Temporal Correlations between Spam and Phishing Websites

Tyler Moore
Center for Research on Computation and Society, Harvard University
tmoore@seas.harvard.edu

Richard Clayton
Computer Laboratory, University of Cambridge
rnc1@cl.cam.ac.uk

Henry Stern
Cisco IronPort Systems LLC
hstern@ironport.com

Abstract

To implement a phishing scam, attackers must create a fake website and send spam to attract visitors. To date, empirical research into phishing's impact has studied either the spam being sent or the website lifetimes. In this paper, we examine both phishing websites and the associated spam to gauge the overall effectiveness of phishing attack and defense. We find that while the bulk of spam is sent around the time of the website's first appearance, spam continues to be sent for many longer lived websites until they are finally removed. We also find that attackers using 'fast-flux' techniques are savvier than ordinary attackers, sending out more spam prior to detection and stopping faster once the websites are taken down. Finally, we conclude that fast-flux attacks pose the greatest phishing threat since they account for 68% of spam despite comprising just 3% of hosts.

1 Introduction

Phishing is the criminal activity of enticing people into visiting websites that impersonate the real thing, to dupe them into revealing passwords and other credentials, which will later be used for fraudulent activities. Although a wide range of companies are attacked, the targets are usually financial institutions; hence for simplicity, we will describe them as 'banks'.

One of the key countermeasures to phishing is the prompt removal of the imitation bank websites. This may be achieved by removing the web pages from the hosting machine, or in complex 'fast-flux' schemes where page requests are relayed by ever-changing armies of proxy machines, a registrar must suspend a domain name from the DNS so it can no longer be resolved.

In prior work [5] we measured phishing website lifetimes and showed how particular groups of attackers use sophisticated strategies to extend the lifetime of their websites. We used publicly visible logging data to estimate how many people were being defrauded, calculating a figure that is very similar to that obtained by Florêncio and Herley, who measured the incidence of submission of the same password to different websites [1]. In a later paper [6] we demonstrated that the lack of data sharing by the take-down industry caused significantly extended website lifetimes; banks do not remove websites of which they are unaware.

However, the significance of these long-lived websites has always been unclear, because people do not find them by random chance and fill in their details. Rather, phishing websites are advertised using spam email, purportedly sent by the bank. The email will demand that action be taken to prevent account suspension, or will insist upon new personal details, or will offer a cash prize for completing a marketing questionnaire.

It is generally assumed that attackers do not send phishing emails long before their website is ready, or continue to advertise a site long after it is taken down. But it is not known if attackers move on to new websites when the old one remains active, perhaps to evade spam filtering mechanisms, or whether they continue to advertise the same site until it is eventually removed.

This paper's key contribution is a first-ever analysis of the temporal relationship between the sending of spam and the lifetime of phishing websites. In particular, we find evidence for the following:

- website take-down is necessary: spam continues to be sent for up to 75% of the phishing websites that are alive after one week;
- fast-flux attacks comprise only 3% of distinct attacks, but 68% of spam volume, suggesting they are a far more serious threat than the more numerous websites hosted on compromised web servers;
- fast-flux attackers manage spam campaigns more efficiently, sending out most spam before a website is discovered and stopping shortly after its removal.

2 Data collection methodology

2.1 The phishing spam dataset

For our spam data we use IronPort's email corpus of messages received directly by IronPort's SpamCop [10] spam traps, a selection of third-party spam traps and customer submissions. The biggest contribution is made by the large number of geographically diverse SpamCop spam traps. We decode, de-obfuscate and extract the URLs from the messages using IronPort Anti-Spam [3].

We consider the time that IronPort's corpus receives an email message to be the time at which the message was sent. This is accurate for all messages received by the SpamCop spam traps, and delayed by a few seconds for third-party spam traps. There can be delays of several hours for customer submissions, but these are only a small fraction of the overall dataset.

2.2 Phishing URL feeds

We receive a number of disparate 'feeds' of phishing website URLs. We fetch data from PhishTank,[1] the community site. We receive a feed from a major brand owner, who is subject to very widespread impersonation, and a feed collated from numerous sources by the Anti-Phishing Working Group (APWG).[2] We also receive feeds from two 'brand protection' companies who offer specialist phishing website take-down services. These companies amalgamate feeds from numerous other sources, and add data from proprietary phishing email monitoring systems.

Although these feeds overlap substantially, in practice each contains a number of URLs that we do not receive from any other source. We believe that our database of URLs is one of the most comprehensive available, and that the overwhelming majority of phishing websites will come to our attention.

2.3 Types of phishing website

There are two very different ways of creating phishing websites. In most cases fraudulent HTML is loaded by the attacker onto a compromised machine. In past work we have shown that this accounts for around 75% of all phishing websites [8]. A similar, though less popular approach (about 20% of cases), is to load the phishing web page onto a 'free' web host where anyone can register and upload pages.

A different method is the creation of a so-called 'fast-flux' systems, where the attackers use rapidly changing (hence the 'fast-flux' term) pools of malware infected machines as proxies to hide the location of their web servers [5, 2]. The attackers generally use lengthy URLs containing random sections, and mount simultaneous attacks against multiple targets using the same domain names. For this work we ignore these specious variations, and the multiple target banks, and just consider the canonical domain names that were used.

2.4 Measuring phishing website lifetimes

In all cases except PhishTank and APWG, the URLs are passed to us after they have been determined to be fraudulent sites. We deem the sites to have been 'up' at the time at which this determination was made or, if we are told when the site was initially reported, we use that (earlier) time instead. In practice, the take-down companies process URLs almost as fast as they are received.

In the case of PhishTank, we use the time of the first appearance on the PhishTank website as the earliest time that the site is known to have existed, and so we are unaffected by the speed at which volunteers vote to determine its status. For the APWG feed, we use the time that the site appears in the feed.

We run a monitoring system which checks the websites reported to us for any changes to their content. Each phishing web page is accessed some 15 to 20 times per day to determine what content is being served. A website that returns a '404' error is removed from testing, but other failures are retried for several days to ensure that minor outages are not mistaken for permanent removal.

We deem a phishing website to be 'up' while the content is unchanged, but any significant change (beyond session identifiers, cookies, etc.) is treated as a removal. In practice, fraudulent sites are created from 'kits' and do not change after installation. The only manual step we apply is to ensure that the site has not been taken down before we first monitor it. We measure the website lifetime from the earliest time we knew it was 'up' until the time of the last access to the fraudulent content.

We identify fast-flux domains because they continually resolve to new IP addresses. For these domains we treat the 'up time' as the period during which the domain can be resolved. In practice, almost all of the IP addresses are functional proxies, and so the domain lifetime exactly corresponds to the period during which victims can be fooled into disclosing their banking credentials.

2.5 The phishing website dataset

For this paper, we selected all of the URLs for phishing websites that we received during the last week of September 2008 (Wed 24 – Tue 30), some 12 693 in total. We constructed 4 475 regular expressions, which allowed

[1] http://www.phishtank.com
[2] http://www.apwg.org/

	Phishing feeds		Spam feed	
	Total	Visited	Total	Visited
Ordinary	4084	3227	430	396
Fast-flux	120	113	103	100

Table 1: Comparing the coverage of phishing URL feeds and the occurrence of phishing URLs in email spam.

	Website lifetime				Spam duration	
	mean		median		mean	median
Ordinary	52	(134)	18	(30)	106	0
Fast-flux	97	(97)	21	(22)	97	28

Table 2: Comparing lifetimes (in hours) of phishing websites to the duration of phishing spam campaigns. For website lifetime, first value is computed using website data only; second value reflects earlier spam data.

us to group together duplicate reports and treat all reports of the same fast-flux domain as a single incident. We then processed our entire phishing URL database (which goes back several years) and excluded long-running websites, viz: any URL or fast-flux domain that was first reported prior to 24 Sep 2008. This allowed us to determine which phishing websites and fast-flux domains were 'new' during the seven day period.

The same set of regular expressions was used to determine which of the phishing website URLs and fast-flux domains were present in the spam dataset covering the period 1 Jun 2008 – 31 Dec 2008. We summarize the totals in Table 1.

Of the 4 084 ordinary phishing websites in our feed for the week, our monitoring system successfully visited 3 227 before the sites were taken down. Around 10% of these websites (430 of 4 084) were also identified in the spam feed. This leaves us with 396 ordinary phishing websites, that we successfully visited prior to take-down, which also appeared in Ironport's spam feed.

Note that there are far fewer unique fast-flux domains in use (120), and furthermore that a much large proportion of these fast-flux domains can be found in the spam feed (103). Why might this be? As we will show in Section 4, despite fast-flux websites being a diminutive part of the total, the spam that advertises them comprises a very large proportion of the overall spam volume. IronPort's spam archive, although extensive, is more likely to exclude smaller spam campaigns. Because spam for fast-flux phishing hosts is sent in such high volume it is far more likely to be retained by IronPort.

3 Spam campaign duration

In Section 2.4, we explained how we compute phishing website lifetimes by recording the first and last time a phishing website is viewable. We now consider relationships with the times that phishing spam is sent.

We start by associating the individual spam email records with their respective phishing URLs to establish what 'campaigns' we have detected for a particular website or a particular fast-flux domain. We can measure the spam campaigns' impact in two ways:

- **Spam campaign duration** the time difference between the first and last spam email sent advertising the same phishing URL – sometimes zero if we only saw one such email;
- **Spam campaign volume** the weighted estimate of the number of spam messages advertising a URL or domain (over time and in summation).

Table 2 compares the duration of spam campaigns and website lifetimes for both fast-flux and ordinary attacks, the latter being those that are hosted on compromised web servers or free web space.

Banks and take-down companies can only measure the average lifetime of phishing websites from the time they first learn of the site until the point at which they get it removed. Our measurements improve on this because we receive feeds from multiple sources and so we can sometimes establish an earlier start time. If we now assume, as seems plausible, that the website is alive at the point that the very first email spam is detected then we can refine our lifetime estimate still further. Ignoring the email data the average lifetime of ordinary phishing websites is 52 hours; but by accounting for the timing of spam transmission, the average more than doubles to 134 hours.

For both attack types, the adjusted average website lifetime is as long or longer than the spam campaign duration (e.g., 134 hours for ordinary websites against 106 hours for the associated spam). However, in every case the mean times are much longer than the median times. We observed this effect for phishing websites in prior work [5], finding that the distribution is highly skewed. If median values are compared, the lifetime measurements for both spam and websites become much closer. However this is a little misleading for our dataset because the median lifetime for spam campaigns is actually 0 hours, because in over half of all campaigns for ordinary websites, the URL is only observed in spam on a single occasion.

The spam durations we have identified are comparatively shorter than campaigns studied by others. For instance, Kanich et al. tracked a 26-day-long spam campaign advertising online pharmacies [4]. However,

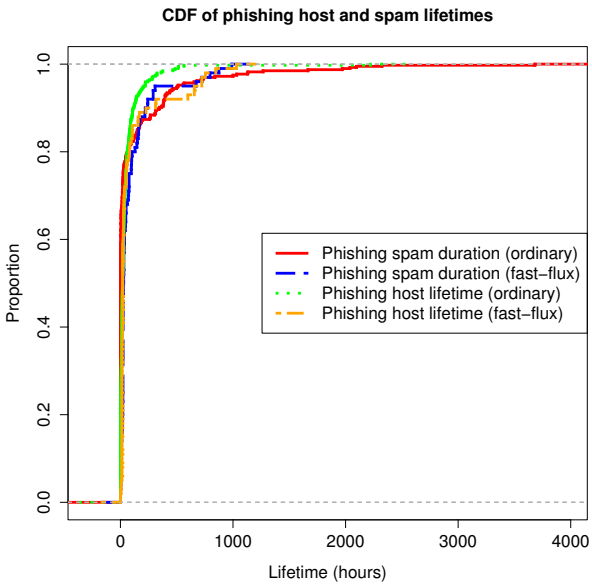

Figure 1: Cumulative distribution function of 'lifetimes' for phishing websites and associated spam.

the lifetime of online pharmacies is much longer than that of the average phishing website (weeks rather than hours) [7]; consequently, we might expect a corresponding difference in spam campaign length.

Given the mean-median disparity, it is instructive to examine the cumulative distribution functions (CDFs) for spam campaigns relative to website lifetimes. Figure 1 plots the CDF for spam campaign duration over time, along with website lifetime as a reference. The red and blue lines indicate spam campaign duration, revealing that the most spam campaigns do not last very long, save for a small but substantial fraction of campaigns that last far longer. For instance, 75% of spam campaigns for ordinary phishing websites last one day or less, while 13% of campaigns continue for more than ten days. Meanwhile, the distribution of spam campaign durations for fast-flux attacks is subtly different. Fast-flux campaigns initially last longer – just 42% are finished within one day. But they also die out faster – no fast-flux campaign lasts more than 41 days, while a few ordinary spam campaigns can last three months or longer.

The other interesting lesson to be learned from Figure 1 is the relationship between phishing website lifetimes (the green and orange lines) and spam duration. The distributions for website lifetime and spam campaign durations are quite similar which suggests that the duration of spam campaigns and website lifetimes may be linked.

Comparing the CDFs of phishing spam and websites gives a feel for the relationship between spam campaigns and the hosting websites. Phishing websites tend to last about as long as spam runs; fast-flux campaigns initially last longer than campaigns for ordinary phishing attacks, but they peter out faster. Yet, although this is a promising start, it does not get us to the heart of the matter. We need a more direct comparison of the temporal relationship between spam and websites. Figure 2 presents another step forward.

The left-hand graph presents a CDF of the time difference between when the phishing website is first detected and when the associated spam is first observed. Positive numbers indicate that the website appeared before spam was detected, while negative numbers indicate that the spam was detected before the website appeared.

For ordinary phishing attacks, there is great variation between when the first spam arrives and when a website appears. 21% of the time, the spam arrives more than a day before the phishing website is detected; 33% of the time, the spam arrives more than a day *after* the phishing website is detected. The remaining half of the time, spam arrives within a day of the website's detection. This means that, for around a fifth of ordinary phishing websites, there is scope for a marked improvement in the speed of detection.

For fast-flux attacks, the difference couldn't be more stark. *All* fast-flux spam is detected within a few hours before or after the phishing website is identified. In fact, the biggest outliers are one domain appearing 4.6 hours after spam is found and one domain appearing 5.4 hours before the spam is identified.

What is the likely explanation for the huge difference between ordinary and fast-flux detection? As we will show in the next section, fast-flux attacks generate far more spam than ordinary phishing attacks. Given that the take-down companies rely in part on spam traps to build their feeds of phishing websites, it is no surprise that higher-spammed fast-flux websites are highly correlated with the timing of spam transmission.

The right-hand graph in Figure 2 shows the time difference between when the last spam is sent advertising a website and when that website is removed. Negative values occur when the final spam was prior to website removal, while positive values are positive when the website was removed and spam was still sent thereafter.

For ordinary phishing websites, there is a substantial variation between when the last spam is sent and when the website is finally removed. 29% of spam campaigns send their final message more than one day *before* the website is ultimately removed, while 35% send their final message more than one day *after* the website has been removed. The remaining 37% of websites are removed within a day of the final associated spam transmission.

Fast-flux spam transmission is not so tightly correlated

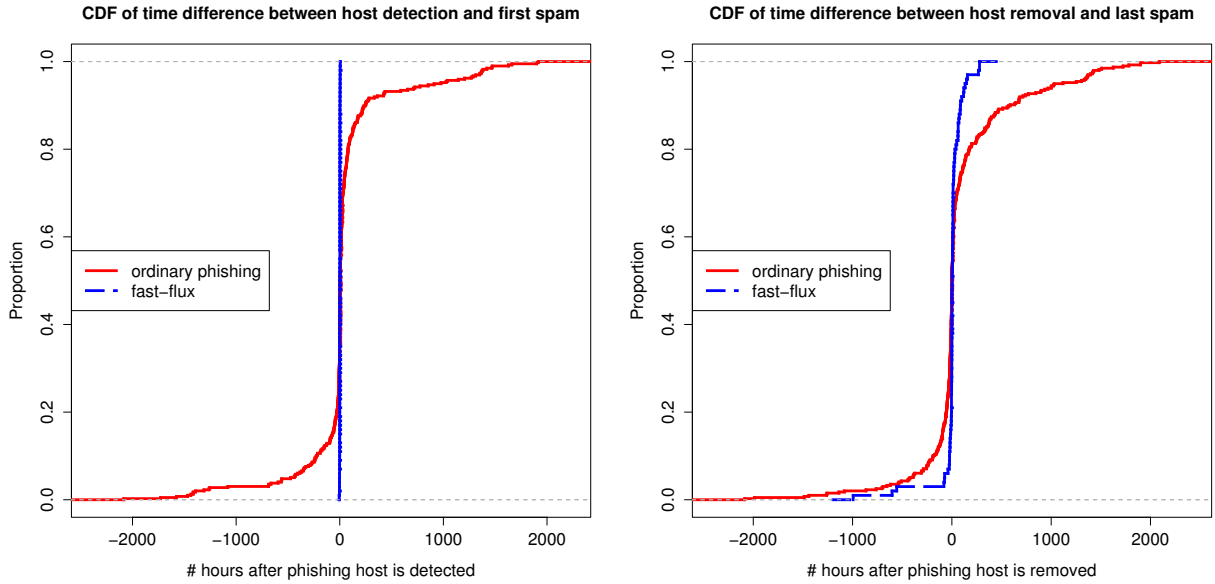

Figure 2: Cumulative distribution function of the difference in time between the first appearance of the phishing host and associated spam (left). Positive numbers indicate that the website appeared before spam was detected, negative numbers indicate that the spam was detected before the website appeared. The right-hand graph gives a similar measure for the difference between the last spam sent and the website's removal.

with website removal as was the case with phishing website creation. 8% of spam campaigns send their final message more than one day before the web domain is ultimately removed, while 23% send their final message more than one day after the fast-flux domain has been removed. However, more than two thirds of fast-flux spam campaigns end within one day of the associated domain's termination.

What lessons can we learn from the right-hand graph? Fast-flux website removal is more more highly correlated with the termination of spam transmission than ordinary phishing websites. Attackers using fast-flux have much better control over the transmission of advertising spam than do attackers using ordinary tactics. Spam is more likely to continue until the website has been removed, and to stop once termination happens.

However, the correlation is not absolute for either type of phishing attack. A substantial portion of phishing websites continue to operate long after the spam has stopped, while another portion of attackers continue to send spam long after the website has been shut down. It is somewhat reassuring to know, given the frequent mistakes and oversights of defenders (as discussed in [6]), that the attackers are not perfect either.

4 Phishing spam volume over the lifetime of its associated website

In the previous section, we examined the relationship between the *timing* of phishing spam transmission and website detection and removal. We now consider how the *volume* of spam transmission compares to a phishing website's lifetime.

The record of spam kept by IronPort does not match to the sending of a single spam message. Rather, the existence of a record at a particular time means that certain threshold of spam messages has been detected. To estimate volume, we might simply add up the number of entries in the IronPort spam archive each phishing website received. However, we can actually do better. IronPort adds a weight to each entry that captures the spam's prevalence. By adding up the weighted entries, we arrive at a measure of the volume of spam transmitted. This volume roughly corresponds to the number of spam messages observed.

Table 3 compares the volume of spam produced for each type of phishing attack. Each fast-flux attack generates nearly nine times as much spam as an ordinary phishing attack on average, and 400 times as much if we compare median values! While the disparity in volume is shocking, it is not too surprising, given that fast-flux attacks often target multiple banks simultaneously and it

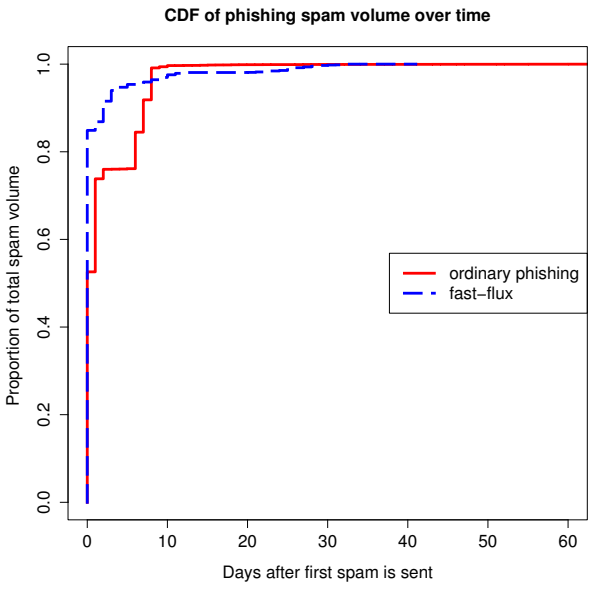

Figure 3: Cumulative distribution function of phishing spam volume over time.

	Spam volume (# messages)	
	mean	median
Ordinary	2276	34
Fast-flux	20 194	15 159

Table 3: Spam volume for different phishing types.

is widely believed that they are run by highly organized criminal gangs.

Just as average lifetimes did not tell the entire story in the last section, we can learn a bit more by examining the temporal distribution of spam volume. Figure 3 plots the CDF of aggregate spam volume over time. For both phishing types, the vast majority of spam (in terms of volume) is sent on the first day of the campaign: 53% for ordinary phishing attacks and 85% for fast-flux. The remaining spam is sent out over the course of several days, and in some cases, weeks. Overall, the spam campaigns advertising fast-flux phishing domains die out sooner – 95% of campaigns advertising fast-flux attacks finish within 5 days, compared to 76% for ordinary phishing attacks.

These findings are compatible with the CDF of spam campaign durations in Figure 1. Just as fast-flux spam campaigns have shorter lifetimes, they also distribute the bulk of their spam more quickly.

Figure 4 plots CDFs relative to the timing of phishing website availability. The left-hand graph plots the CDF of the volume of phishing spam relative to the time the associated phishing website was *first detected*. Negative numbers indicate that the spam was sent before the website was detected, and positive numbers indicate that the spam continued to be sent after the website was detected. For ordinary phishing websites, 16% of spam volume was sent more than a day prior to the website's detection, 3% of the total volume was sent more than a day after the website was detected, and the remaining 81% was sent within a day before or after the website was detected. For fast-flux websites, the breakdown is similar. However, the tails of the spam distribution are far more skewed for ordinary phishing than for fast-flux attacks: all of the 15% of spam advertising fast-flux more than one day prior to detection was sent on the *second* day before detection; by contrast, spam advertising ordinary phishing websites was spread out over two months before the website was identified.

The right-hand graph in Figure 4 plots the CDF of the volume of phishing spam relative to the time the associated phishing website was *removed*. The pattern identified here is somewhat different. 99.997% of spam advertising fast-flux attacks is sent out prior to the websites' removal. Ordinary phishing campaigns also mainly stop prior to the host's removal – just 4% of all spam messages are sent after clean-up.

5 Discussion

5.1 What metric of phishing harm is best?

Ideally, we would measure the harm caused by phishing directly, either by counting its victims or the resulting financial losses. Previous research has attempted to count victims, by using either a custom browser extension [1] or collecting server logs from phishing websites [5]. However, such an approach is not always feasible and the sample size is usually small.

Alternatively, we can infer harm from indirect measurements such as counting spam or phishing websites. Table 4 presents three possible indirect metrics of phishing harm. First, we could simply count the number of phishing websites, as has been done by the APWG [9]. By this measure, ordinary phishing attacks constitute the vast bulk (97%) of harm.

However, some phishing websites are removed within minutes, while others stay up for months. Perhaps it would be better to tally the aggregate lifetime of phishing websites. By this measure, ordinary phishing attacks still cause the most harm, accounting for 68% of the total website uptimes.

Another option is to consider the amount of spam dedicated to each of the attacks, since it is spam that drives victims to the fake websites. By this measure, the out-

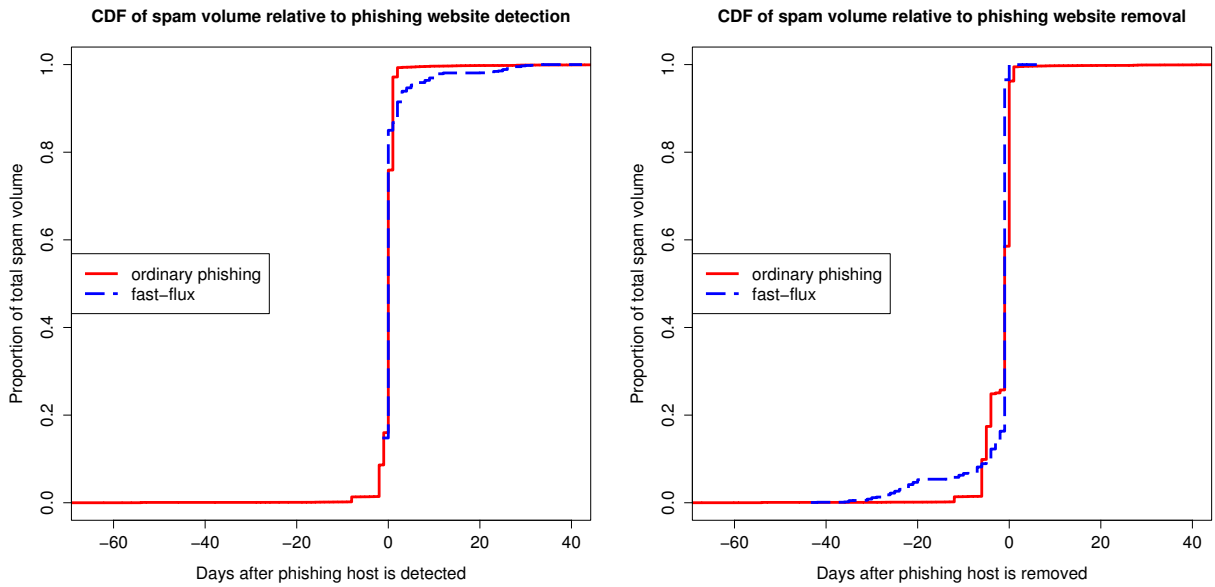

Figure 4: At left, cumulative distribution function of the volume of phishing spam relative to the *detection time* of the associated phishing host. At right, the cumulative distribution function of the volume of phishing spam is given relative to the *removal time* for the associated host. For both graphs, negative numbers indicate that the spam was sent before the website was detected (or removed, respectively), and positive numbers indicate that the spam continued to be sent after the website was detected (or removed, respectively).

	Websites		Website lifetime		Spam volume	
	#	%	Hrs	%	#	%
Ordinary	4 084	97.0	20 602.7	68.0	978 693.1	32.0
Fast-flux	120	3.0	9 673.8	32.0	2 080 035.7	68.0

Table 4: Which phishing type is more effective? Candidate metrics include the number of websites used, spam volume and aggregate website lifetimes.

come is reversed: spam advertising fast-flux websites accounts for 68% of the total! While all three measures have merit, it seems that spam volumes capture best the relative exposure to phishing.

5.2 Does phishing website take-down help?

Having explored the relationship between the timing of spam transmission and phishing website lifetimes, we can draw some conclusions on the effectiveness of phishing website take-down. From Figure 4 (right), we can see that the bulk of spam is sent prior to the phishing website's removal. However, we cannot conclude from this that website take-down is futile; we can only conclude that most spamming stops once the website has been removed. From Figure 4 (left), we can see that most spam is sent before around the time the website is first detected. However, there is a sizable fraction of spam that continues after detection. Is this for websites that remain alive, or is it haphazardly sent for long-dead phishing websites?

Figure 5 attempts to answer this question. It plots the proportion of *live* websites that continue to send spam x days after being detected. For instance, 75% of fast-flux and 35% of ordinary phishing attacks whose websites continue to operate after one week send new spam before being removed. A substantial fraction of phishing websites are advertised by spam up to one month after the website has first been detected. The few websites that remain for more than one month (some lasting up to 4 months in our sample) do not receive additional spam. Live fast-flux websites are around twice as likely to be advertised in fresh spam compared to ordinary phishing, further reinforcing the view that fast-flux attackers

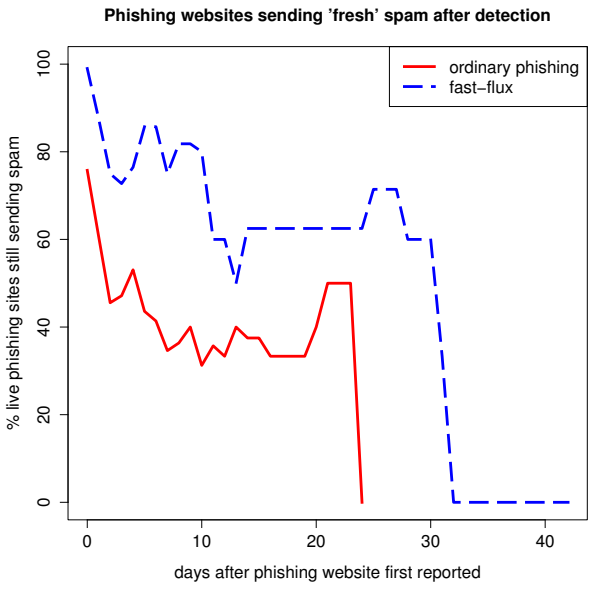

Figure 5: Proportion of live phishing websites sending 'fresh' spam after detection. 'Fresh' means that new spam messages continue to be sent after x days.

are more competent. Because many unremoved websites continue to be the subject of spam advertisements, we conclude that take-down does indeed appear to have a positive effect.

6 Conclusion

Empirical study of malicious online activity is hard. Attackers remain elusive, compromises happen fast, and strategies change frequently. Unfortunately, each of these factors cannot be changed. However, one way researchers can help themselves is to bring together different data sources on attacks. In this paper, we have combined phishing website lifetimes with detailed spam data, and consequently we have provided several new insights.

First, we have demonstrated the gravity of the threat posed by attackers using fast-flux techniques. They send out 68% of spam while hosting only 3% of all phishing websites. They also transmit spam effectively: the bulk is sent out early, it stops once the site is removed, and it keeps going whenever websites are overlooked by the take-down companies. In this respect, we also conclude that long-lived phishing websites continue to cause harm and should therefore be taken down.

In the future, we aim to examine a longer time period to reinforce our conclusions. We could link together even more data, using web usage statistics to close the loop between spam transmission and individual harm. From our efforts, we can safely conclude that combining disparate data leads to a greater understanding of attacker behavior, and this approach could usefully be extended to other areas such as web-based malware.

References

[1] D. Florêncio and C. Herley, Evaluating a trial deployment of password re-use for phishing prevention, in *Anti-Phishing Working Group eCrime Researchers Summit (APWG eCrime)*, 2007, pp. 26–36.

[2] T. Holz, C. Gorecki and F. Freiling: Detection and Mitigation of Fast-Flux Service Networks. In *15th Network & Distributed System Security Symposium (NDSS)*, 2008.

[3] IronPort Anti-Spam. `http://www.ironport.com/technology/ironport_antispam.html`

[4] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, and S. Savage: Spamalytics: An Empirical Analysis of Spam Marketing Conversion. In *15th ACM Conference on Computer and Communications Security (CCS)*, 2008.

[5] T. Moore and R. Clayton: Examining the impact of website take-down on phishing. In *APWG eCrime*, 2007, pp. 1–13.

[6] T. Moore and R. Clayton: The consequence of non-cooperation in the fight against phishing. In *APWG eCrime*, 2008, pp. 1–14.

[7] T. Moore and R. Clayton: The impact of incentives on notice and take-down. In M. E. Johnson (ed.): *Managing Information Risk and the Economics of Security*, pp. 199–223. Springer, New York, 2008.

[8] T. Moore and R. Clayton: Evil Searching: Compromise and Recompromise of Internet Hosts for Phishing. To appear in *FC09*, February 2009.

[9] R. Rasmussen and G. Aaron: Global Phishing Survey: Domain Name Use and Trends 1H2008. `http://www.antiphishing.org/reports/APWG_GlobalPhishingSurvey1H2008.pdf`

[10] SpamCopWiki: SpamTrap. 21 July 2006. `http://forum.spamcop.net/scwik/SpamTrap?time=2006-07-21+21\%3A17\%3A40`