

Temporal Correlations between Spam and Phishing Websites

Tyler Moore, Richard Clayton and Henry Stern

Center for Research on Computation and Society
Harvard University

USENIX LEET '09
Boston, MA
April 21, 2009



Outline

- 1 Data collection methodology
 - Motivation
 - Phishing website and spam data sources
 - Examining our datasets
- 2 Comparing website lifetimes and spam campaigns
 - Spam campaign duration
 - Phishing spam volume over time
- 3 Discussion
 - What metric of phishing harm is best?
 - Does phishing website take-down help?

Outline

- 1 Data collection methodology
 - Motivation
 - Phishing website and spam data sources
 - Examining our datasets
- 2 Comparing website lifetimes and spam campaigns
 - Spam campaign duration
 - Phishing spam volume over time
- 3 Discussion
 - What metric of phishing harm is best?
 - Does phishing website take-down help?



Motivation

- Removing impersonating content from hosting website is key phishing countermeasure
 - Banks, or 3rd party **take-down companies**, collect 'feeds' of phishing URLs
 - Verify URLs in feed, then issue take-down notices to relevant ISPs and/or registrars
 - Average phishing website lifetime (eCrime '07): 62 to 95 hours
 - Long tail of long-lived phishing websites (7% > 1 week)
- Motivating question: **do long-lived phishing websites matter?**
 - Our view: only if victims still visit the website
 - Since spam drives victims to phishing websites, we must compare the **timing of phishing spam** to the time phishing websites are alive



Motivation

- Removing impersonating content from hosting website is key phishing countermeasure
 - Banks, or 3rd party **take-down companies**, collect 'feeds' of phishing URLs
 - Verify URLs in feed, then issue take-down notices to relevant ISPs and/or registrars
 - Average phishing website lifetime (eCrime '07): 62 to 95 hours
 - Long tail of long-lived phishing websites (7% > 1 week)
- Motivating question: **do long-lived phishing websites matter?**
 - Our view: only if victims still visit the website
 - Since spam drives victims to phishing websites, we must compare the **timing of phishing spam** to the time phishing websites are alive



Motivation

- Removing impersonating content from hosting website is key phishing countermeasure
 - Banks, or 3rd party **take-down companies**, collect 'feeds' of phishing URLs
 - Verify URLs in feed, then issue take-down notices to relevant ISPs and/or registrars
 - Average phishing website lifetime (eCrime '07): 62 to 95 hours
 - Long tail of long-lived phishing websites (7% > 1 week)
- Motivating question: **do long-lived phishing websites matter?**
 - Our view: only if victims still visit the website
 - Since spam drives victims to phishing websites, we must compare the **timing of phishing spam** to the time phishing websites are alive

Data sources

- Phishing website lifetimes
 - Amalgamate several feeds: PhishTank, APWG, one large brand owner, and two take-down companies (each a combination of outside feeds and proprietary collection)
 - Automated testing system continuously queries sites until they stop responding or change
- Phishing spam campaigns
 - Subset of Ironport's spam corpus marked as phishing
 - Primarily 3rd party spam traps, but also customer submissions
 - Define **spam campaign** as all spam associated with a single phishing host
 - Define **spam campaign duration** as the time difference between first and last email advertising phishing host

Types of phishing websites

- Ordinary phishing website hosting
 - Free webspace
(<http://www.bankname.freespacesitename.com/signin/>)
 - Compromised machine
(<http://www.example.com/user/images/www.bankname.com/>)
- Fast-flux-hosted phishing websites
 - Register many innocuous-sounding domains (lof80.info)
 - Send out phishing email with URL
<http://www.volksbank.de.netw.oid3614061.lof80.info/vr>
 - Resolve domain to random selection of 5 or 10 botnet-infected machines, which proxy to a back-end server



Phishing datasets for our study

- Phishing website dataset
 - 12 693 phishing URLs for last week of September 2008
 - Pares down to 4 084 ordinary websites and 120 fast-flux domains
- Phishing spam dataset
 - Checked phishing spam sent Jun – Dec 2008
 - 430 ordinary phishing hosts appeared in spam list
 - 103 fast-flux phishing domains appeared in spam list



Questions we examine

- What can we measure?
 - 1 Spam campaign volume
 - 2 Spam campaign duration
 - 3 Phishing website lifetimes
- How should we measure it?
 - 1 On its own
 - 2 Phishing website lifetimes: relative to the start/end of spam campaigns
 - 3 Phishing spam campaigns: relative to first appearance/take-down time

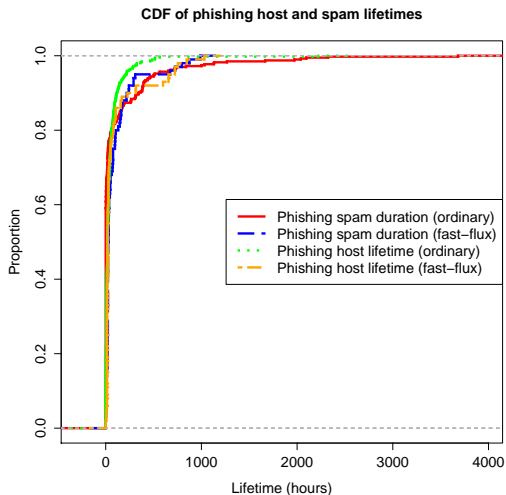
Outline

- 1 Data collection methodology
 - Motivation
 - Phishing website and spam data sources
 - Examining our datasets
- 2 Comparing website lifetimes and spam campaigns
 - Spam campaign duration
 - Phishing spam volume over time
- 3 Discussion
 - What metric of phishing harm is best?
 - Does phishing website take-down help?

Lifetimes of phishing websites and spam campaigns

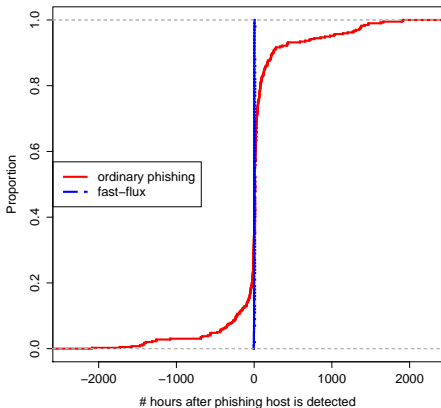
	Website lifetime (hrs)		Spam duration (hrs)	
	mean	median	mean	median
Ordinary	52	18	106	0
Fast-flux	97	21	97	28

CDF of phishing website and spam campaign lifetimes

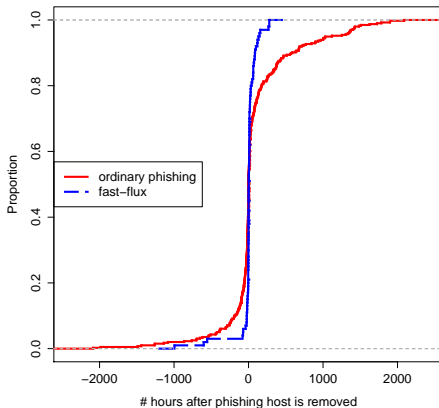


Spam timing relative to phishing website uptime

CDF of time difference between host detection and first spam



CDF of time difference between host removal and last spam

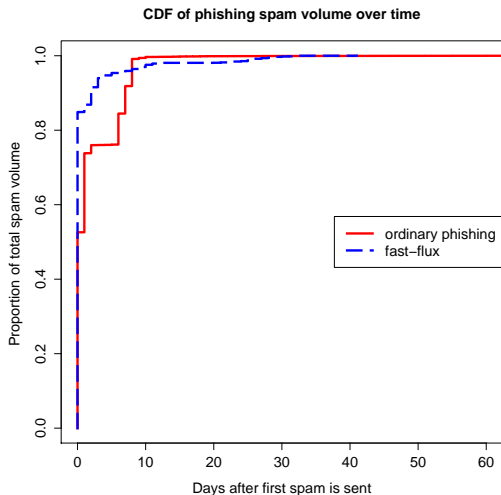


- 0 is time host appears (left graph) or is removed (right graph)
- Left graph: time for **first** spam in campaign
- Right graph: time for **last** spam in campaign

Initial observations

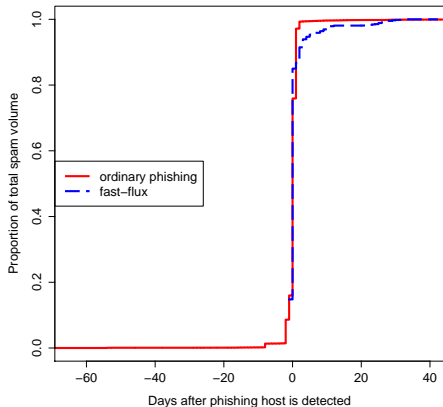
- Fast-flux hosting more tightly correlated with spam transmission
 - Most spam starts around time website appears
 - Most spam stops around time website is removed
- Ordinary phishing websites exhibit higher variance
 - First spam may be sent well before or after website appears
 - 29% of spam campaigns send final message more than a day **before** the website is removed
 - 35% of spam campaigns send final message more than a day **after** the website is removed

Spam volume over time

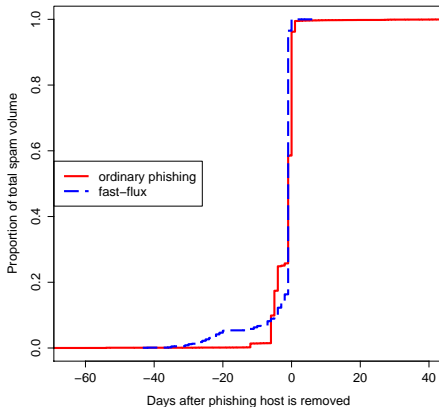


Spam volume relative to phishing website uptime

CDF of spam volume relative to phishing website detection



CDF of spam volume relative to phishing website removal



- 0 is time host appears (left graph) or is removed (right graph)
- Both graph: volume of spam sent by time t

Observations on spam volume

- Most spam is sent around the time the website appears
 - Almost all fast-flux spam appears on the day surrounding website appearance
 - 16% of ordinary phishing spam sent more than a day before detection, 3% more than a day after detection
- Most spam stops once the website is removed
 - 99.997% of fast-flux spam is sent prior to website removal
 - 4% of spam advertising ordinary phishing websites are sent out after removal

Outline

- 1 Data collection methodology
 - Motivation
 - Phishing website and spam data sources
 - Examining our datasets
- 2 Comparing website lifetimes and spam campaigns
 - Spam campaign duration
 - Phishing spam volume over time
- 3 Discussion
 - What metric of phishing harm is best?
 - Does phishing website take-down help?

How do you measure the impact of phishing?

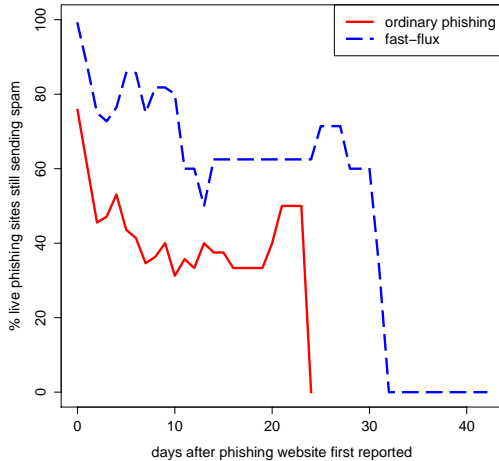
	Websites		Website lifetime		Spam volume	
	#	%	Hrs	%	#	%
Ordinary	4 084	97.0	20 602.7	68.0	978 693.1	32.0
Fast-flux	120	3.0	9 673.8	32.0	2 080 035.7	68.0

Do long-lived phishing websites matter?

- Some phishing websites are very long-lived
- Should we bother removing them?
- If spam is still being sent for the websites, then users may still be at risk

Phishing website take-down does help!

Phishing websites sending 'fresh' spam after detection



Conclusion

- We have brought together data on phishing website lifetimes and the spam campaigns which advertise them
- Most spam is sent around the time when the website is alive
- Phishing attacks using fast-flux techniques transmit spam more effectively
- Phishing website take-down helps, because spam transmission continues until removal
- <http://people.seas.harvard.edu/~tmoore/>