Spam and Phishing

Richard Clayton

joint work with
Tyler Moore [Harvard] & Henry Stern [Cisco Ironport]

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Yahoo!
15th July 2009

Phishing websites

- Compromised webhosts (76% in Jan 2008)
  - – vulnerable sites found by “evil” search
  - – website uploaded as a ZIPfile “kit”
  - – PHP pages generate email to @gmail drop address
  - – many sites exploited time and again
- Free webspace (17% in Jan 2008)
  - – as above, but “free” account signed up for
- Remaining 7% are specials…

“Fast-flux hosting”

- HTTP relays hosted on compromised end-user machines (part of a “botnet”)
- Back-end “mothership” remains invisible
- DNS regularly resolves to new IP addresses
  - hence the “fast-flux” name
  - previous “rock phish” scheme used small number of static relays that were pre-qualified, nowadays approach is to use 5 or 10 A records in parallel

Take-down

- Main phishing countermeasure is “take down”
- Banks & “take-down companies” collect “feeds” of phishing URLs (mainly from spam)
- Hosting sites are asked to remove bad pages
- For fast-flux, registrars must remove domain
- We’ve been using the feeds (since early 2007) to track the effectiveness of take-down and to measure the impact of fast-flux techniques

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

Do long lifetimes matter?

- Many sites removed fast
  - when bank knows about site, 4.3 hours
  - when bank does not know about site, 4.3 days
- Our measurements show a longggg tail!
- Does this matter?
  - only if people are still visiting the website
  - hence to assess the harm of long-lived site, we should determine email spam “campaign” lifetimes

Email data from Cisco IronPort

- IronPort handles many millions of emails for many thousands of customers
- They operate spam-traps & receive spam reports 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

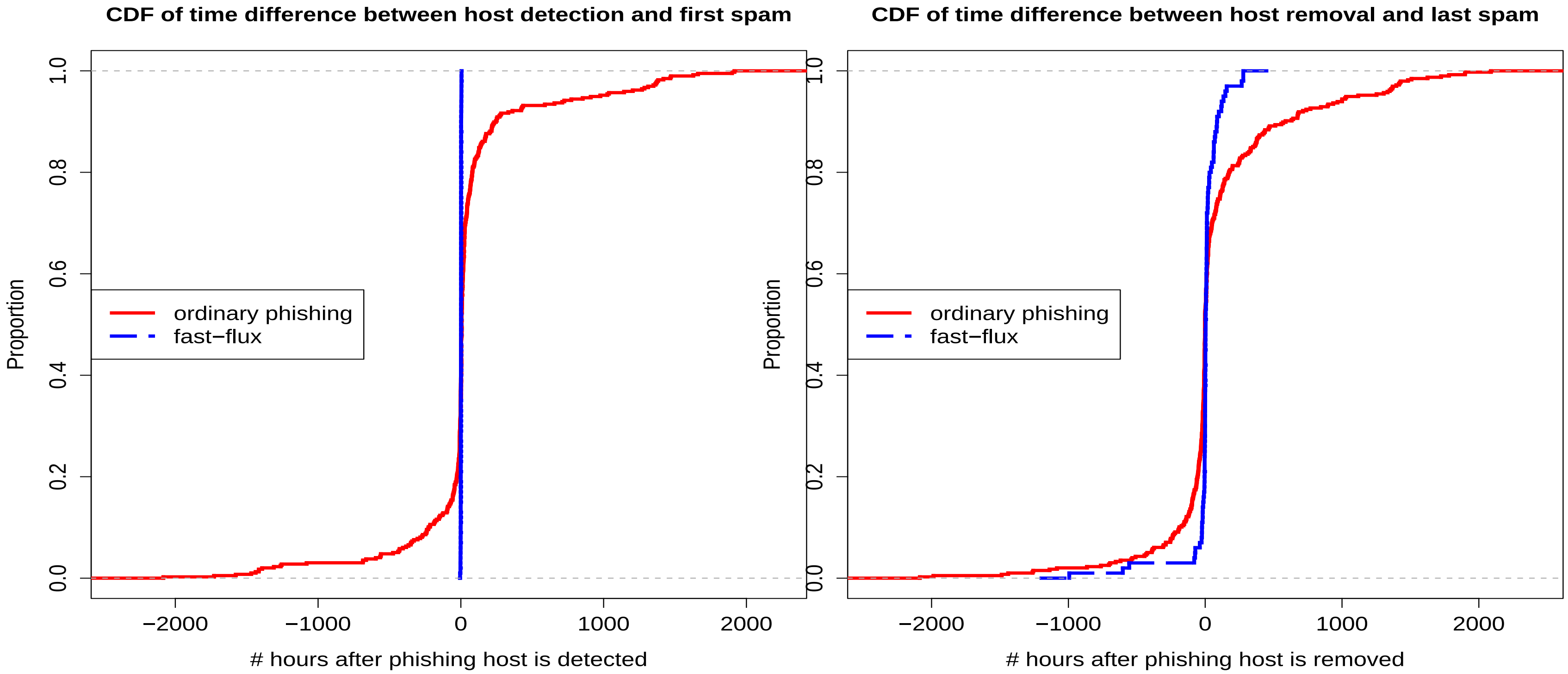
Phishing websites

- 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

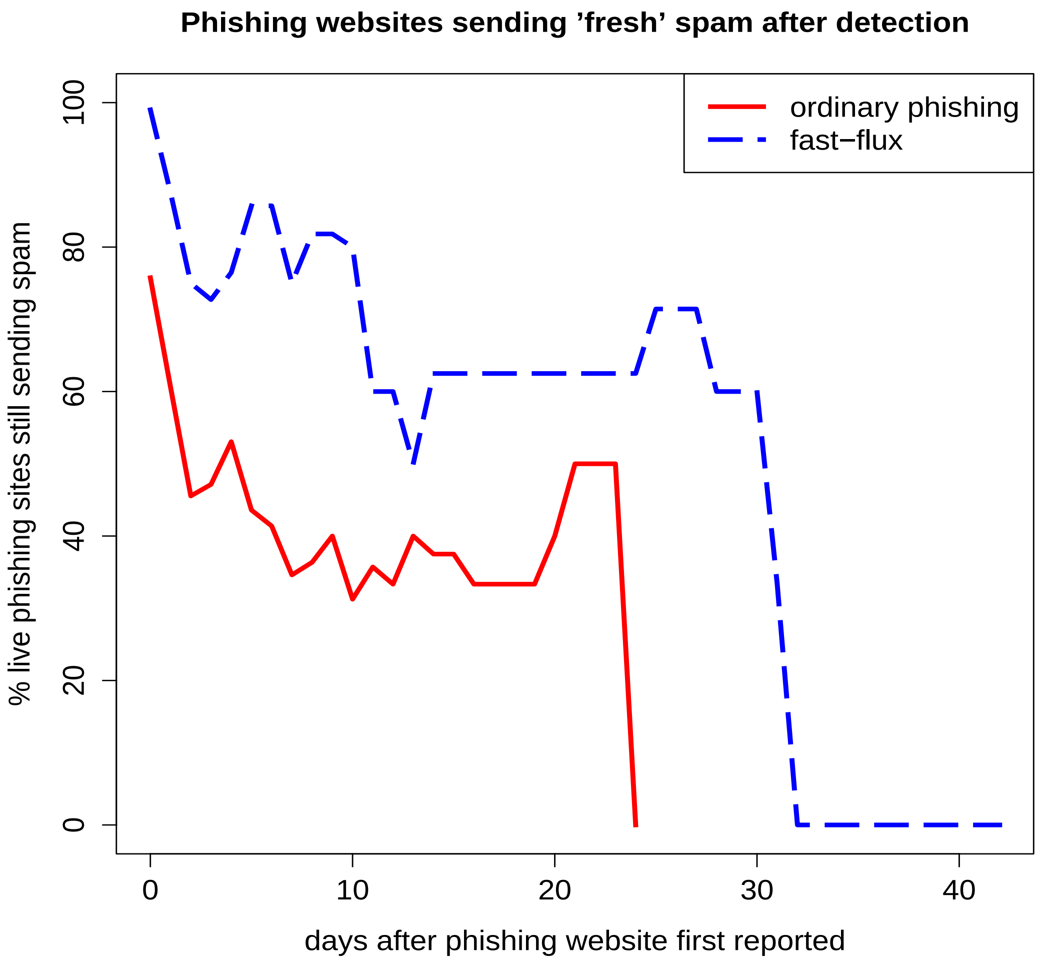
Fast-flux domains appear in phishing feeds almost immediately after first email; and spam ceases promptly when site removed.

Far less correlation occurring for “ordinary” phishing websites.

Volume of phishing spam

- 68.3% of the spam was for fast-flux domains
  - for 103 domains (17 domains weren’t seen in spam)
- 31.7% of the spam was for other sites
  - NB only had spam sample for 430 websites (11%)
- See paper for the volume/time distribution
  - the take-homes are: fast-flux campaigns often slow down before removal; ordinary sites often at a low volume before detection occurs

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 damage measures: loss of money/confidence
- Website lifetime approximates to loss of money (***if*** spam equally convincing); Spam volume approximates to loss of confidence (***if*** spam delivery equally likely)
- In practice, police choose the high profile targets (! ?)

How important is phishing?

- Losses may be from phishing, MITM malware, ATM skimmers, or merchant compromise
- We measured (Spring 2007) average website lifetimes & average visitors to estimate losses
  - “non-rock” was, at that time, $178 million
  - we doubled this to include rock-phish => $350m
  - this was based on $572 per victim
  - compare this with Gartners’ overall $2 billion

The toolbar data

- Dinei Florêncio and Cormac Herley (APWG 2007) considered password re-use
- Customised IE7 add-on spotted when same password used at two different websites
- Saw 101 events from 436K users in 3 weeks
- This is a rate of 0.40% per year
- Our data equates to 0.34% per year (US only)
  - so pretty close, all things considered

Is there over-phishing?

- Cormac Herley & Dinei Florêncio (NSPW 08)
  - argue phishing is a “tragedy of the commons”
  - viz: too many players leads to over-phishing
  - key question: have we reached equilibrium?
- They critically examine victimisation studies
  - Gartner (2005: 0.5%, 2006: 1.05%, 2008: 2.18%)
  - but margin of error just about as big (*c*.1.4%) !
  - huge issues of refusal rates, and “telescoping”
  - also weren’t distinguishing “lottery scams”

H&F also unimpressed by $572

- Average loss figures calculated from surveys
  - small numbers scaled up to US population
  - then rounded ? (losses close to $2bn, $3bn, $4bn)
- But figures are dominated by outliers
  - e.g. one individual losing $485K
  - mean can be $800, median $200
- cf UK figures £23m in 2007, £53m in 2008
  - NB: figures don’t include money clawed back

Nobel Prize for Economics

- “Market for Lemons”, George Akerlof, 1970
  - 2001 Laureate for “asymmetric information” work
- Town with good cars and “lemons”
  - a good car (a cherry) is worth $3000
  - a lemon is only worth $1000
  - the equilibrium price for cars in this town will be around $1000, because buyers take the cynical view that they’re likely to get a lemon…
  - various real world fixes for this (warranties etc)

The Underground Economy

- Open outcry IRC channels where phishing proceeds are traded (along with “ciscos”, “roots”, “drops”, “scam pages” etc)
- Described by Thomas & Martin (Team Cymru) in ;login paper in 2006, and measured by others since
  - Ross Anderson compares this with Adam Smith’s pin factory: efficiency from specialisation
- Symantec regularly quotes figures in reports

UE prices are rather low

- Going rate for credit card details is circa $1
  - rarer cards (Sweden/Belgium) maybe $20
- But is a low price good or bad?
  - maybe prices are low because of over-supply?
  - maybe prices are low because no buyers?
  - maybe prices are low because hard to monetize?
  - maybe these are just “price points”?
  - Herley & Florêncio (WEIS 2009) suggest that the explanation is that it’s a “lemons market” !

Are we encouraging phishing?

- When I give talks I regularly suggest to the audience that they should take up phishing, it pays well I say, it’s not very hard, and the chances of being caught are about zero.
  - my lawyer says I should stress I am not serious!
- Herley & Florêncio say I’m wrong about how well it pays – but new entrants are encouraged by the impression given of a share of billions
  - I think we need more work on phishing incomes

The hard questions

- Can we better quantify phishing losses?
- How much damage is there to “confidence”?
- What does a brand lose from being phished?
- Given limited investigative resources, what part of phishing should we tackle?
- How much do phishers earn?
- How do we discourage new criminals?
- How much have we still left to learn?

Spam and Phishing

BLOG:

http://www.lightbluetouchpaper.org/

PAPERS:

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

UNIVERSITY OF
CAMBRIDGE
Computer Laboratory