Temporal Correlations Between Spam and Phishing Websites

Richard Clayton

**Joint work with Tyler Moore (Harvard)
& Henry Stern (Cisco IronPort)**

Phishing site take-down

- Removing website content is a key countermeasure to phishing
- Banks & “take-down companies” collect “feeds” of phishing URLs, then approach hosting sites (or registrars)
- We use this data to track website lifetimes
  - data from PhishTank, two take-down companies, one large brand-owner plus the APWG feed (note that all of these are amalgamations of many other sources)

Do long lifetimes matter?

- Many sites removed within a day, but our measurements show a longggg tail!
- Does this matter?
  - – only if people are still visiting the website
  - – hence to assess the harm of a long-lived site, we should examine email spam data to determine email spam “campaign” lifetimes

Take-down measurements (Jan08)

	Total	Mean (hours)	Median (hours)
Free webhosting	395	48	0
when brand owner aware	240	4.3	0
when brand owner unaware	155	115	29
Compromised machines	193	49	0
when brand owner aware	105	3.5	0
when brand owner unaware	155	104	10
Rock-phish domains	821	70	33
Fast-flux domains	314	96	25

Email data from Cisco IronPort

- IronPort handles many millions of emails for many thousands of customers
- They operate spam-traps & receive spam from customers & others
- All the “spam URLs” are extracted (and decoded & de-obfuscated)
- We considered a dataset of all URLs seen between June and December 2008

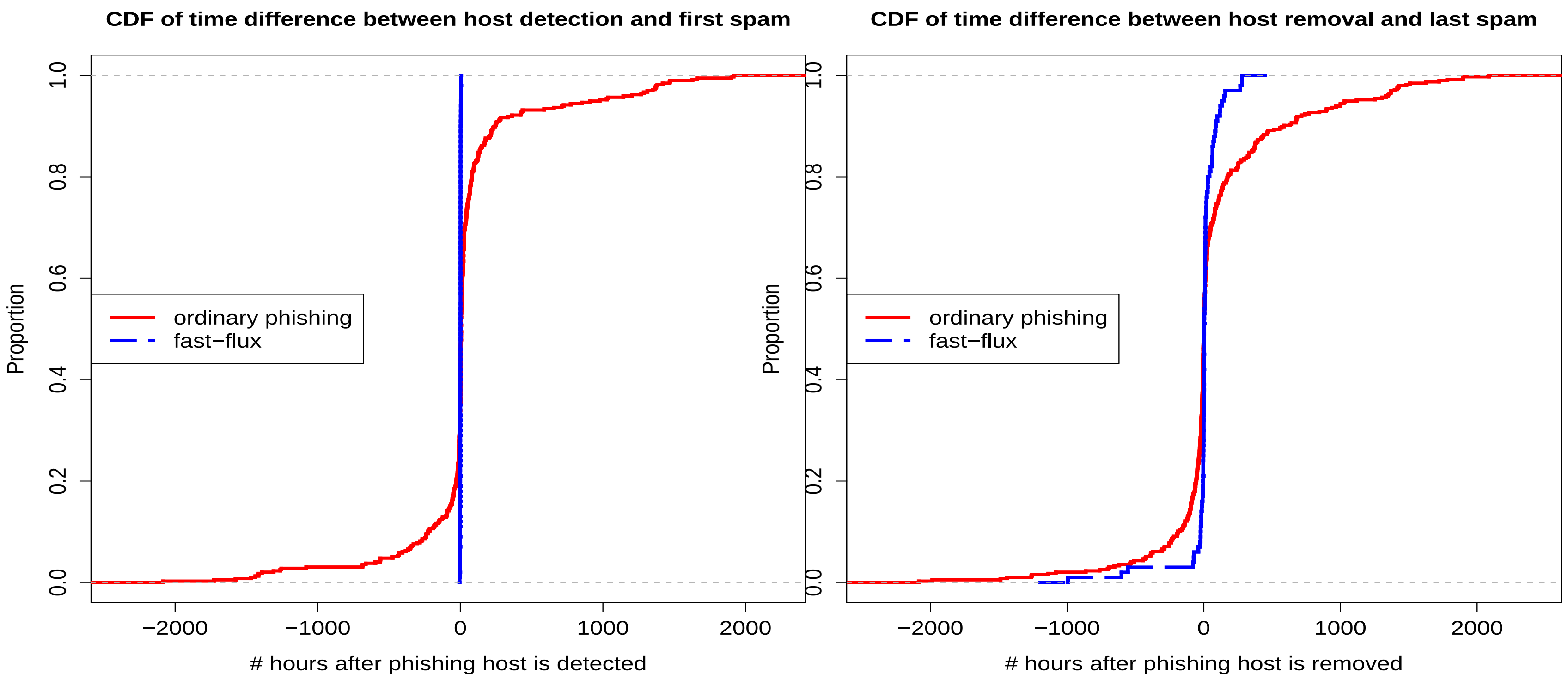
Phishing websites

- Considered all new sites 24–30 Sep 2008
  - 12693 URLs => 4084 websites (compromised & free hosting), 120 fast-flux domains
- Matched (generic) URL in the email dataset
  - “spam campaign” is time from first to last sighting
  - some were zero length (URL only seen once)
- Limited spam coverage (surprisingly!?!)
  - 430 sites (11%), 103 fast-flux domains (86%)

Lifetimes (Sep 08; awareness not considered)

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

Correlation of lifetimes


CDF of time difference between host detection and first spam

CDF of time difference between host removal and last spam

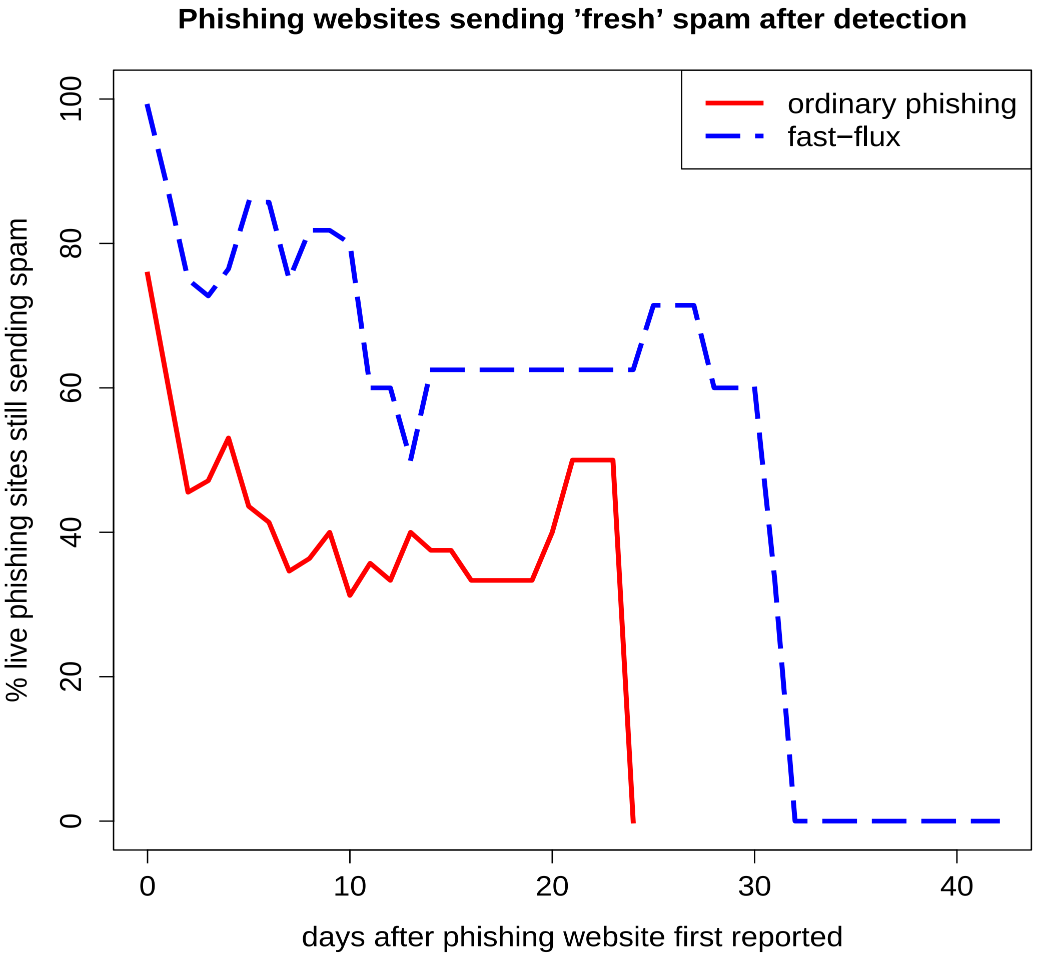
Fast-flux domains appear in phishing feeds almost immediately after first email; and spam ceases promptly when site removed.

Far less correlation occurring for “ordinary” phishing websites.

Volume of phishing spam

- 68.3% of spam was for fast-flux domains
  - just for 103 domains
- 31.7% of spam was for other sites
  - for the 430 websites which had any spam at all
- See paper for the volume/time distribution
  - take-home: fast-flux campaigns often slow before removal; ordinary sites often low volume before detection

So, do long-lived sites matter?



If website remains up then email is still being sent (for weeks).

Hence website removal really does seem to be important!

NB: very long-lived fast-flux sites were in Ecuador TLD

What’s causing most damage ?

	Websites		Lifetime (hrs)		Spam volume
	Total	%	Total	%	
Ordinary	4084	97%	20603	68%	32%
Fast-flux	120	3%	9674	32%	68%

Two sane measures of damage: loss of money/confidence

Website lifetime approximates to loss of money (assuming spam equally convincing); Spam volume approximates to loss of confidence (assuming spam equally likely to reach inbox).

In practice, law enforcement just chase high profile targets (?)

Temporal Correlations Between Spam and Phishing Websites

BLOG:

http://www.lightbluetouchpaper.org/

PAPERS:

http://www.cl.cam.ac.uk/~rnc1/publications.html