Evil Searching: Compromise and Recompromise of Internet Hosts for Phishing

Tyler Moore and Richard Clayton

CRCS, Harvard University
Computer Laboratory, University of Cambridge

Financial Crypto
Accra Beach Hotel, Barbados
February 25, 2009

HARVARD
School of Engineering and Applied Sciences

Outline

HARVARD
School of Engineering and Applied Sciences

Outline

HARVARD
School of Engineering and Applied Sciences

Data collection methodology

- We empirically examine phishing website 'take-down'
  - Widely-used countermeasure in fight against phishing
  - Banks, or 3rd party take-down companies, collect 'feeds' of phishing URLs
  - Feeds obtained from banks, third parties and using proprietary spam traps
  - Verify URLs in feed, then issue take-down notices to relevant ISPs and/or registrars
- Amalgamate several phishing 'feeds'
  - One large brand owner
  - PhishTank
  - APWG
  - Two take-down companies (each a combination of outside feeds and proprietary collection)

HARVARD
School of Engineering and Applied Sciences

Phishing-website demographics (Oct '07–Mar '08)

Type of phishing attack	Count	%
Compromised web servers	88 102	75.8
Free web hosting	20 164	17.4
Rock-phish domains	4 680	4.0
Fast-flux domains	1 672	1.4
'Ark' domains	1 575	1.4
Total	116 193	100

- Questions we seek to answer
  - What % of web servers used to host phishing are later recompromised?
  - How are vulnerable web servers found?
  - Does the way vulnerable web servers are found influence the likelihood of later recompromise?

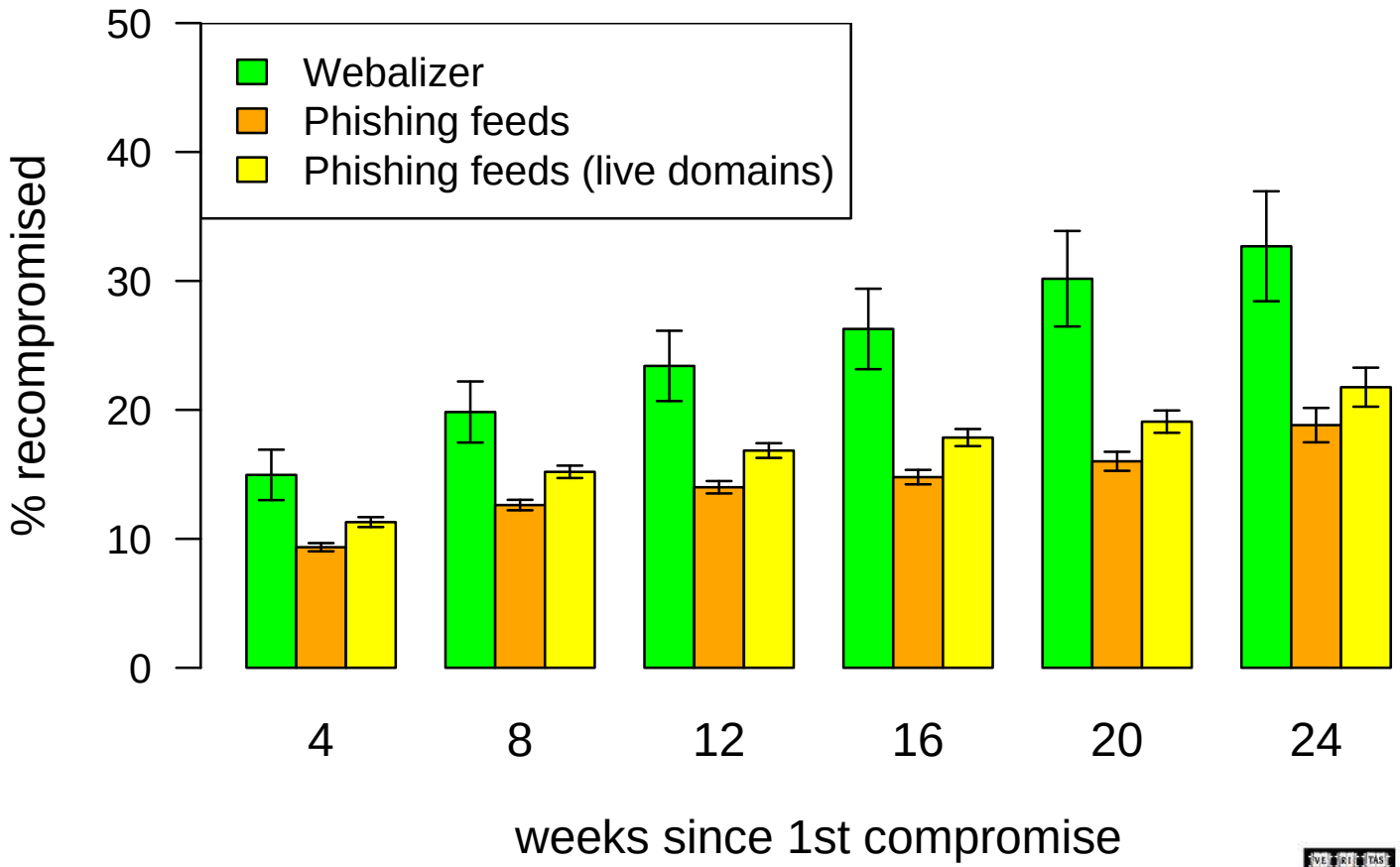
HARVARD
School of Engineering and Applied Sciences

Phishing website recompromise

- What constitutes recompromise?
  - If one attacker loads two phishing websites on the same server a few hours apart, we classify it as one compromise
  - If the phishing pages are placed into different directories, it is more likely two distinct compromises
- For simplicity, we define website recompromise as distinct attacks on the same host occurring ≥ 7 days apart
- 83% of phishing websites with recompromises ≥ 7 days apart are placed in different directories on the server

HARVARD
School of Engineering and Applied Sciences

Phishing website recompromise



HARVARD
School of Engineering and Applied Sciences

Outline

HARVARD
School of Engineering and Applied Sciences

The Webalizer

- Webalizer data
  - Web page usage statistics are sometimes set up by default in a world-readable state
  - We automatically checked all sites reported to our feeds for the Webalizer package, revealing over 2 486 sites from June 2007–March 2008
  - 1 320 (53%) recorded search terms obtained from '`Referrer`' header in the HTTP request
- Using these logs, we can determine whether a host used for phishing had been discovered using targeted search

HARVARD
School of Engineering and Applied Sciences

Types of evil search

- Vulnerability searches: `phpizabi v0.848b c1 hfp1` (unrestricted file upload vuln.), `inurl: com_juser` (arbitrary PHP execution vuln.)
- Compromise searches: `allintitle: welcome paypal`
- Shell searches: `intitle: ''index of'' r57.php`, `c99shell drwxrwx`

Search type	Websites	Phrases	Visits
Any evil search	204	456	1 207
Vulnerability search	126	206	582
Compromise search	56	99	265
Shell search	47	151	360

HARVARD
School of Engineering and Applied Sciences

One phishing website compromised using evil search

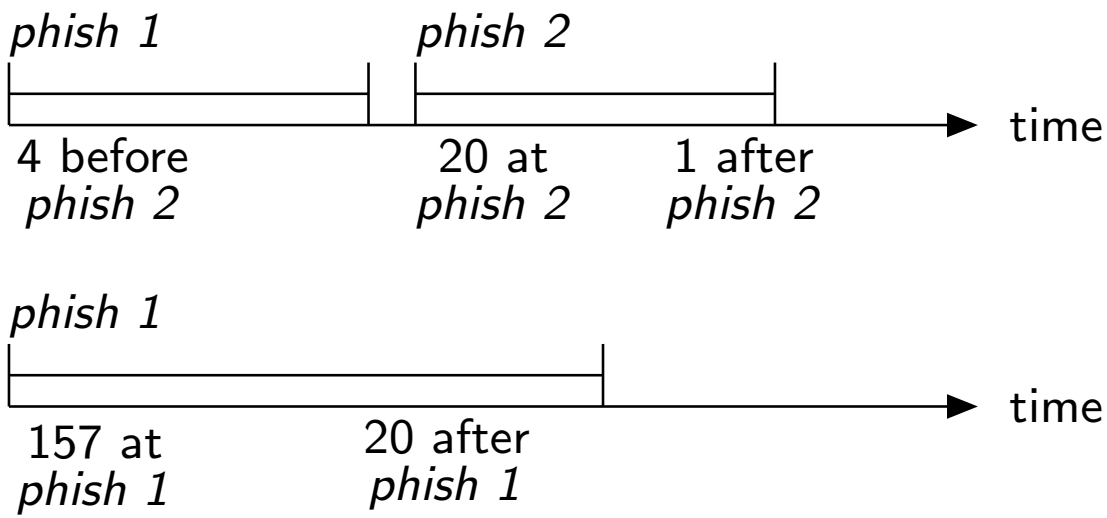
HARVARD
School of Engineering and Applied Sciences

One phishing website compromised using evil search

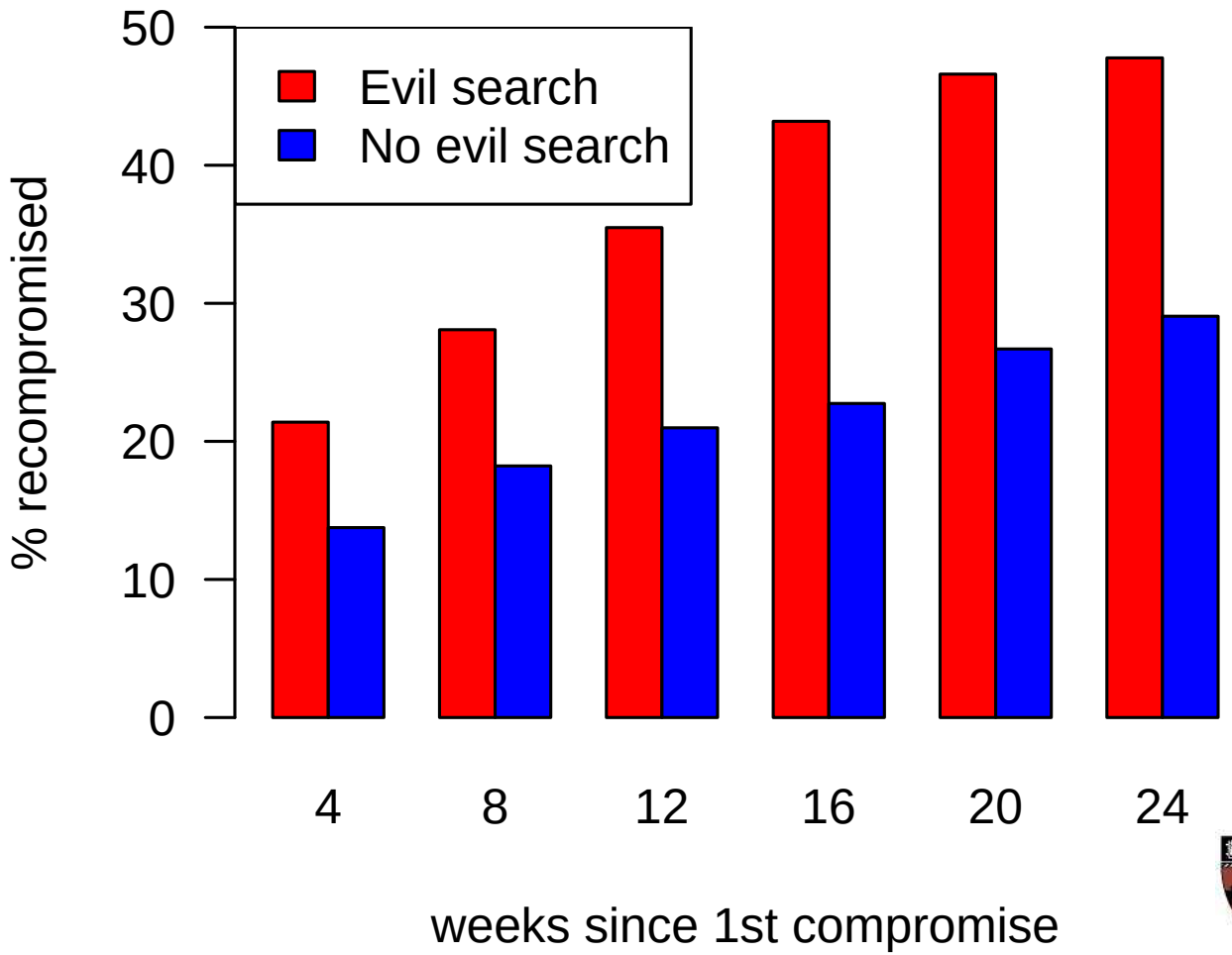
1: 2007-11-30 10:31:33 phishing URL reported: http://chat2me247.com/stat/q-mono/pro/www.lloydstsb.co.uk/lloyds_tsb/logon.ibc.html
2: 2007-11-30 no evil search term 0 hits
3: 2007-12-01 no evil search term 0 hits
4: 2007-12-02 phpizabi v0.415b r3 1 hit
5: 2007-12-03 phpizabi v0.415b r3 1 hit
6: 2007-12-04 21:14:06 phishing URL reported: http://chat2me247.com/seasalter/www.usbank.com/online_banking/index.html
7: 2007-12-04 phpizabi v0.415b r3 1 hit

HARVARD
School of Engineering and Applied Sciences

Timeline of evil web search terms appearing in Webalizer logs



Evil searching makes recompromise more likely



HARVARD
School of Engineering and Applied Sciences

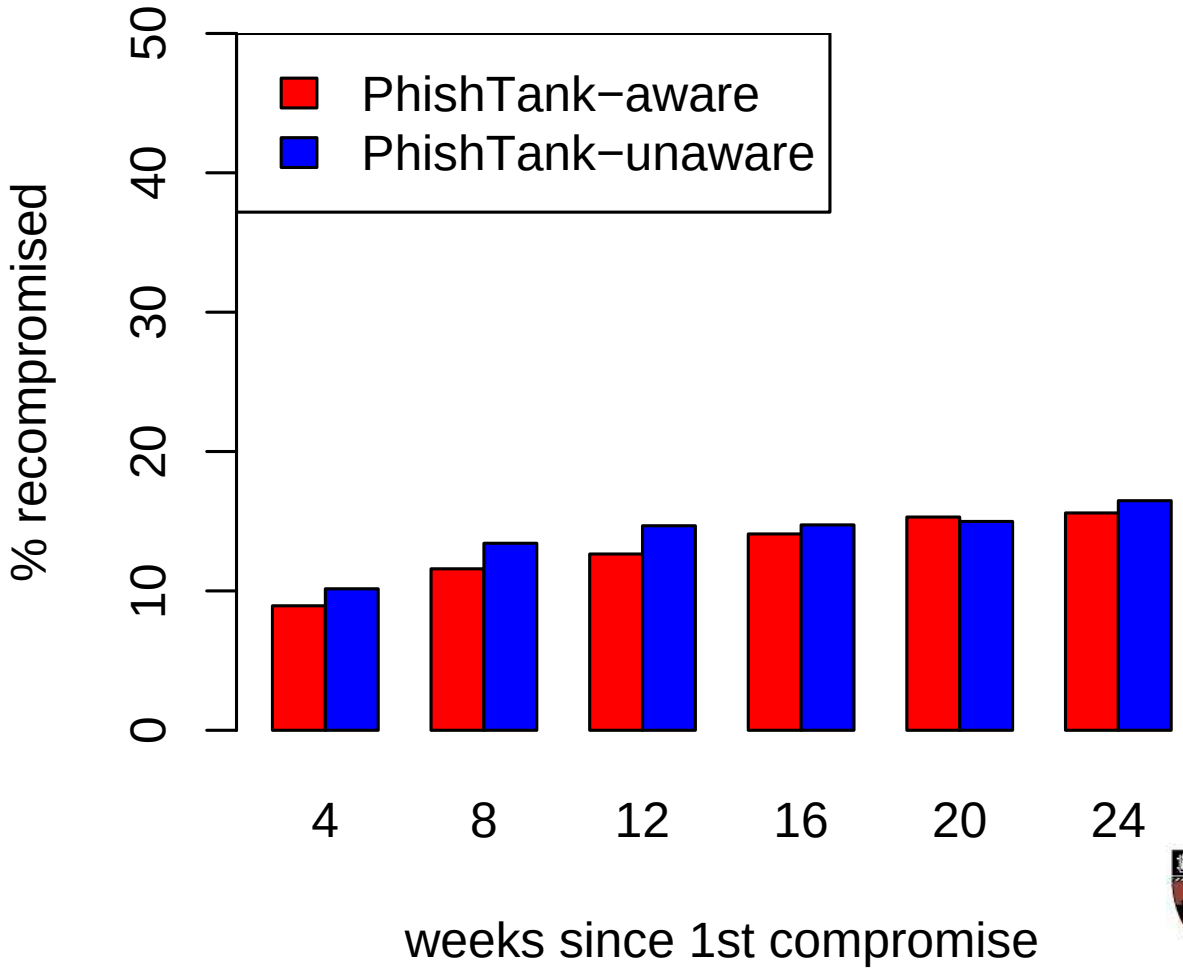
Outline

Public versus private blacklists

- Is it better to hide or publish blacklists of vulnerable hosts?
  - Many fear publishing could help attackers find hosts to recompromise
  - Google's Safe Browsing API only allows verification of known URLs; APWG only shares with trusted parties
  - But might the good from public dissemination (e.g., greater awareness to defenders) outweigh the bad?
  - PhishTank and CastleCops publish lists of phishing URLs
- Fortunately, the data can give us an answer
  - Our test: do websites appearing in PhishTank get recompromised more or less frequently than websites not appearing in PhishTank
  - Caveat: we only compare recompromise rates of new hosts following their first compromise

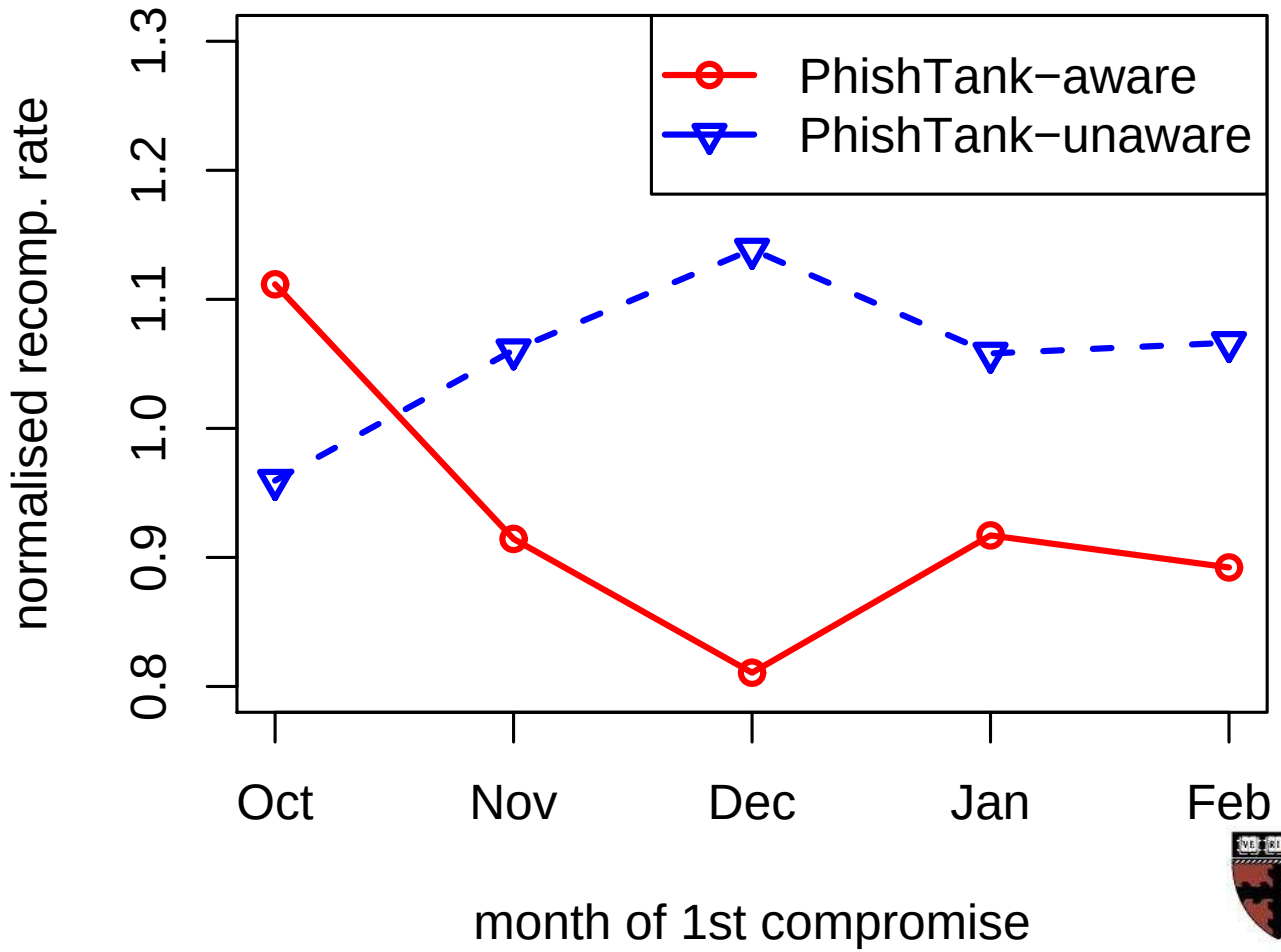
HARVARD
School of Engineering and Applied Sciences

Recompromise rates similar for public and private blacklists



HARVARD
School of Engineering and Applied Sciences

Recompromise rates slightly lower for public blacklists



HARVARD
School of Engineering and Applied Sciences

Outline

HARVARD
School of Engineering and Applied Sciences

Mitigating the impact of evil searches

1. Obfuscating target details
   - Strip out version numbers, etc.
   - But: most searches contained no version numbers; defenders also use searches
2. Evil search penetration testing
   - Run evil search terms and warn affected sites
   - But: searches are only hints; confirming suspicions often illegal
3. Blocking evil search queries
   - But: constructing up-to-date blacklist hard; no incentive for search engines to block
4. Lower reputation of previously phished hosts discoverable by evil search terms
   - SiteAdvisor warns about websites consistently hosting malicious content; why not warn about hosts findable by evil search terms?

HARVARD
School of Engineering and Applied Sciences

Concluding remarks

- We have provided clear evidence that criminals who compromise web servers to host phishing websites use search engines to find them ($\geq 18\%$ of hosts found by evil search)
- 19% of all phishing websites recompromised within 24 weeks, rising to 48% when evil search terms found in the logs
- Phishing hosts disclosed on a public blacklist are slightly less likely to be recompromised than hosts kept hidden