blocklist Chrome extension](#), which we launched last week. However, we did compare the Blocklist data we gathered with the sites identified by our algorithm, and we were very pleased that the preferences our users expressed by using the extension are well represented. If you take the top several dozen or so most-blocked domains from the Chrome extension, then this algorithmic change addresses 84% of them, which is strong independent confirmation of the user benefits.

### Search This Blog

powered by ™

### Follow us on Twitter



[Site Feed](#)



[Google](#)

578K readers

BY FEEDBURNER

[Make Google your homepage](#)

### Blog Archive

Blog Archive ▾

### Labels

[accessibility](#) (30)

[acquisition](#) (19)

[ads](#) (101)

[Africa](#) (17)

[Android](#) (27)

[apps](#) (370)

[April 1](#) (4)

[Asia](#) (31)

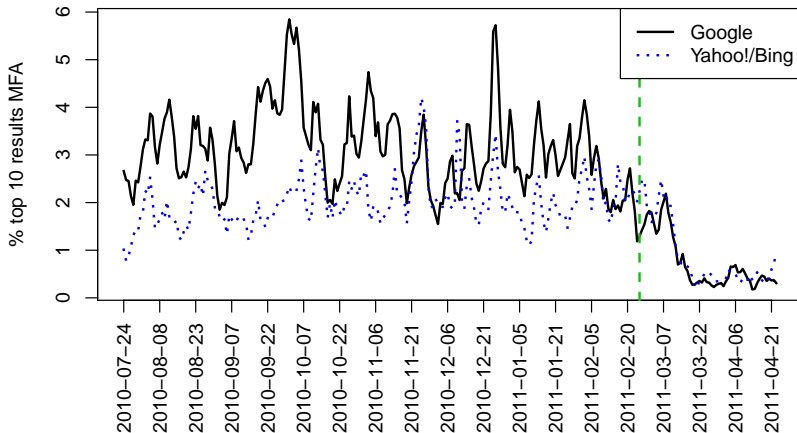
[books + book search](#) (44)

[crisis response](#) (24)

[developers](#) (111)



# Measuring the effect of Google's intervention

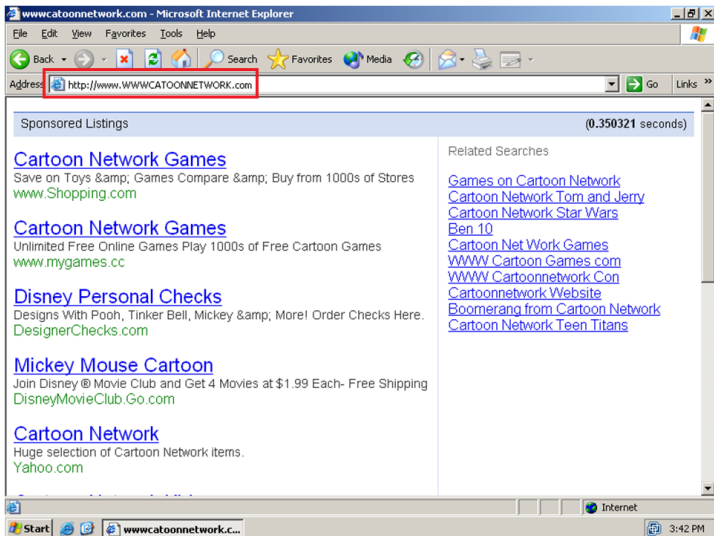




# Implications of Google's intervention for search engines

	Monthly MFA visits		
	Pre-intervention	Post-intervention	% change
Google search	3 364 402	1 788 480	-47%
Yahoo!/Bing search	1 302 314	1 448 058	+11%
<b>Total</b>	<b>4 666 716</b>	<b>3 236 538</b>	<b>-31%</b>

# Cautionary tale on crackdowns: typosquatting



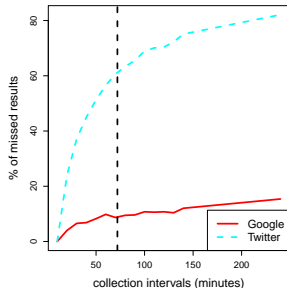
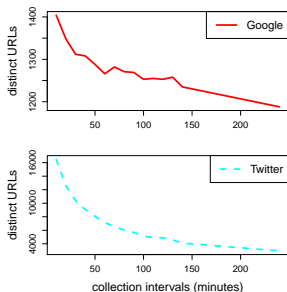
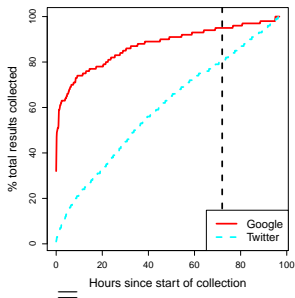
# 1 278 cartoonnetwork.com typos, including...

fartoonnetwork.com	cagtoonnetwork.com	cartiionnetwork.com	cartoonneswork.com	cartoonnework.com
cartoonnetsork.com	cartoinnetwork.com	cartolnnetwork.com	cartoonnftwork.com	cartoonneywork.com
cartoonntewrk.com	cartoonnetlorc.com	cartoonnetowok.com	crtonnetwork.com	cartoonnegwork.com
cargoonnetwork.com	carttoonnnetwork.com	cartoonnetwork.com	cartoonnetwork.com	cartoonnetworkrk.com
cartoonnetwkor.com	catoonnnetwork.com	cartoooonnetwork.com	caoonnetwork.com	cartonbetwork.com
cartoonetgork.com	cartoonnetqork.com	cartoonneetwort.com	cartoonnetwork.com	catoonnetwork.com
cartoonnetworkk.com	caryoonnetwork.com	cartoonnetwork.com	cartoonnetwork.com	cargoonnetwork.com
caretoonnetwork.com	cartoonnetwork.com	cartoonnetworkoer.com	cartoonnetworker.com	chartoonnetwork.com
cartoonnetwokr.com	cartoonnetwokl.com	cartoonnetwoke.com	cartoownnetwork.com	cartoobetwork.com
cartoonnetworkcom.com	nartoonnetwork.com	cartoonnstwork.com	cartoounnetwork.com	cartoonetwork.com
carfoonnetwork.com	cartoonnotwork.com	cartoonnetwok.com	cartoonnnetwor.com	cartonnetworkk.com
cartoopnetwork.com	cartoonnetwogk.com	cartoonnetwaork.com	cartoonntewrok.com	cartoonetweek.com
caqrtoonnetwork.com	cartoonneework.com	cartppnnetwork.com	cartoonnetnetwork.com	cartoonetwork.com
cartoonntwoork.com	catoonneetwork.com	crattoonnetwork.com	cartoonnetweark.com	cartoonnetwork.com
cartoonnetwoirk.com	cartoonznetwork.com	cartoobnetwork.com	catoonnetwork.com	cartiinnetwork.com
cartoonnnetwrk.com	cartoomnetwork.com	cartoonnetwart.com	wwwcartonetwork.com	cartoonntwork.com
cartoonhetwork.com	fcartoonnetwork.com	catoonnetworkk.com	artoonnetwor.com	cartoonnetwock.com
cartoonnetook.com	cartoonnetkwork.com	cartonnetwokr.com	cartonnetwork.com	cartoonetowrk.com
catoonnetworkk.com	cartoo0network.com	cacrtoonnetwork.com	cartoonnetworkorkl.com	cartoonetdwork.com
cartoonnetwcrk.com	cartoonetwrk.com	cartoonnetworkk.com	cartoonnetwoirk.com	cartoknnetwork.com
cartoonnnetwrk.com	cartoonnetbwork.com	caetoonnetwork.com	cartoonknetwork.com	catoonnetwork.com
cartoonnexwork.com	carooonnetwork.com	dartoonnetwork.com	certoonnetwork.com	cartoonnetwork.com
cartoonetworg.com	cartoonetworl.com	cartoonetworj.com	cartoonnetwork.com	cartoonetwort.com
crattonnetwork.com	cartoonnettokr.com	carntoonnetwork.com	caretoonnetwork.com	cartooonnetwork.com
cartoonnerwort.com	cartoonnerwork.com	cartoonnerworl.com	cartoonnetfork.com	cartoonnetttwork.com
cartoonnetwar.com	cartoonnetwak.com	cartoonnekwork.com	cartooknetwork.com	cartoonegwork.com
catoonnetwok.com	cartoonnetworkwork.com	cartoonnetgor.com	cartoonnetwokk.com	wwwcatoonnetwork.com
cartoolnnetwork.com	cartoonnetworkcom.com	casrtoonnetwork.com	cartoonnetworkk.com	cartoonnedwort.com
cartoonnedword.com	cartoonnedwork.com	wwwcarttoonnetwork.com	cartoonnerwork.com	catoonnetworkk.com
carttoonnetwork.com	cartoonnetwovrk.com	cartoonnetwqork.com	cratoonnetwork.com	czrtoonnetwork.com

## Conclusions

- Trending terms are successfully exploited by miscreants more often than consistently popular terms
- Pointing users to malware and ad-filled sites yields similar revenue levels
- Either way, miscreants are only modestly successful
- Google's quality crackdown worked, but beware unintended consequences for the attraction of malware to the bad guys
- For more, see <http://cs.wellesley.edu/~tmoore/>

# Calibration tests balance comprehensiveness and efficiency



## Regression model for malware prevalence

$$\text{logit}(p_{\text{HasMalware}}) = \beta + \text{AdPrice}x_1 + \log_2(\text{Popularity})x_2$$

	coef.	odds	Std. Err.	Significance
AdPrice	-0.509	.601	0.091	$p < 0.001$
$\log_2(\text{Popularity})$	-0.117	0.889	0.012	$p < 0.001$

## Regression model for MFA prevalence

$$\text{FracTop10MFA} = \beta + \text{AdPrice}x_1 + \log_2(\text{Popularity})x_2 + \text{Category}x_3 .$$

	coef.	Std. Err.	Significance
AdPrice	-0.0091	0.091	$p < 0.001$
$\log_2(\text{Popularity})$	-0.004	0.012	$p < 0.001$
Coefficients for category variables in earlier Table, $R^2$ : 0.1373			