Fashion Crimes: Trending-Term Exploitation on the Web

Tyler Moore
Harvard University
tmoore@seas.harvard.edu

Nektarios Leontiadis
Carnegie Mellon University
leontiadis@cmu.edu

Nicolas Christin
Carnegie Mellon University
nicolasc@cmu.edu

ABSTRACT

Online service providers are engaged in constant conflict with miscreants who try to siphon a portion of legitimate traffic to make illicit profits. We study the abuse of "trending" search terms, in which miscreants place links to malware-distributing or ad-filled web sites in web search and Twitter results, by collecting and analyzing measurements over nine months from multiple sources. We devise heuristics to identify ad-filled sites, report on the prevalence of malware and ad-filled sites in trending-term search results, and measure the success in blocking such content. We uncover collusion across offending domains using network analysis, and use regression analysis to conclude that both malware and ad-filled sites thrive on less popular, and less profitable trending terms. We build an economic model informed by our measurements and conclude that ad-filled sites and malware distribution may be economic substitutes. Finally, because our measurement interval spans February 2011, when Google announced changes to its ranking algorithm to root out low-quality sites, we can assess the impact of search-engine intervention on the profits miscreants can achieve.

Categories and Subject Descriptors

K.4.1 [**Public Policy Issues**]: Abuse and crime involving computers

General Terms

Measurement, Security, Economics

Keywords

Online crime, search engines, malware, advertisements

1. INTRODUCTION

News travels fast. Blogs and other websites pick up a news story only about 2.5 hours on average after it has been reported by traditional media [21]. This leads to an almost continuous supply of new "trending" topics, which are then amplified across the Internet, before fading away relatively quickly.

CCS'11, October 17–21, 2011, Chicago, Illinois, USA.

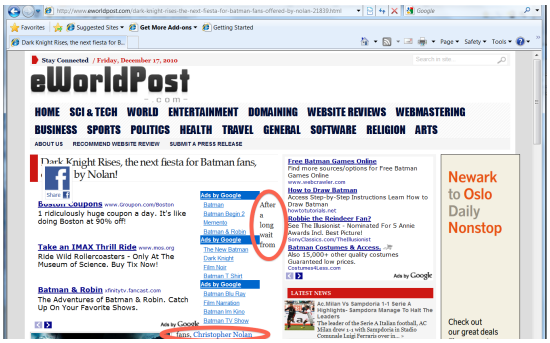


Figure 1: Ad-filled website appearing in the results for trending terms (only 8 words from the article, circled, appear on screen).

However narrow, these first moments after a story breaks present a window of opportunity for attackers to infiltrate web and social network search results in response. The motivation for doing so is primarily financial. Websites that rank high in response to a search for a trending term are likely to receive considerable amounts of traffic, regardless of their quality. Web traffic can in turn be monetized in a number of ways, as shown in related work [6, 10, 17, 20]. In short, manipulation of web or social network search engine results can be a profitable enterprise for its perpetrators.

In particular, the sole goal of many sites designed in response to trending terms is to produce revenue through the advertisements that they display in their pages, without providing any original content or services. Figure 1 presents a screenshot for `eworldpost.com`, which has appeared in response to 549 trending terms between July 2010 and March 2011. The actual article (circled) is hard to find, when compared to the amount of screen real estate dedicated to ads. Such sites are often referred to as "Made for AdSense" (MFA) after the name of the Google advertising platform they are often targeting. Whether such activity is deemed to be criminal or merely a nuisance remains an open question, and largely depends on the tactics used to prop the sites up in the search-engine rankings. Some other sites devised to respond to trending terms have more overtly sinister motives. For instance, a number of malicious sites serve malware in hopes of infecting visitors' machines [30], or peddle fake anti-virus software [2, 8, 36].

Both MFA and malware-hosting sites are enough of a scourge to trigger response from search engine operators. Google modified its search algorithm in February 2011 in part to combat MFA sites [35], and has long been offering the Google Safe Browsing API to block malware-distribution sites. Trending-term exploitation makes both MFA and malware sites even more dynamic than they used to be, thereby complicating the defenders' task.

This paper provides the first large-scale measurement and analysis of trending-term exploitation on the web. Based on a collection of over 60 million search results and tweets gathered over nine months, we characterize how trending terms are used to perform web search-engine manipulation and social-network spam. An important feature of our work is that we bring an outsider's perspective. Instead of relying on proprietary data tied to a specific search engine, we use comparative measurements of publicly observable data across different web search engines (Google, Yahoo!/Bing) and social network (Twitter) posts.

Our specific contributions are as follows. We (1) provide a methodology to automate classification of websites as MFA, (2) show salient differences between tactics used by MFA site operators and malware peddlers, (3) construct an economic model to characterize the trade-offs between advertising and malware as monetization vectors, quantifying the potential profit to the perpetrators, and (4) examine the impact of possible intervention strategies.

The rest of this paper is organized as follows. We introduce our measurement and classification methodology in Section 2. We analyze the measurements collected in Section 3 to characterize trending-term exploitation on the web. Notably, we uncover collusion across offending domains using network analysis, and we use regression analysis to conclude that both malware and MFA sites thrive on less popular and profitable trending terms. We then use these findings to build an economic model of attacker revenue in Section 4, and examine the effect of search-engine intervention in Section 5. We compare our study with related work in Section 6, before drawing brief conclusions in Section 7.

2. METHODOLOGY

We start by describing our methodology for data collection and website classification. At a high level, we need to issue a number of queries on various search engines for current trending terms, follow the links obtained in response to these queries, and classify the websites we eventually reach as malicious or benign. Within the collection of malicious sites so obtained, we have to further distinguish between malware-hosting sites and ad-laden sites. Moreover, we need to compare the results obtained with those collected from "ordinary," rather than trending, terms.

The data collection hinges on a number of design choices that we discuss and motivate here. Specifically, we must determine how to build the corpus of trending terms to use in queries ("*trending set*"); identify a set of control queries ("*control set*") against which we can compare responses to queries based on trending terms; decide on how frequently, and for how long, we issue each set of queries; and find mechanisms to classify sites as benign, malware-distributing, and MFA.

2.1 Building query corpora

Building a corpus of trending terms is not in itself a challenging exercise. Google, through Google Hot Trends [15], provides a list of twenty current "hot searches," which we determined, through pilot experiments, to be updated hourly. Likewise, Twitter avails a list of ten trending topics [37] and Yahoo! gives a "buzz log" [38] containing the 20 most popular searches over the past 24 hours.

These different lists sometimes have very little overlap. For instance, combining the 20 Yahoo! Buzz logs, 20 Google Hot Trends, and 10 Twitter Trending Topics, it is not uncommon to find more than 40 distinct trending terms over short time intervals. This would seem to make the case for aggregating all sources to build our query corpus. However, all search APIs limit the rate at which queries can be issued. We thus face a trade-off between the time granularity of our measurements and the size of our query corpus.

Trending set. Fortunately, we can capture most of the interesting patterns we seek to characterize by solely focusing on Google Hot Trends. Indeed, a recent measurement study conducted by John et al. [17] shows that over 95% of the terms used in search engine manipulation belong to the Google Hot Trends. However, because Twitter abuse may not necessarily follow the typical search engine manipulation patterns, we use both Google Hot Trends and the Twitter current trending topics in our Twitter measurements.

Hot trends, by definition, are constantly changing. We update our trending term corpus every hour by simply adding the current Google Hot Trends to it. Determining when a term has "cooled" and should be removed from the query corpus is slightly less straightforward. We could simply remove terms from our query corpus as soon as they disappear from the list of Google Hot Trends. However, unless all miscreants stop poisoning search results with a given term as soon as this term has "cooled," we would likely miss a number of attempts to manipulate search engine results. Furthermore, Hot Trends are selected based upon their rate of growth in query popularity. Terms that have fallen out of the list in most cases still enjoy a sustained period of popularity before falling.

We ran a pilot experiment collecting Google and Twitter search results on 20 hot terms for up to four days. As Figure 2(a) shows, 95% of all unique Google search results and 81% of Twitter results are collected within three days. Thus, we settled on searching for trending terms while they remain in the rankings, plus up to three days after they drop out of the rankings.

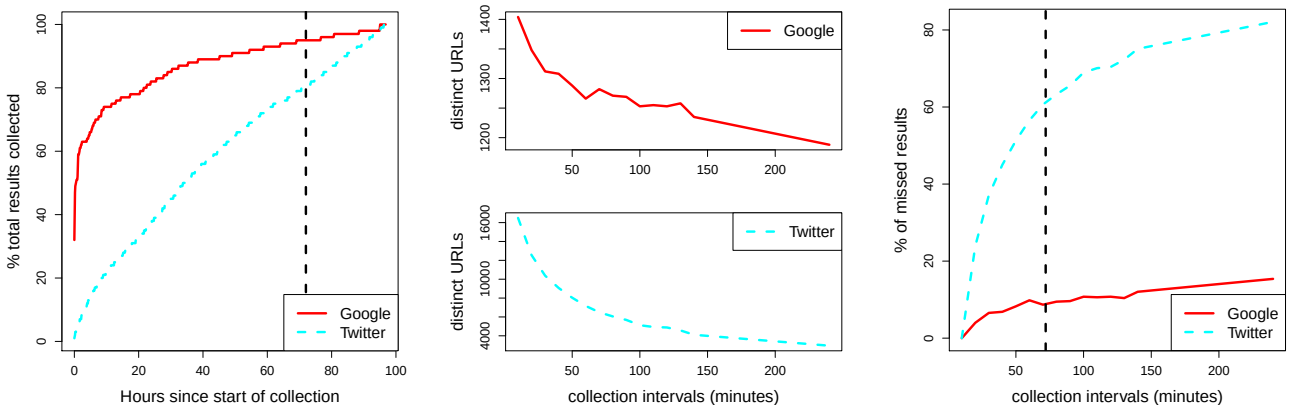
Control set. It is necessary to compare results from the trending set to a control set of consistently popular search terms, to identify which phenomena are unique to the trending nature of the terms as opposed to their overall popularity. We build a control list of the most popular search terms in 2010 according to Google Insights for Search [13]. Google lists the top 20 most popular search terms for 27 categories. These reduce to 495 unique search terms, which we use as a control set.

2.2 Data collection

For each term in our trending and control sets, we run automated searches on Google and Yahoo! between July 24, 2010 and April 24, 2011. We investigate MFA results throughout the sample, and study the timeliness of malware identification between January 26 and April 24, 2011. We study Twitter results gathered between March 10 and April 18, 2011.

We use the Google Web Search API [1] to pull the top 32 search results for each term from the Google search engine, and the Yahoo! BOSS API to fetch its top 100 Yahoo! results for each term. Since the summer of 2010 Yahoo! and Bing search results are identical [23]. Consequently, while in the paper we refer to Yahoo! results, they should also be interpreted as those appearing on Bing. Likewise, we use the Twitter Atom API to retrieve the top 16 tweets for each term in Google's Hot Trends list and Twitter's Current Trends list. We resolve and record URLs linked from tweets, as well as the authors of these tweets linking to other sites.

Because all these APIs limit the number of queries that can be run, we had to limit the frequency with which we ran the search queries. To better understand the trade-offs between search frequency and comprehensiveness of coverage, we selected 20 terms from a single trending list and ran searches using the Google API every 10 minutes for one week. We then compared the results we could obtain using the high-frequency sampling to what we found when sampling less often. The results are presented in Figures 2(b) and 2(c). Sampling once every 20 minutes, rather than every 10 minutes, caused 4% of the Google search results to be missed. Slower intervals caused more sites to be missed, but only slightly:



(a) New search results as a function of time in Twitter and Google. More than 80% of Google results appear within 3 days, while Twitter continuously produces new results.

(b) Number of distinct URLs collected using different collection intervals. The measurement lasted for two weeks using a fixed set of terms that was trending at the beginning of the experiment.

(c) Number of distinct URLs that we failed to collect using different collection intervals. The measurement lasted for two weeks using a fixed set of terms that was trending at the beginning of the experiment

Figure 2: Calibration tests weigh trade-offs between comprehensiveness and efficiency for collecting trending-term results.

85% of the search results found when reissuing the query every 10 minutes could also be retrieved by sampling only once every 4 hours. So, even for trending topics, searching for the hot terms once every four hours provides adequate coverage of Google results. For consistency, we used the same interval on Twitter despite the higher miss rate. Twitter indeed continues producing new results over a longer time interval, primarily due to the "Retweet" function which allows users to simply repost existing contents.

2.3 Website classification

We next discuss how we classified websites as benign, malware-distributing, or ad-filled. We define a website as a set of pages hosted on the same second-level DNS domain. That is, `this.example.com` and `that.example.com` belong to the same website.[1] While we realize that different websites may be hosted on the same second-level domains, they are ultimately operated or endorsed by the same entity – the owner of the domain. Hence, in a slight abuse of terminology we will equivalently use "website" and "domain" in the rest of this discussion.

Malware-distributing sites. We pass all search results to Google's Safe Browsing API, which indicates whether a URL is currently infected with malware by checking it against a blacklist. Because the search results deal with timely topics, we are only interested in finding which URLs are infected near the time when the trending topic is reported. However, there may be delays in the blacklist updates, so we keep checking the results against the blacklist for 14 days after the term is no longer hot.

When a URL appears in the results and is only later added to the blacklist, we assume that the URL was already malicious but not yet detected as such. It is, of course, also possible that the reason the URL was not in the blacklist is that the site had not yet been infected. In the case of trending terms, however, a site appearing in results indicates a likely compromise, since the attacker's modus operandi is to populate compromised web servers with content that reflects trending results [17].

The possibility of later compromise further justifies our decision to stop checking the search results against the blacklist after two weeks have passed. While it is certainly possible that some malware takes more than two weeks to be detected, the potential for prematurely flagging a site as compromised also grows with time. Indeed, in a study of spam on Twitter [10], the majority of tweets flagged by the Google Safe Browsing API as malicious were not added to the blacklist until around a month had passed. We suspect that many of the domains marked as malicious were in fact only compromised much later. Consequently, our decision to only flag malware detected within two weeks is a conservative one that minimize false positives while slightly increasing false negatives.

Dealing with long-delayed reports of malware poses an additional issue for terms from the control set, because these search results are more stable over time. Sometimes a URL appears in the results of a term for years. If that website becomes infected, then it would clearly be incorrect to claim that the website was infected but undetected the entire time. In fact, most malware appearing in the results for the control set are for websites that have only recently "pushed" their way into the top search results after having been infected. For these sites, delays in detection do represent harm.

We thus exclude from our analysis of malware in the control set URLs that appeared in the results between December 20-31, 2010, when we began collecting results for the control set. To eliminate the potential for edge effects, our analysis of malware does not begin until January 26, 2011. As in the trending set, we also only flag results as malware when they are detected within 14 days.

Finally, we note that sometimes malware is undetected by the SafeBrowsing API on the top-level URL, but that URLs loaded externally by the website are blocked. Consequently, our analysis provides an upper bound on malware success.

MFA sites. Automated identification of MFA sites is a daunting task. There are no clear rules for absolutely positive identification, and even human inspection suggests a certain degree of subjectivity in the classification. We discuss here a set of heuristics we use in determining whether a site is MFA or not.

While 182 741 different domains appeared in the top 32 Google and Yahoo! search results for trending terms over 9 months, only 6 558 (3.6%) appeared in the search results for at least 20 different trending terms. Because the goal of MFA sites is to appear high in the search results for as many terms as possible, we investigate further which of these 6 558 websites are in fact legitimate sources of information, and which are low-quality, ad-laden sites. To that effect we selected a statistically significant (95% confidence interval)

[1] So do `this.example.co.uk` and `that.example.co.uk`, as `co.uk` is considered a top-level domain; as are a few others (e.g., `ac.jp`) for which we maintain an exhaustive list.

random sample of 363 websites for manual inspection. From this sample, we identified five broad categories of websites indicative of MFA sites. All MFA sites appear to include a mechanism for automatically updating the topics they cover; differences emerge in how the resulting content is presented.

1. Sites which reuse snippets created by search engines and provide direct links to external sites with original content (e.g., `http://newsblogged.com/tornado-news-latest-real-time`).

2. Sites in blog-style format, containing a short paragraph of content that is likely copied from other sources and only slightly tweaked – usually by a machine algorithm, rather than a human editor (e.g, `http://toptodaynews.com/water-for-elephants-review`).

3. Sites that automatically update to new products for sale pointing to stores through paid advertisements (for instance, `http://tgiblackfriday.com/Online-Deals/-261-up-Europe-On-Sale-Each-Way-R-T-required--deal`).

4. Sites aggregating content by loading external websites into a frame so that they keep the user on the website along with their own overlaid ads (e.g., `http://baltimore-county-news.newslib.com/`).

5. Sites containing shoddily, but seemingly manually written content based on popular topics informed by trending terms (e.g., `http://snarkfood.com/mel-gibsons-mistress-says-hes-not-racist/310962/`).

Based on manual inspection of our random sample of 363 sites, we decided to classify websites in any of the first four categories as MFA, while rejecting sites in the fifth category. (Including those would have driven up the false positive rate to unacceptable levels.) This results in 44 of the 363 websites being tagged as MFA.

Subsequently, we used a supervised machine-learning algorithm (Bayesian Network [29] constructed using the K2 algorithm [7]) to automatically categorize the remaining 6 195 candidate websites.

The set of measures used to describe each page is a combination of structural and behavioral characteristics: (1) the number of internal links, i.e. links to the same domain as the web page under examination; (2) the number of external links, i.e. links directed to external domains; and (3) the existence of advertisements in the web page. We calculate these three quantities for each of the 6 558 domains by parsing the front page of the domain and a set of five additional web pages within the same domain, randomly chosen among the direct links existing in the front page.

We experimented with many more features in the classifier (e.g,, time since the website was registered, private WHOIS registration, number of trending terms where a website appears in the search results, presence of JavaScript, etc). As manual inspection confirmed, this did not improve classification accuracy beyond the three features described in the paper. MFA sites exhibit large numbers of external links but few internal links, because unlike external links to ads, internal links do not (directly) generate revenue.

We determine whether a website has advertisements by looking for known advertising domains in the collected HTML. Because these domains often appear in JavaScript, we use regular expressions to search throughout the page. We use manually-collected lists of known advertising domains used by Google and Yahoo!, complemented by the "Easy List" maintained by AdBlock Plus [3] (Jan. 12, 2011).

We used a subset of the 363 sample domains as a training set for the machine learning algorithm. We did not use the entire set because it is overcrowded with non-MFA domains (87% non-MFA vs. 13% MFA), which would lead to over-training the model towards non-MFA websites. By using fewer non-MFA websites in the training set (80% vs. 20%), we kept our model biased towards non-MFA websites, thereby maintaining the assumption of innocence while remaining able to identify obvious MFA instances.

We assessed the quality of our predictive model by performing 10 rounds of cross-validation [19], yielding a 87.3% rate of successful classifications. In the end, the algorithm classified 838 websites (0.46% of all collected domains) that appear in the trending set results as MFAs. The relatively small number of positive identifications allows for manual inspection to root out false positives. We find that 120 of the websites (consistent with the predicted 87.3% success rate) are likely false positives. We remove these websites from consideration when conducting the subsequent quantitative analysis of MFA behavior.

	Terms			**Results**			**URLs**			**Domains**		
	Total	Inf.	%	Total	Inf.	%	Total	Inf.	%	Total	Inf.	%
Malware												
Web Search												
Trending set	6 946	1 232	18	9.8M	7 889	.08	607K	1 905	.30	109K	495	.50
Control set	495	123	25	16.8M	7 332	.04	231K	302	.13	86K	123	.14
Twitter												
Trending set	1 950	46	2.4	466K	137	.03	355K	101	.03	43K	13	.03
Control set	495	53	11	1M	139	.01	825K	129	.02	98K	101	.02
Twitter trnd.	1 176	20	1.7	180K	24	.01	139K	21	.02	26K	9	.03
MFA sites												
Web Search												
Trending set	19 792	15 181	76.7	32.3M	954K	3.0	1.35M	83 920	6.2	183K	629	.34
Twitter												
Trending set	1 950	1 833	94	466K	32 152	6.9	355K	32 130	9.0	43K	141	.3
Twitter trnd.	1 176	1 012	86	179K	12 145	6.6	139K	12 144	8.7	26K	42	.2

Table 1: Total incidence of malware and MFA in web search and Twitter results.

3. MEASURING TRENDING-TERM ABUSE

3.1 Incidence of abuse

We now discuss the prevalence of malware and MFA in the trending search results. There are many plausible ways to summarize tens of millions of search results for tens of thousands of trending terms gathered over several months. We consider four categories: terms affected, search results, URLs and domains.

Table 1 presents totals for each of these categories. For web search, we observed malware in the search results of 1 232 of the 6 946 terms in the trending set. Running queries six times a day over three months yielded 9.8 million search results. Only 7 889 of these results were infected with malware – 0.08% of the total. These results corresponded to 607 156 unique URLs, only 1 905 of which were infected with malware. Finally, 495 of the 108 815 domains were infected.

How does this compare to popular search terms? As a percentage, more control terms were infected with malware, but that is due to their persistent popularity. Around the same number of search results were infected, but the control set included nearly twice as many overall results (because there were around 300 trending terms "hot" at any one time compared to the 495 terms always checked in the control set). 1 905 URLs were infected in the trending set, compared to only 302 in the control set.

The prevalence of malware on Twitter is markedly lower: only 2.4% of terms in the trending set were found to have malware, compared to 18% for search, and only 101 URLs on 13 distinct domains were found infected. While the number of infections observed is very small (0.03%), it is consistent with the proportion of malicious URLs observed by Grier et al. [10] on a significantly larger dataset of 25 million unique URLs. The control and Twitter-trending sets also reveal similarly low levels of infection.

	Terms		Results	Domains		URLs	
	#	%	#	#	%	#	%
Trending terms – web search (point in time)							
detected	12.8	4.4	14.8	13.8	0.089	8.7	0.146
top 10	2.9	1.0	3.2	3.1	0.020	2.4	0.040
undetected	6.2	2.1	7.6	6.7	0.0	3.718	0.061
top 10	1.2	0.4	1.5	1.4	0.009	0.9	0.015
Control terms – web search (point in time)							
detected	9.5	1.9	14.1	11.5	0.043	8.9	0.067
top 10	3.1	0.6	3.9	3.7	0.014	3.1	0.023
undetected	1.0	0.2	1.0	1.0	0.0	0.856	0.006
top 10	0.1	0.0	0.1	0.1	0.000	0.1	0.001

Table 2: Prevalence of malware in trending and control terms, presented as the average prevalence of malware at every point in time when searches are issued.

Grier et al. observed a much higher proportion of "spammy" behavior on Twitter. Likewise, we observe substantial promotion of MFA websites on Twitter: 94% of trending terms contained tweets with MFA domains. While most terms are targeted, only a small number of domains are promoted – 141 in the trending set and 42 in the Twitter-trending set. Web search is also targeted substantially by MFA sites. 77% of terms in the trending set included one or more of the 629 MFA domains in at least one result.

From the figures in Table 1 alone, it would appear that malware on trending terms is largely under control, while MFA sites are relatively rampant. However, aggregating figures across a large period of time can obscure the potential harm of malware distributed via trending terms. Table 2 presents the malware infection rate at a single *point in time*: counting the number of terms and search results that are infected with malware for each of the trending terms within a 3-day window of rising. For example, on average, 12.8 trending terms are infected with malware that has already been flagged by the Safe Browsing API, which corresponds to 4.4% of recently hot terms at any given moment. A further 6.2 trending terms are infected but not yet detected by the blacklist. On average, 1.2 terms include a top 10 result that distributes malware and has not yet been detected by the Safe Browsing API. Viewed in this manner, the threat from web-based malware appears more worrisome.

But is the threat worse for trending terms? 9.5 control terms include detected malware at a given point in time, with one term infected but not yet detected. Hence, popular terms are still targeted for malware, but less frequently and with less success. Finally, the false negative rate for the trending set is much higher than for the control set: 34% (7.6 results undetected compared to 14.8 detected) vs. 7% (1 undetected result compared to 14.1 detected).

3.2 Network characteristics

We next turn to characterizing how sites preying on trending terms are connected to each other. To prop up their rankings in Google, one would expect a group of sites operated by a same entity to link to each other – essentially building a "link farm [11]." Thus, we conjecture that looking at the network structure of both MFA and malware-serving sites may yield some insight on both the actors behind these attacks, and the way campaigns are orchestrated.

MFA domains. We build a directed graph G_{MFA} where each node corresponds to one of the 629 domains we identified as MFA sites, and each of the 3 221 (directed) edge corresponds to an HTML link between two domains. We construct the graph by fetching 1 000 backlinks for each of the sites from Yahoo! Site Explorer [39]. Extracting the strongly connected components from G_{MFA} yields family of sites that link to each other. We find 407 distinct strongly connected components, most (392) only contain singletons. More interestingly, 193 sites (30.7% of all MFA sites) form a strongly connected component. These nodes have on average a degree (in- and out-links) of 12.83, and an average path length between two nodes of 3.92, indicating a quite tightly connected network. It thus appears that a significant portion of all MFA domains may be operated by the same entity – or at the very least, by a unique group of affiliates all linking to each other. Further inspecting where these sites are hosted indicates that 130 of the 193 sites belong to one of only seven distinct Autonomous Systems (AS); here, sites within a same AS are usually hosted by the same provider, which confirms the presence of a fairly large, collusive, MFA operation.

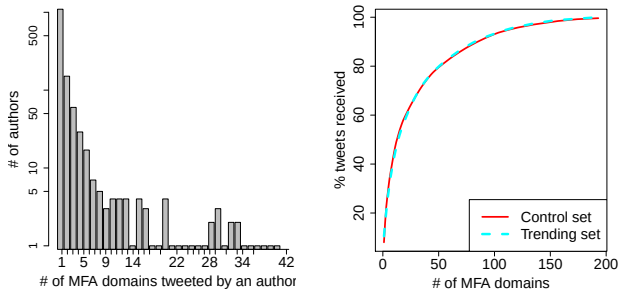
Campaign ID	# Domains	Duration	Distinct ASes
949	590	>1 year	>200
5100	36	>8 months	1
5101	25	>8 months	1
5041	11	4 days	2
5053	10	2 days	1
4979	9	11 days	2
4988	9	8 days	2

Table 3: Malware campaigns observed.

Malware-serving sites. Examining the network characteristics of malware-distributing sites serves a slightly different purpose. Here, sites connected to each other are unlikely to be operated by the same entity, but are likely to have been *compromised* by the same group or as part of the same campaign. This is consistent with the behavior observed by John et al. [17], who found that miscreants add links between malicious websites to elevate PageRank. As with MFA sites, we build a directed graph G_{mal} where each node corresponds to one of the 6 133 domains we identified as malware-serving based on a longer collection of trending terms gathered from April 6, 2010 to April 27, 2011. Each (directed) edge corresponds to an HTML link between two malware-serving domains. G_{mal} contains 6 133 nodes and 18 864 edges, and 5 125 distinct strongly connected components, only 216 of which contain more than one node. Table 3 lists the largest strongly connected components ("campaigns") in G_{mal}. For each of the nodes in these campaigns, we look up the time at which they were first listed as infected. By comparing the first and last nodes to be infected within a given campaign, we can infer the campaign's duration. We also look up the number of distinct ASes in each campaign.

We observe divergent campaign behaviors, each characterized by markedly different attacker tactics. The largest campaign (949) is still ongoing at the time of this writing: nodes are compromised at a relatively constant rate, and are hosted on various ASes. This indicates a long-term, sustained effort. This campaign affects at least 9.6% of all the malware-infested sites we observed. Campaigns 5100 and 5101 are likely part of the same effort: all nodes share the same set of servers, and seem compromised by the same exploit. Interestingly, this campaign went unabated for at least 8 months (until Dec. 2010). Finally, the other four notable campaigns we observed target small sets of servers, that are compromised almost simultaneously, and all immediately link to each other.

Our definition of a campaign is extremely conservative: we are only looking for strongly connected components in the graph we have built. It is thus likely that many of the singletons we observed are in fact part of larger campaigns. Further detection of such campaigns would require more complex clustering analysis. For instance, one could try to use the feature set of the classifi-



(a) Number of Twitter authors posting unique URLs containing MFA domains. Most authors post up to 14 distinct domains.

(b) CDF of tweets associated with MFA domains. The x-axis shows the number of domains associated with a portion of the tweets.

Figure 3: Trending-term exploitation on Twitter.

cation algorithm as a coordinate system, and cluster nodes with nearby coordinates. However, it is unclear that this specific coordinate system would provide definitive evidence of collusion.

3.3 MFA in Twitter

We turn our attention now to the use of MFA links in Twitter posts. We are interested in measuring the amount of unique MFA-related URLs each malicious user posts, and the popularity of the MFA websites among them.

Figure 3(a) shows that 95% of the authors who post MFA URLs link to 5 domains or less – this amounts to about 20 000 posts. However, the remaining 5% is responsible for about 55 000 posts, and links to 870 domains. The control set gives similar numbers.

In other words, a small number of authors are responsible for wide promotional campaigns of MFA websites. The vast majority of authors post a small number of MFA links, and it is unclear whether they are actually malicious or not.

Similarly, the number of MFA domains that receive the majority of related tweets is small as Figure 3(b) shows. 50% of the MFA infected tweets direct users to 14 MFA domains, with the remaining 50% distributed across 180 MFA domains.

3.4 Search-term characteristics

We now examine how characteristics of the trending terms themselves influence the prevalence of malware and MFA sites in their search results. We focus on the importance of the term's category, popularity in searches, and expected advertising revenue.

Measuring term category, popularity and ad prices. We combine results from several Google tools in order to learn more about the characteristics of each of the trending terms. First, we classify the trending terms into categories using Google Insights for Search, which assigns arbitrary search terms to the most likely category, out of the same 27 categories used for constructing the control sample described in Section 2.1 above.

Second, Google offers a free service called Traffic Estimator that estimates for any phrase the number of global monthly searches averaged over the past year [14]. For trending terms, averaging over the course of a year significantly underestimates the search traffic when a term is peaking in popularity. Fortunately, Google also offers a measure of the relative popularity of terms through Google Trends [15], provided at the granularity of one week. The relative measure is normalized against the average number of searches for the past year, precisely the figure returned by the Traffic Estimator. We obtain the *peak-popularity estimate* $\mathrm{Pop}(s)$ for a term s by multiplying the relative estimate for the week when the term peaked by the absolute long-run popularity estimate.

Category name	%	CPC	Malware % terms	MFA % terms	MFA % top 10	coef.
Arts & Humanities	2.7	$0.44	20.1	40.6	6.8	
Automotive	1.3	$0.67	16.0	29.2	5.2	−0.0062
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	−0.0043
Finance & Insurance	1.4	$1.26	20.2	30.4	5.6	
Food & Drink	2.9	$0.43	17.1	49.5	7.9	+0.0105
Games	2.3	$0.32	13.4	30.0	5.6	−0.0073
Health	2.5	$0.85	14.1	27.6	5.9	−0.0046
Home & Garden	0.5	$0.76	7.1	29.7	7.2	
Industries	1.6	$0.50	26.1	38.6	6.6	−0.0072
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	−0.0027
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	+0.0203
Science	1.4	$0.40	16.0	44.9	9.1	+0.0095
Shopping	3.2	$0.56	11.6	43.7	8.8	+0.0106
Social Networks	0.5	$0.19	27.8	59.1	6.4	
Society	5.1	$0.62	15.2	33.7	5.6	−0.0085
Sports	15.4	$0.38	20.7	44.9	6.9	−0.0044
Telecommunications	0.8	$0.91	10.9	36.4	4.6	
Travel	1.7	$0.88	10.1	29.3	6.4	
Average (category)	**3.7**	**$0.59**	**18.4**	**38.3**	**6.6**	

Table 4: Malware and MFA incidence broken down by trending term category.

The Google Traffic Estimator also indicates the advertising value of trending terms, by providing estimates of the anticipated cost per click (CPC) for keywords. We collect the CPC for all trending and control terms. Many trending terms are only briefly popular and return the minimum CPC estimate of US$0.05. We use the CPC to approximate the relative revenue that might be obtained for search results on each term. The CPC is a natural proxy for the prospective advertising value of user traffic because websites that show ads are likely to present ads similar to the referring term.

Empirical analysis. Table 4 breaks down the relative prevalence of trending terms, and their abuse, by category. Over half of the terms fall into three categories – Entertainment, Sports and Local. These categories feature topics that change frequently and briefly rise from prior obscurity. 18% of trending terms include malware in their results, while 38% feature MFA websites in at least 1% of the top 10 results.

We observe some variation in malware and MFA incidence across categories. However, perhaps the most striking result from examining the table is that all categories are targeted, irrespective of the category's propensity to "trendiness." Miscreants do not seem to be specializing yet by focusing on particular keyword categories.

If we instead look at popularity and ad prices, substantial differences emerge. Figure 4 shows how the incidence of malware and MFA varies according to the peak popularity and ad price of the trending term. The left-most graph shows how malware varies according to the term's peak popularity. The least popular terms (less than 1 000 searches per day at their peak) attract the most malware in their top results. 38% of such terms include malware, while 9% of these terms include malware that is not initially detected. As

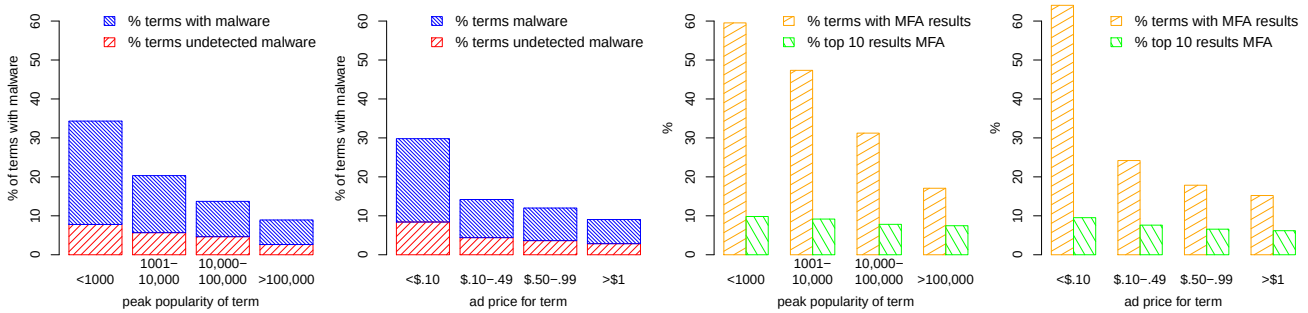


Figure 4: Exploring how popularity and ad price of trending terms affects the prevalence of malware (left) and ad-laden sites (right).

terms increase in peak popularity, fewer are afflicted by malware: only 6.2% of terms with peak popularity greater than 100 000 daily searches include malware in their results, and only 2% of terms include malware that is not immediately detected. A similar pattern follows for malware incidence according to the term's ad price. 30% of terms with ad prices under 10 cents per click had malware in their results, compared to 8.8% of terms with ad prices greater than $1 per click.

A greater fraction of terms overall include MFA websites in their results than malware (37% vs. 19%). Consequently, all proportions are larger in the two graphs on the right side of Figure 4. 60% of terms with peak popularity of less than 1 000 daily searches include MFA sites in their results. This proportion drops steadily until only 17.4% of terms attracting over 100 000 daily visits include MFA in the top 10 search results. A similar reduction can be seen for varying ad prices in the right-most figure. The two right figures show the percentage of all terms that have MFA, followed by the percentage of top 10 results that are MFA, for only those terms that have MFA terms present. Here we can see that the percentages remain relatively steady irrespective of term popularity and price. For unpopular terms, 10% of their results point to MFA, dropping modestly to 8% for the most popular terms. The drop is more significant for ad prices – from 10% to 6%. Consequently, while the success in appearing in results diminishes with popularity and rising ad prices, when a term does have MFA, a similar proportion of its results are polluted.

Of course, ad prices and term popularity are correlated – more popular search terms tend to attract higher ad prices, and vice versa. Consequently, we use linear regression to disentangle the effect both have on the prevalence of abuse.

Because the dependent variable is binary in the case of malware (either the term has malware present or does not), we use a logit model for the regression of the following form:

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

We also ran a logit regression with the term's categories, but none of the category values were statistically significant. Thus, we have settled on this simpler model. The results of the regression reveal that a term's ad price and search popularity are both negatively correlated with the presence of malware in a term's search results, and the relationship is statistically significant:

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

These coefficients mean that a $1 increase in the ad price corresponds to a 40% decrease in the odds of having malware in the term's results. Likewise, when the popularity of a term doubles, the odds of having malware in the term's results decreases by 11%.

We also devised a linear regression using the fraction of a term's top 10 results classified as MFA as the dependent variable:

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

The Category variable is encoded as a 27-part categorical variable using deviation coding. Deviation coding is used to measure each categories' deviation from the overall mean value, rather than deviations across categories.

For this regression, the term's ad price and search popularity are both statistically significant and negatively correlated with the fraction of a trending term's top 10 results classified as MFA:

	coef.	Std. Err.	Significance
AdPrice	-0.0091	0.091	$p < 0.001$
$\log_2$(Popularity)	-0.004	0.012	$p < 0.001$
Coefficients for category variables in Tab. 4, R^2: 0.1373			

A $1 increase in the ad price corresponds to a 0.9% decrease in the MFA rate, while a doubling in the popularity of a search term matches a 0.4 percentage point decrease. This may not seem much, but recall that, on average, 6.6% of a term's top 10 results link to MFA sites. A 0.9% decrease in MFA prevalence represents a 13.2% decrease from the average rate.

Each of the coefficients listed in the right-most column in Table 4 are statistically significant (all have p values less than 0.001, except Local, Health, and Automotive, where $p < 0.05$). For instance, Food & Drink terms correspond to a 1 percentage point increase in the rate of MFA domains in their top 10 results, while Reference terms suffer a 2% higher MFA rate.

Implications of analysis. The results just presented demonstrate that, for both malware and MFA sites, miscreants are struggling to successfully target the more lucrative terms. An optimistic interpretation is that defenders manage to relegate the abuse to the more obscure terms that have less overall impact. A more pessimistic interpretation is that miscreants are having success in the tail of hot terms, which are more difficult to eradicate.

It is not very surprising that malware tends to be located in the results of terms that demand lower ad prices, given that higher ad prices do not benefit malware distribution. However, it is quite unexpected that the prevalence of MFA terms is negatively correlated with a term's ad price, since those promoting MFA sites would much prefer to appear in the search results of more expensive terms. One reason why malware and MFA appears less frequently on pages with higher ad prices could be that there is stronger legitimate competition in these results than for results fetching lower ad prices.

	# Visitors		
	Total	Period	Monthly Rate
MFA	39 274 200	275 days	4 284 458
Malware (trending set)			
detected	454 198	88 days	154 840
Bing, Yahoo!	189 511	88 days	64 606
undetected	143 662	88 days	48 975
Malware (control set)			
detected	12 825 332	88 days	4 372 272
Bing, Yahoo!	6 352 378	88 days	2 165 583
undetected	83615	88 days	28505

Table 5: Estimated number of visits to MFA and malware sites for trending terms.

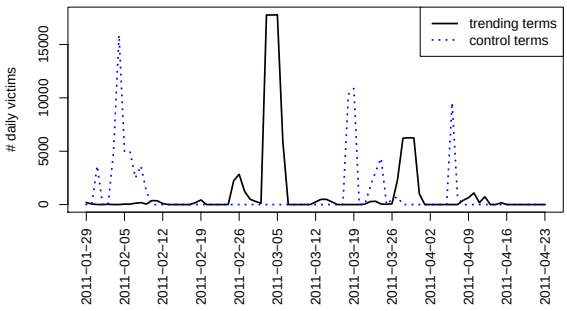


Figure 5: Number of estimated daily victims for malware appearing in trending and control terms.

Furthermore, there is a potential incentive conflict for search engines to eradicate ad-laden sites, when many of the pages run advertisements for the ad platforms maintained by the search engines. It is therefore encouraging that the evidence suggests that search engines do a better job at expelling MFA sites from the results of terms that attract higher ad prices.

Finally, the data helps to answer an important question: are malware and ad abuse websites competitors, or do they serve different parts of the market? The evidence suggests that, in terms of being a technique to monetize search traffic, malware and MFA behave more like substitutes, rather than complements. Both approaches thrive on the same types of terms, low-volume terms where ads are less attractive. Consequently, a purely profit-motivated attacker not fearful of arrest might choose between the two approaches, depending on which method generates more revenue.

4. ECONOMICS OF TRENDING-TERM EXPLOITATION

We next examine the revenues possible for both malware and ads, by first characterizing the volumes of population affected, before deriving actual expected revenues.

4.1 Exposed population

We first estimate the number of visits malware and MFA sites attract from trending-term searches. The cumulative number of visits over an interval t to a website w for a search term s is given by

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

where $\text{Pop}(s)$ is the monthly peak popularity of the term, as defined in Section 3.4. $\text{Rank}(w, s)$ is the position in search results website w occupies in response to a query for s, and $C(r)$ defines a click probability function for search rank $1 \leq r \leq 10$ following the empirical distribution observed by Joachims et al. [16]. They found that 43% of users clicked on the first result, 17% on the second result, and 98.9% of users only clicked on results in the first page. We ignore results in ranks above 10 (i.e., $C(r) = 0$ for $r > 10$).

$\text{Pop}(s)$ is measured at a monthly rate, so we normalize the visits to the four-hour interval between each search. We also weigh Google and Yahoo! search results differently. Google has reportedly an 64.4% market share in search, while Yahoo! and Bing have a combined market share of 30% [12]. Since our estimates are based on what Google observes, we anticipate that Yahoo! and Bing attract $\frac{30\%}{64.4\%} = 46.5\%$ of the searches that Google does.

The results are given in Table 5. MFA sites attract 39 million visits over nine months, or 4.3 million visits per month. For the malware results, we compare the estimated visits for both control and trending terms. While more users see malware in the results of control terms than trending terms (about 4.4 million versus about 200 000 per month over three months), over 99% of the visits from control terms are blocked by the Safe Browsing API. By contrast, 24% of the visits triggered from the results of trending terms are not blocked by the Safe Browsing API. In aggregate, trending terms expose around 49 000 victims per month to undetected malware, compared to about 28 000 for control terms.

The table also lists the number of Bing and Yahoo! users that encounter malware detected by Google's SafeBrowsing API. We cannot say for certain whether or not these users will be exposed to malware. If they attempt to visit the malicious site using the Chrome or Firefox browser then they would be protected, since Google's SafeBrowsing API is integrated into those browsers. Internet Explorer users would be protected only if the sites appear in IE's internal blacklist. Unfortunately, we could not verify this since the blacklist is not made publicly accessible.

The sums presented in the table mask several peculiarities of the data. First, for malware, the number of visitors exposed is highly variable. Figure 5 plots the number of daily victims over time. Most days the number of victims exposed is very small, often zero. Because terms in the control set are always very popular, successful attacks cause large spikes, but tend to be rare. On the other hand, trending terms exhibit frequent spikes, but many of the spikes are small. This is because many trending terms are in fact not very popular, even at their peak. A big spike, as happened around March 5, results from the conjunction of three factors: (1) the attacker must get their result towards the top of the search results; (2) the result cannot be immediately spotted and flagged; and (3) the trending term has to be popular enough to draw in many victims. Consequently, there is a downside to the constantly replenishing pool of trending terms for the attacker – they are often not popular enough for the attacker to do much damage. This is further exacerbated by the finding from the last section – more popular terms are less likely to be manipulated. At the same time, the figure demonstrates that even the odd success can reel in many victims.

Figure 6 plots the cumulative distribution of user visits compared to the affected domains. The graph indicates high concentration – most of the traffic is drawn to a small number of domains. The concentration of visitors is particularly extreme for malware, which makes sense given the spikes observed in Figure 5. The concentration in MFA sites shows that a few websites profit handsomely from trending terms, and that many more are less successful. This is consistent with our earlier finding that there are only a few large connected clusters of MFA sites linking to each other. One conse-

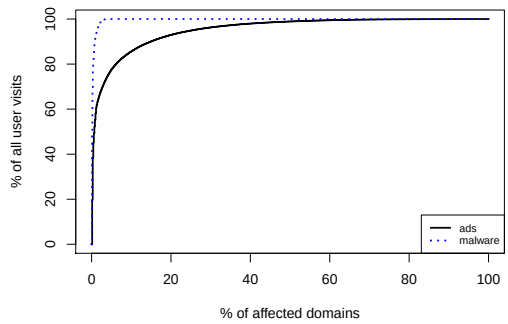


Figure 6: CDF of visits for domains used to transmit malware or ads in the search results of trending terms.

quence of this concentration is that we can approximate the revenue to the biggest players simply by considering aggregate figures.

4.2 Revenue analysis

We next compare revenues miscreants generate from MFA sites and from malware-hosting sites.

MFA revenue. Essentially, the aggregate revenue for MFA websites is a sum of the revenues generated by all MFA sites w obtained in response to all the search terms s considered. Each website generates a revenue equal to the number of website visitors times the advertising revenue that can be obtained from these visitors:

$$R_{\text{MFA}}(t) = \sum_{w \in \text{MFA}(s)} \sum_{s} V(w,s,t) \cdot (p_{\text{PPC}} \cdot p_{\text{clk}} \cdot r_{\text{PPC}} + p_{\text{banner}} \cdot r_{\text{banner}} + p_{\text{aff}} \cdot p'_{\text{clk}} \cdot r_{\text{aff}}) \; .$$

There are three broad classes of online advertising in use on MFA domains – pay-per-click (PPC) (e.g., Google AdSense), banners (e.g., Yahoo! Right Media) and affiliate marketing (e.g., Commission Junction). Banner advertisements are paid r_{banner} by the visit, PPC only pays r_{PPC} when the user clicks on an ad (which happens with probability p_{clk}), and affiliate marketing pays r_{aff} whenever a visitor clicks the ad and then buys something (which happens with probability p'_{clk}). By inspecting our corpus of MFA sites, we discover that 83% include PPC ads, 66% use banner ads, and 16% include affiliate ads. 50% of sites use two types of advertising, and 7% use all three. We include each type of advertisement in the revenue calculation with probability $p_{\text{ad type}}$, and we assign the probability according to the percentage of MFA site visits that include each class of ad. For the MFA websites we have identified, $p_{\text{PPC}} = 0.94$, $p_{\text{banner}} = 0.53$, and $p_{\text{aff}} = 0.33$.

To calculate the earning potential of each ad type, we piece together rough measures gathered from outside sources. Estimating the "click-through rate" (CTR) p_{clk} is difficult, as click-through rates vary greatly, and ad platforms such as Google keep very tight-lipped on average click-through rates. One Google employee reported that an average CTR is "in the neighborhood of 2%" [32]. We anticipate that the CTR for MFA sites is substantially higher than 2%, since sites have multiple ads aggressively displayed and little original content. Nonetheless, we assign $p_{\text{clk}} = 0.02$.

To measure per-click ad revenue r_{PPC}, we turn to the CPC estimates Google provides for advertising keywords. We expect that more persistent search terms are likely to appear as keywords for ads, even on websites about trending terms. Hence, we assume that advertising revenue for trending terms matches the CPC for most popular keywords in the corresponding category. We assign the expected advertising revenue to the mean of ad prices for the 20 most popular search terms weighted by the amount each category is represented in the results from the trending set (see Table 4, column 1). This yields $r_{\text{PPC}} = \$0.97$.

Calculating banner advertising revenue is a bit easier, since no clicks are required to earn money. Public estimates of average revenue are hard to come by, but the ad network Adify issued a press release stating that its median cost per 1 000 impressions in Q2 2010 was \$5.29 [4], so we assign $r_{\text{banner}} = \$0.00529$.

For affiliate marketing, we assume assume that $p'_{\text{clk}} = p_{\text{clk}} = 0.02$, the same as for PPC ads. To estimate the revenue r_{aff} that can be earned, we turn to Commission Junction (CJ), one of the largest affiliate marketing networks that matches over 2 500 advertisers with affiliates. CJ provides an estimate of expected earnings from advertisers per 100 clicks; we collected this estimate for all advertisers on Commission Junction in December 2009, and found it to be \$26.49. Consequently, we estimate that $r_{\text{aff}} = \$0.265$.

Putting it all together, we estimate the monthly revenue to MFA sites to be:

$$\begin{aligned} R_{\text{MFA}}(1 \text{ month}) =& 4\,284\,458 \times (0.94 \times 0.02 \times \$0.97 \\ &+ 0.53 \times \$0.00529 + 0.33 \times 0.02 \times \$0.265) \\ =& \$97\,637 \; . \end{aligned}$$

So, MFA sites gross roughly \$100,000 per month from trending-term exploitation. There are, however, costs that are not factored into the above derivation, which makes it an upper bound. For instance, Google generally imposes a 32% fee on advertising revenues [24]. Furthermore, servers have to be hosted and maintained. As an example, most sites in the largest cluster in Section 3.2 are hosted by the same service provider, which charges \$140/server/month. That cluster contains 193 nodes hosted on 155 unique servers, which, ignoring economies of scale, would come up to \$21 700/month in maintenance. Nevertheless, it is worth noting that these costs can be amortized over other businesses – it is unlikely that such servers are only set up for the purpose of trending-term exploitation.

Malware revenue. Attackers have experimented with several different business models to monetize drive-by-downloads, from adware to credential-stealing trojans [31]. However, researchers have observed that attackers exploiting trending terms have tended to rely on fake antivirus software [2, 8, 36]. We therefore define the revenue due to malware in trending results as:

$$R_{\text{mal}}(t) = \sum_{\text{w} \in \text{mal(s)}} \sum_{s} V(w,s,t) \cdot p_{\text{exp}} \cdot p_{\text{pay}} * r_{\text{AV}} \; ,$$

where we multiply the number of visits times the likelihood of exposure, the probability of a victim paying for the software, and the amount paid. For these figures, we turn to the analysis of Stone-Gross et al. [36], who acquired a copy of back-end databases detailing the revenues and expenses of three large fake antivirus programs, each of which were advertised by compromising trending search results. They found that 2.16% of all users exposed to fake antivirus ultimately paid for a "license," at an average cost of \$58. We can use these figures directly in our model for the revenues due to malware, setting $p_{\text{pay}} = 0.0216$ and $r_{\text{AV}} = \$58$.

Unlike most drive-by-downloads, fake antivirus software does not need to exploit a vulnerability in the client visiting the infected search result in order for a user to be exposed. Instead, the server will use a server-side warning designed to appear as though it is on the client's machine, and then prompt a user to install software [2]. Because of this, every user that visits a link distributing fake AV is exposed, and so we assign $p_{\text{exp}} = 1$. These parameters yield a

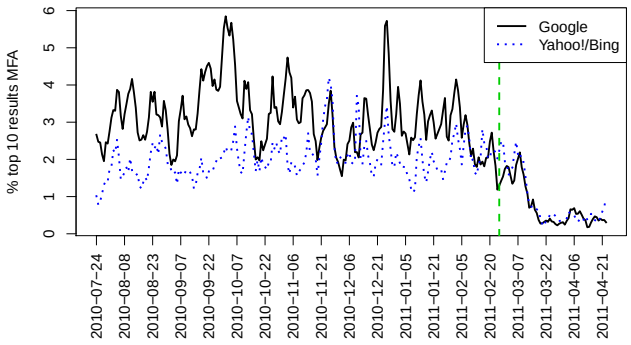


Figure 7: MFA prevalence in the top 10 search results fell after Google announced changes to its ranking algorithm on February 24, 2011, designed to counter "low-quality" results.

monthly revenue from malware of:

$$R_{\text{mal}}(1 \text{ month}) = 48\,975 \times 1 \times 0.0216 \times 58 \approx \$61\,356\ .$$

Thus, malware sites (e.g., fake anti-virus sites) generate roughly $60,000/month just from trending-term exploitation.

Here too, there are costs associated with deploying these sites, but server maintenance is a lot cheaper than in the case of MFA sites, given that most machines hosting malware have been compromised rather than purchased. Bots go for less than a dollar [5, and references therein], while a compromised server (presumably with high quality network access) goes at most for $25 according to Franklin et al. [9]. Note that we do not adjust the returns on malware for the risk of being caught because the likelihood of being arrested for cyber-criminal activity is currently negligible in many jurisdictions where cyber-criminals operate.

One conclusion of this analysis is that malware and MFA hosting have quite different revenue models, but yield surprisingly similar amounts of money to their perpetrators. This lends further support to the hypothesis that they could be treated as substitutes.

5. SEARCH-ENGINE INTERVENTION

On February 24, 2011, following a series of high-profile reports of manipulation of its search engine (e.g., [33, 34]), Google announced changes to its search ranking algorithm designed to eradicate "low-quality" results [35]. Google defined low-quality sites as those which are "low-value add for users, copy content from other websites or sites that are just not very useful." The MFA sites examined in this paper certainly appear to match that definition. Because we were already collecting search results on the trending set, we can measure the effectiveness of the intervention in eradicating abuse targeting trending terms.

Figure 7 plots over time the average percentage of top 10 search results marked as MFA for terms in the trending set. From July to February, 3.1% of Google's top 10 results (solid line) for trending terms pointed to MFA sites, compared to 2.0% for Yahoo!'s top 10 results (dotted line). The vertical dashed line marks February 24, 2011, the day of Google's announcement. The proportion of MFA sites quickly fell, stabilizing a month later at a rate of 0.47% for Google. Curiously, Yahoo!'s share of top 10 MFA results also fell, to an average of 0.56%.

Landing in the top results tells only part of the story. The underlying popularity of the trending terms is also important. We compute the estimated site visits to MFA sites, which is more directly tied to revenue. Table 6 compares the number of visits referred to by Google and Yahoo! search results before and after the intervention. Between July 24, 2010 and February 24, 2011, MFA sites attracted 4.67 million monthly visits on average. Between March 10 and April 24, 2011, the monthly rate fell 31% to 3.2 million.

	Monthly MFA visits		
	Pre-intervention	Post-intervention	% change
Google search	3 364 402	1 788 480	-47%
Google ads	2 989 821	1 763 709	-41%
Other ads	374 556	24 770	-93%
Yahoo!/Bing search	1 302 314	1 448 058	+11%
Google ads	1 204 928	1 424 323	+18%
Other ads	95 363	23 734	-75%
Total	**4 666 716**	**3 236 538**	**-31%**

Table 6: Estimated number of visits to MFA and malware sites for trending terms.

However, the changes differed greatly across search engines. Referrals from Google search results fell by 47%, while on Yahoo! and Bing the visits increased by 11%. The table also distinguishes between whether the MFA site uses Google ads or another provider. 81% of MFA sites show Google ads, which is not surprising given Google's dominance in pay-per-click advertising. It is an open question whether Google might treat MFA sites hosting its own ads differently than sites with other ads. Striking them from the search results reduces Google's own advertising revenue. However, it is in Google's interest to provide high-quality search results, the amount of foregone revenue is small, and is likely to be partly replaced by other search results. Our figures support the latter rationale. Sites with Google ads fell by 1.2 million visits, or 41%. Visits to sites not using AdSense fell by 91%, but, in absolute terms, the reduction was smaller than for sites with Google ads. By contrast, Yahoo! results with Google ads rose by 18%.

Using the pre- and post-intervention MFA visit rates into the revenue equations developed in Section 4.2, the average monthly take for MFA sites has fallen from $106 000 to $74 000. If this reduction holds over time, what are the implications for miscreants? First, they may decide to devote more effort to manipulating Yahoo! and Bing, despite their lower market penetration, since the MFA revenues are growing more equitable in absolute terms. Second, malware becomes more attractive as an alternative source of revenue, so one unintended consequence of the intervention to improve search quality could be to foster more overtly criminal activities harming consumers. Third, revenue models based on advertising require volume, and external efforts that reduce traffic levels can cause significant pain to the miscreant. By contrast, malware offers substantially more expected revenue per visitor, and is therefore likely to be much more difficult to eradicate.

Given the striking change in MFA prevalence following Google's intervention, it is worth checking whether this intervention alters the significance of the empirical conclusions reached in Section 3.4. We included a dummy variable into the MFA regression reflecting whether Google's intervention had yet occurred, and found that this inclusion does not alter the significance of the dependent variables presented in Section 3.4.

6. RELATED WORK

Our work inscribes itself in the body of literature on understanding the underground online economy. Some of the early econometric work in that domain revolves around quantities bartered in underground forums [9], and on email spam campaigns [18, 22]. Grier et al. [10] extend this literature to Twitter spam. Along the

same lines, Moore and Clayton have published a series of papers characterizing phishing campaigns [25, 26, 27].

More recently, a number of papers have also started to investigate web-based scams. Christin et al. [6] study a specific web-based social engineering scam ("one click fraud"). Provos et al. describe in details how so called "drive-by-downloads" are used to automatically install malware [30, 31]. Cova et al. [8] and Stone-Gross et al. [36] focus on fake anti-virus malware, and provide estimates of the amount of money they generate. Stone-Gross et al. calculate, through recovery of the miscreants' transactions logs, that anti-virus campaigns gross between $3.8 and $48.4 million a year. Affiliates funneling traffic to miscreants get between $50,000 and $1.8 million in over two months. These totals are markedly higher than what we obtain, but they consider all possible sources of malware (botnets, search engine manipulation, drive-by-downloads) whereas we only look at the much smaller subset of search engine manipulation based on trending term exploitation.

A few recent works consider search engine manipulation. Leontiadis et al. [20] investigate search engine manipulation to promote potentially illicit online drugs. John et al. [17] present a case study of recent search engine manipulation campaigns, confirming that trending-term exploitation is an attack vector of choice. They then devise countermeasures to thwart search engine manipulation.

Our approach differs from the related work in that we focus on a specific phenomenon – trending-term exploitation – by investigating how it is carried out (e.g., search-engine manipulation, Twitter spam), as well as its purpose: malware distribution and monetization through advertisements. Our analysis thus sheds light on a specific technique used by miscreants that search-engine operators are battling to fend off.

7. CONCLUSION

We have undertaken a large-scale investigation into the abuse of "trending" terms, focusing on the two primary methods of monetization: malware and ads. We have found that the the dynamic nature of the trends creates a narrow opportunity that is being effectively exploited on web search engines and social-media platforms. We have presented statistical evidence that the less popular and less financially lucrative terms are exploited most effectively, and that the spoils of abuse are highly concentrated among a few players. We have developed an empirically grounded model of the earnings potential of both malware and ads, finding that each attracts aggregate revenues on the order of $100 000 per month. Finally, we have found that Google's intervention to combat low-quality sites has likely reduced revenues from trend exploitation by more than 30%.

There is a connection in our economic modeling to the battle over how to profit from typosquatting [28]. In both cases, Internet "bottom feeders" seek to siphon off a fraction of legitimate traffic at large scale. Several years ago, typosquatting was used in phishing attacks and to distribute malware. Today, however, typosquatting is almost exclusively monetized through pay-per-click and affiliate marketing ads [28], attracting hundreds of millions of dollars in advertising revenue to domain squatters via ad platforms.

The open question is whether a significant crackdown on, say, fake antivirus sales, will simply shift the economics in favor of low-quality advertising. However, while ad platforms might tolerate placing ads on typosquatted websites, advertising that lowers the quality of search results directly threatens the ad platform's core business of web search. Consequently, we are more optimistic that search engines might be willing to crack down on all abuses of trending terms, as we have found in our initial data analysis.

8. ACKNOWLEDGMENTS

We thank our anonymous reviewers for helpful feedback on this paper. This research was partially supported by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from the Army Research Office, and by the National Science Foundation under ITR award CCF-0424422 (TRUST).

9. REFERENCES

[1] Google Web Search API. https://code.google.com/apis/websearch/.

[2] M. Abu Rajab, L. Ballard, P. Mavrommatis, N. Provos, and X. Zhao. The nocebo effect on the web: an analysis of fake anti-virus distribution. In *Proc. USENIX LEET'10*, San Jose, CA, April 2010.

[3] AdBlock. Adblock easy list. https://easylist-downloads.adblockplus.org/easylist.txt.

[4] Adify. Adify vertical gauge shows steady growth in seven of eleven critical verticals. http://www.smartbrief.com/news/aaaa/industryMW-detail.jsp?id=732F69A7-9192-4E05-A261-52C068021634. Last accessed May 5, 2011.

[5] N. Christin, S. Egelman, T. Vidas, and J. Grossklags. It's all about the Benjamins: Incentivizing users to ignore security advice. In *Proc. Financial Crypto.'11*, St. Lucia, Feb. 2011.

[6] N. Christin, S. Yanagihara, and K. Kamataki. Dissecting one click frauds. In *Proc. ACM CCS'10*, pages 15–26, Chicago, IL, Oct. 2010.

[7] G.F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4):309–347, 1992.

[8] M. Cova, C. Leita, O. Thonnard, A. Keromytis, and M. Dacier. An analysis of rogue AV campaigns. In *Proc. RAID 2010*, Ottawa, ON, Canada, September 2010.

[9] J. Franklin, V. Paxson, A. Perrig, and S. Savage. An inquiry into the nature and causes of the wealth of internet miscreants. In *Proc. ACM CCS'07*, pages 375–388, Alexandria, VA, October 2007.

[10] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: The underground in 140 characters or less. In *Proc. ACM CCS'10*, pages 27–37, Chicago, IL, October 2010.

[11] Z. Gyöngyi and H. Garcia-Mollina. Link spam alliances. In *Proc. VLDB'05*, pages 517–528, Trondheim, Norway, Aug. 2005.

[12] Experian Hitwise. Experian hitwise reports bing-powered share of searches reaches 30 percent in march 2011, April 2011. http://www.hitwise.com/us/press-center/press-releases/experian-hitwise-reports-bing-powered-share-of-s/.

[13] Google Inc. Google insights for search. http://www.google.com/insights/search/.

[14] Google Inc. Google traffic estimator. https://adwords.google.com/select/TrafficEstimatorSandbox.

[15] Google Inc. Google trends. http://www.google.com/trends/.

[16] T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proc. ACM SIGIR'05*, pages 154–161, Salvador, Brazil, Aug. 2005.

[17] J. John, F. Yu, Y. Xie, M. Abadi, and A. Krishnamurthy. deSEO: Combating search-result poisoning. In *Proc. USENIX Security'11*, San Francisco, CA, August 2011.

[18] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, and S. Savage. Spamalytics: An empirical analysis of spam marketing conversion. In *Proc. ACM CCS'08*, pages 3–14, Alexandria, VA, Oct. 2008.

[19] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proc. IJCAI'95*, pages 1137–1145, Montreal, QC, Canada, Aug. 1995.

[20] N. Leontiadis, T. Moore, and N. Christin. Measuring and analyzing search-redirection attacks in the illicit online prescription drug trade. In *Proc. USENIX Security'11*, San Francisco, CA, August 2011.

[21] J. Leskovec, L. Backstrom, and R. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proc. ACM KDD'09*, Paris, France, June 2009.

[22] K. Levchenko, N. Chachra, B. Enright, M. Felegyhazi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, D. McCoy, A. Pitsillidis, N. Weaver, V. Paxson, G. Voelker, and S. Savage. Click trajectories: End-to-end analysis of the spam value chain. In *Proc. IEEE Symp. Security & Privacy*, pages 431–446, Oakland, CA, May 2011.

[23] Microsoft. Microsoft, yahoo! change search landscape. `http://www.microsoft.com/presspass/press/2009/jul09/07-29release.mspx`.

[24] N. Mohan. The AdSense revenue share, May 2010. `http://adsense.blogspot.com/2010/05/adsense-revenue-share.html`.

[25] T. Moore and R. Clayton. Examining the impact of website take-down on phishing. In *Proc. APWG eCrime'07*, Pittsburgh, PA, October 2007.

[26] T. Moore and R. Clayton. Evil searching: Compromise and recompromise of internet hosts for phishing. In *Proc. Financial Crypto'09*, pages 256–272, Barbados, Feb. 2009.

[27] T. Moore, R. Clayton, and H. Stern. Temporal correlations between spam and phishing websites. In *Proc. USENIX LEET'09*, Boston, MA, April 2009.

[28] T. Moore and B. Edelman. Measuring the perpetrators and funders of typosquatting. In *Proc. Financial Crypto.'10*, pages 175–191, Tenerife, Spain, Jan. 2010.

[29] J. Pearl. Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning. In *Proc. 7th Conf. of the Cognitive Science Society*, pages 329–334, Irvine, CA, Aug. 1985.

[30] N. Provos, P. Mavrommatis, M. Abu Rajab, and F. Monrose. All your iFrames point to us. In *Proc. USENIX Security'08*, San Jose, CA, August 2008.

[31] N. Provos, D. McNamee, P. Mavrommatis, K. Wang, and N. Modadugu. The ghost in the browser: Analysis of web-based malware. In *Proc. USENIX HotBots'07*, Cambridge, MA, April 2007.

[32] B. Schwarz. Google adwords click through rates: 2% is average but double digits is great, January 2010. `http://www.seroundtable.com/archives/021514.html`. Last accessed May 3, 2011.

[33] D. Segal. A bully finds a pulpit on the web. *New York Times*, November 2010. Article appeared in print on November 28, 2010, on page BU1 of the New York edition. Available online at `http://www.nytimes.com/2010/11/28/business/28borker.html`.

[34] D. Segal. The dirty little secrets of search. *New York Times*, February 2011. Article appeared in print on February 13, 2011, on page BU1 of the New York edition. Available online at `http://www.nytimes.com/2011/02/13/business/13search.html`.

[35] A. Singha. Finding more high-quality sites in search, February 2011. `http://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html`.

[36] B. Stone-Gross, R. Abman, R. Kemmerer, C. Kruegel, D. Steigerwald, and G. Vigna. The underground economy of fake antivirus software. In *Proc. (online) WEIS 2011*, Fairfax, VA, June 2011.

[37] Twitter. Twitter developers trends resources. `http://dev.twitter.com/doc/get/trends/`.

[38] Yahoo! Inc. Yahoo buzzlog. `http://buzzlog.yahoo.com/overall/`.

[39] Yahoo! Inc. Yahoo site explorer. `http://siteexplorer.search.yahoo.com/`.