TrackBack Spam: Abuse and Prevention

Elie Bursztein, Peifung E. Lam, John C. Mitchell
Stanford University
Introduction

- Many users nowadays post information on cloud computing sites
- Sites sometimes need to link to each other
- However, cross-referencing can become a vehicle for abuses (such as spamming)
- This calls for a study of security issues on cross-referencing between cloud sites
• Blog cross-referencing offers one such example
• Blogs have automated mechanisms, called Linkbacks, to facilitate cross-referencing, and this has been exploited by spammers
We carried out a 1-year study of a major spamming platform, and analyzed 10 million spams.

Gained insight on attacker’s method of operation and resources.

Propose a defense against blog spams.
Outline

- Blog Spam
- Experiment setup : Honey blog!
- Results
- Defense
General Stats on Blogs

- 184 Million blogs world-wide
- 73% of internet users have read a blog
- 50% post comments

Source: universalmccann
Common Blog Platforms

- WordPress
- TypePad
- Movable Type
- Blogger
Why blogs are special

- Blog are designed around the idea of user pushing content
- As an example, Linkbacks allow cross-linking between blogs.
- More specifically, when blog A cites another blog B, a notification of the citation can be sent to B, which can then link back to blog A automatically.
# TrackBack - a type of LinkBack

<table>
<thead>
<tr>
<th>TrackBack URL</th>
<th>The URL of TrackBack capture script</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto discovery of TrackBack URL</td>
<td>Resource Description Framework (RDF)</td>
</tr>
<tr>
<td>Trigger</td>
<td>Code on blog site extracts citations to other blogs</td>
</tr>
<tr>
<td>Notification</td>
<td>HTTP Post</td>
</tr>
</tbody>
</table>
Spectrogram Link:
http://erebus.nmt.edu/blog/spec/20071023_090404.jpg

Mail this    Printer friendly

Trackbacks
The trackback uri for this entry is
http://erebus.nmt.edu/blog/trackback.php/1/12773

Listed below are the weblogs that reference this post

full tilt poker
full tilt poker
Blog : full tilt poker
Tracked on : Sun, 19 Apr 2009 01:06:01 +0000
Trackback Post variables

[title] => Title of the referencing blog entry
[excerpt] => Post excerpt ....
[blog_name] => Mysite blog
Problem

- Trackbacks are used to
  - push spam
  - do malevolent Search Engine Optimization
- One blog spam can reach thousand of users
How big is the problem?

Source: Akismet.com

Total spam: 13,275,940,950
Total ham: 2,701,440,026

Today (UTC, 1 hours left)
Spam today: 12,141,992
Ham today: 2,567,818
Honey Blog

- A blog acting as a potential target for spamming
- Instrumented our blog site and analyzed spams
Setup

- Hosted a real blog (dotclear) with a modified TrackBack mechanism
- Record TrackBacks
- Passive fingerprinting
- Sample the lure site
Unique Spammer IPs

Unique Spammer IPs

User Agents in Spamming

- Mar-Apr 2007
- May-Jun 2007
- July 2007-Apr 2008

- WordPress/1.9
- WordPress 2.1.2
- Opera
- WordPress 1.9
- WordPress 2.1
- IE 6 XP
- WordPress/2.0
- WordPress/2.1.2
- Firefox

Percentage %
Trackback content

- Random keywords revolving around adult theme
- Blog URLs in the Trackback pings are of the form random-words.nx.cn
Trackback Post sample

Apparent Bayesian poisoning against spam filters:

[title] => Please teacher hentai pics


[excerpt] => pics Please teacher hentai pics ...

[blog_name] => Please teacher hentai pics
Spam Workflow

1. Servers submit Trackback spam
2. Spam points to Social network site exploited as relay site
3. Relay site links to lure sites with purported adult content
4. Lure site badgers user to download fake video plugins hosted on malware site
• **Www.nx.cn**, a community hosting site at Ningxia province, PRC

• Exploited by attackers as relay

• The hosting site started to use CAPTCHA (some in Chinese) around May, 2008

• We observed a corresponding drop of spam activities using them as relay
Administrative Divisions of the People’s Republic of China (PRC)
Behind the relay

- Lead to various sites
  - selectedclipz.com, gogomovz.com (purported adult site)
  - vidzwares.com (malware distribution site)
- Need an id in the url download.php?id=429
Whois

Domain Name: GOGOMOVZ.COM
Registrar: ONLINENIC, INC.
Whois Server: whois.onlinenic.com
Referral URL: http://www.OnlineNIC.com
Name Server: NS1.GOGOMOVZ.COM
Name Server: NS2.GOGOMOVZ.COM.
Updated Date: 22-oct-2008
Creation Date: 22-oct-2008
Expiration Date: 22-oct-2009
Registrant:

... ul Beketova 3
Nijnii Novgorod,n/a,RUSSIAN FEDERATION 603057
DNS analysis: related domains

- ns1.clipzsaloon.com
- ns1.clipztube.com
- ns1.freexxxmovz.com
- ns1.itunnelz.com
- ns1.vidzselector.com, and more...
Malware

- Binary flagged as
  - TrojanDownloader:Win32/Zlob.gen!dll
  - Trojan.Popuper.origin
  - Downloader.Zlob.LI
TalkBack

• Designed a secure protocol: TalkBack
• Address the root of the problem: prevent spammers to post notifications
• Key ideas:
  • Lightweight PKI
  • Global rate limiting
Goals

- Sender authenticity
- Receiver authenticity
- Notification integrity
- Notification irrefutability
How it works

1. Seed request
2. Auto-Discovery
3. Talkback posting
4. Talkback reporting
Conclusion

- Linking between cloud sites can become a vehicle for spamming
- One such example is blog TrackBacks
- We did a 1 year study of a major blog spamming platform: 10 million spams analyzed
- Gained insight about TrackBack spam and spammers
- Provided us a basis to build better defense
Related work and alternative approaches

- TrackBack Validator [21]
  - Parsing sender page to find the link
- Reputation system
- IP Blacklisting
- Local rate limiting
Questions ?

Thank you!