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



yler Moore and Richard Clayton Compromise and Recompromise of Internet Hosts for Phishing

< 6 >

- ∢ ≣ →

Outline

1 Recompromise of phishing websites Data collection methodology Defining recompromise 2 Evil searching Website-usage summaries Evidence for evil searching Evil searching and recompromise PhishTank and recompromise Public v. private blacklists Mitigation strategies and conclusion



Data collection methodology Defining recompromise

Outline

Recompromise of phishing websites Data collection methodology Defining recompromise

2 Evil searching

- Website-usage summaries
- Evidence for evil searching
- Evil searching and recompromise
- PhishTank and recompromisePublic v. private blacklists
- 4 Mitigation strategies and conclusion



A (1) > A (2) > A

Data collection methodology Defining recompromise

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 Applied Sciences

< 口 > < 同 > < 三 > < 三

Data collection methodology Defining recompromise

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

Type of phishing attack	Count	%
Compromised web servers	88102	75.8
Free web hosting	20164	17.4
Rock-phish domains	4680	4.0
Fast-flux domains	1672	1.4
'Ark' domains	1575	1.4
Total	116193	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?



Data collection methodology Defining recompromise

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



Recompromise of phishing websites Evil searching PhishTank and recompromise

Data collection methodology Defining recompromise

Phishing website recompromise



weeks since 1st compromise



Website-usage summaries Evidence for evil searching Evil searching and recompromise

Outline

Recompromise of phishing websites
 Data collection methodology
 Defining recompromise

2 Evil searching

- Website-usage summaries
- Evidence for evil searching
- Evil searching and recompromise
- PhishTank and recompromisePublic v. private blacklists

Mitigation strategies and conclusion



A (1) > A (2) > A

Website-usage summaries Evidence for evil searching Evil searching and recompromise

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 2486 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



ヘロン 人間と 人間と 人間と

Website-usage summaries Evidence for evil searching Evil searching and recompromise

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	1207	
Vulnerability search	126	206	582	
Compromise search	56	99	265	
Shell search	47	151	360	
				School of Engineering

ler Moore and Richard Clayton Compromise and Recompromise of Internet Hosts for Phishing

(日) (同) (三) (三)

and Applied Sciences

Website-usage summaries Evidence for evil searching Evil searching and recompromise

One phishing website compromised using evil search

🍪 phpizabi v0.415b r3	- Google Search - Mozilla Firefox		
<u>File E</u> dit ⊻iew Hi <u>s</u> to	rry <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp		
- 📄 - 🥰	🔝 🌇 🖸 http://www.goog	le.co.uk/search?q=ph 🔽 🕨	G Google
🌆 Getting Started 🔯 L	atest Headlines		
Web Images Map	<u>s News Shopping Mail m</u>	iore 🔻	<u>Sign in</u>
Googl	Phpizabi v0.415b r3 Search: ● the web ○		arch Advanced Search Preferences
Web	Results 11 - 20 of abo	ut 696 for phpizabi v0.415k	э г3. (0.14 seconds)
LDS Dating World Policy Terms of u: LDSDatingWorld.com www.ldsdatingworld.com	, All Rights Reserved. Running , om/?L=chat.c2 - 24k - <u>Cached</u> -	Abuse. Copyright (C) 2007, on PHPizabi v0.415b R3 . • <u>Similar pages</u>	
Our Getaway Copyright (C) 2006, O BirminghamUK Com. www.ourgetaway.co.u	urGetaway, All rights reserved (Running on PHPizabi v0.415b k/?L=users.profile&id=1261 - 4′	DurGetaway is a registered t R3 . 1k - <u>Cached</u> - <u>Similar pages</u>	trademark of
Dream Date & Pe Copyright © 2006 - 20 www.chat2me247.con	r <u>sonals</u> 07 DREAM DATE & PERSON4 n/dating/?L≕users.friendslist - 2	LS. Running on PHPizabi v 5k - <u>Cached</u> - <u>Similar pages</u>	v0.415b R3.
		х ц <i>т</i>	
	re and Richard Clayton	Compromise and Reco	ompromise of Intern

Evidence for evil searching

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 no evil search term 2: 2007-11-30 0 hits 0 hits 3: 2007-12-01 no evil search term 4: 2007-12-02 phpizabi v0.415b r31 hit phpizabi v0.415b r3 1 hit 5: 2007 - 12 - 036: 2007-12-04 21:14:06 phishing URL reported: http://chat2me247.com /seasalter/www.usbank.com/online_banking/index.html phpizabi v0.415b r3 1 hit 7: 2007-12-04



Website-usage summaries Evidence for evil searching Evil searching and recompromise

Timeline of evil web search terms appearing in Webalizer logs



HARVARD School of Engineering and Applied Sciences

Tyler Moore and Richard Clayton Compromise and Recompromise of Internet Hosts for Phishing

Image: A math a math

Website-usage summaries Evidence for evil searching Evil searching and recompromise

Evil searching makes recompromise more likely



Public v. private blacklists

Outline

Recompromise of phishing websites

 Data collection methodology
 Defining recompromise

 Evil searching

 Website-usage summaries
 Evidence for evil searching
 Evil searching

- Evil searching and recompromise
- PhishTank and recompromisePublic v. private blacklists

Mitigation strategies and conclusion



Public v. private blacklists

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
 School of Engineering and Applied Sciences

• □ > • □ > • □ > • □ > •

Public v. private blacklists

Recompromise rates similar for public and private blacklists



Tyler Moore and Richard Clayton Compromise and Recompromise of Internet Hosts for Phishing

Public v. private blacklists

Recompromise rates slightly lower for public blacklists



Tyler Moore and Richard Clayton Compromise and Recompromise of Internet Hosts for Phishing

Outline

• Data collection methodology Defining recompromise Website-usage summaries Evidence for evil searching • Evil searching and recompromise • Public v. private blacklists

4 Mitigation strategies and conclusion



< 🗇 🕨 < 🖻 🕨 <

Mitigating the impact of evil searches

Obfuscating target details

- Strip out version numbers, etc.
- But: most searches contained no version numbers; defenders also use searches
- Evil search penetration testing
 - Run evil search terms and warn affected sites
 - But: searches are only hints; confirming suspicions often illegal
- Blocking evil search queries
 - But: constructing up-to-date blacklist hard; no incentive for search engines to block
- 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?

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



・ロト ・ 同 ト ・ ヨ ト ・ ヨ