Discovering Phishing Dropboxes Using Email Metadata

Tyler Moore\textsuperscript{1} and Richard Clayton\textsuperscript{2}

Computer Science & Engineering Department, Southern Methodist University, Dallas, TX\textsuperscript{1}
Computer Laboratory, University of Cambridge, Cambridge, UK\textsuperscript{2}

APWG eCrime Researchers Summit
October 23, 2012
Outline

1. Introduction and research approach
   - Phishing kits
   - Incoming email metadata

2. Techniques for identifying dropboxes
   - Direct identification of dropboxes
   - Indirect identification of dropboxes
   - Identifying the source of dropbox email
Phishing kits

DIY phishing kits introducing new features

Summary: What are some of the main factors for the increase of phishing attacks, and their maturity from passive emails to blended threats attempting to not just steal personal information, but also infect with malware by embedding client-side vulnerabilities at the pages? It's all a matter of perspective, which in this post will emphasize on the continuing efforts on behalf of phishers to innovate, and introduce new features within the most recently obtained do-it-yourself phishing page generators.

By Dancho Danchev for Zero Day | May 15, 2008 - 08:02 GMT (01:02 PDT)

Phishers are becoming smarter, more targeted: Webense

Bildsoft releases tool for removing Gauss financial malware

Chinese hackers linked to Canada's Telivnet breach

Help us crack Gauss' encryption: Kaspersky Labs

The best of ZDNet, delivered

ZDNet Newsletters
Get the best of ZDNet delivered straight to your inbox

Enter your email address

✓ ZDNet Must Read News Alerts - US: Major news is breaking. Are you ready? This newsletter has only the most important tech news nothing else.

Subscribe Now

Tyler Moore
Discovering Phishing Dropboxes Using Email Metadata 4 / 21
Phishing kits: typical PHP code

```php
<?php
$ip = getenv("REMOTE_ADDR");

$mess = "Email: ".$_POST['email']."\n";
$mess .= "PWord: ".$_POST['passwd']."\n";
$mess .= "IP: ".$ip."\n";

$dest = "dropbox@example.com";
$subj = "PP ReZuLtZ";

if (mail($dest,$subj,$mess))
    { header("Location: /www.paypal.com/"); }
else
    { echo "ERROR! Please go back retry."); }
?>
```
The static nature of phishing kits

- PHP script invariably included in the ZIP archive
- Almost never edited on the server itself
- Thus it is inconvenient to change the subject line or dropbox email address as criminals move across servers
- This means we can link criminal behavior over time
Incoming email metadata

Timestamp
The time that the email is placed into a mailbox.

Source IP address
The machine that sent the email to the email provider.

SMTP “mail from”
The sender of the email, as declared in the SMTP conversation. This can be forged but usually provides some identification of true origin.

SMTP “mail to”
The destination(s) to which the email is being sent. In this context, this information is always valid.

From
‘From:' email header field. It can be set by the phishing kit and is usually entirely bogus.

Subject
‘Subject:' email header field (invariably set by phishing kit).

URLs
These are the URLs from the body of the email.
Why include URLs in email metadata?

- URLs are a very distinctive way to identify email spam
- Email addresses are treated as mailto:// URLs
- If a phishing victim’s account is an email address, then this will turn up in the metadata of a dropbox email
Finding dropboxes

How phishers use dropboxes:

victim1@example.com

From: spam
To: dropbox1@example.com
Subject: P1 ReZuLtUS
user: victim1@example.com
pass: hamster34

How we identify dropboxes:

<table>
<thead>
<tr>
<th>Phish URL</th>
<th>username</th>
<th>pwd</th>
</tr>
</thead>
<tbody>
<tr>
<td>phish1.com</td>
<td><a href="mailto:daucus12@ex.tld">daucus12@ex.tld</a></td>
<td>foo1</td>
</tr>
<tr>
<td>phish2.com</td>
<td><a href="mailto:daucus13@ex.tld">daucus13@ex.tld</a></td>
<td>foo2</td>
</tr>
<tr>
<td>phish3.net</td>
<td><a href="mailto:daucus14@ex.tld">daucus14@ex.tld</a></td>
<td>bar1</td>
</tr>
</tbody>
</table>

To: dropbox1@example.com
Subject: P1 ReZuLtUS
mailto://daucus12@ex.tld
Direct dropbox discovery

- **Our dataset**
  - Phishing URL source: cleaned amalgamation of APWG, PhishTank, brand owner, and brand protection company feeds
  - On June 1 2012 we sent emails with spurious credentials to 170 different websites targeting PayPal reported in May 2012 and found to still be online

- **Results**
  - 28 / 170 emails found in email metadata logs (16.5%)
  - 17 distinct dropbox email addresses
  - Lots of distinctive Subject lines:
    - P1 ReZuLtUS
    - Paypal Spam Result
    - 10.0.0.1 | New PayPal Account
    - [EMAIL: jim@example.com | secret]
Observed dropbox statistics

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td>victims (emails)</td>
<td>380</td>
<td>177</td>
</tr>
<tr>
<td>sites (senders)</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>victims/sites</td>
<td>68.9</td>
<td>13</td>
</tr>
</tbody>
</table>
Limited correlation between # victims and # sites

![Graph showing limited correlation between the number of victims and the number of sites per dropbox. The median number of victims is 177, and the median number of sites is 6.](image)
Recall those distinctive subject lines?

- P1 ReZuLtUS
- Paypal Spam Result
- 10.0.0.1 | New PayPal Account
- [EMAIL: jim@example.com | secret]

We found 15 distinct patterns from the 28 subject lines

Searching for those subject lines in all email yielded 81 new dropboxes (3 times as many as found through direct probing)
An upper bound on criminals targeting PayPal

- \# dropboxes \sim \# criminals
- Some criminals use multiple dropboxes, and criminals inevitably register new dropboxes as their old ones are shut
- But for a small snapshot in time, the \# of dropboxes can serve as an upper bound for the number of criminals operating
- We found 29 dropboxes used to attack PayPal in July 2012 (17 throughout month and 12 for shorter periods)
- Thus we estimate that we found between 20-29 criminals, and our direct identification technique found dropboxes for 16.4% of PayPal phish
- So we estimate 122–164 criminals attacked PayPal in July 2012 (out of 26,900 distinct URLs on 13,018 domain names)
Intersection method to identify dropbox source URLs

1. Identify dropboxes from subject patterns
   - P1 ReZultUS: dropbox1@example.com
   - Paypal Spam Result: dropbox2@example.com
   - dropbox3@example.com

2. Find victims from mailto: URLs in dropbox emails
   - dropbox1@example.com
   - time email received: 2012-06-08 01:28:10
   - victim email: mailto:victim1@example.com
   - V2
   - time email received: 2012-06-08 21:00:01
   - victim email: mailto:victim2@example.com

3. Find phishing URLs by intersecting URLs in victim emails
   - V1 mailto URLs: amazon.com, twitter.com
   - V2 mailto URLs: nytimes.com, facebook.com
   - acount-15f2vb1n.save-data-supportteam165
   - fgg478521fdsd5d1d6.dnstour.com/Uid=98635/
   - acount-15f2vb1n.save-data-supportteam165
   - fgg478521fdsd5d1d6.dnstour.com/Uid=98635/
Interventions possible using the intersection method

1. Identify phishing victims at the time of credential disclosure
   - Regularly run searches for known dropbox subjects and identify victims from mailto: URLs

2. Identify (and block) phishing URLs faster
   - Once two victims have entered their details, identify the phishing URL and email provider can block its other customers from being phished
   - Can also pass along newly discovered URLs to blacklists
Intersection method proof-of-concept

- Inspected one week’s worth of email for dropbox subjects (15-21 July)
- Found 934 victim credentials sent to dropbox from 114 IP addresses
- Of these, 159 victims had email address with metadata we could inspect coming from 47 IP addresses
- Of 47 IP addresses, 25 had one victim with metadata
- This leaves 22 potential phishing URLs to run intersection method
### Intersection method in action

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-07-19 15:16:22</td>
<td>Phish arrived at V1</td>
</tr>
<tr>
<td>2012-07-19 15:20:02</td>
<td>Phish arrived at V3</td>
</tr>
<tr>
<td>2012-07-19 15:21:32</td>
<td>V1 becomes a victim</td>
</tr>
<tr>
<td>2012-07-19 15:48:30</td>
<td>V6 ← <a href="http://77kids.com">http://77kids.com</a> etc.</td>
</tr>
<tr>
<td>2012-07-19 16:12:56</td>
<td>V7 ← <a href="http://ui.constantcontact.com/">http://ui.constantcontact.com/</a>...</td>
</tr>
<tr>
<td>2012-07-19 16:18:18</td>
<td>Phish arrived at V5</td>
</tr>
<tr>
<td>2012-07-19 16:18:53</td>
<td>Phish arrived at V4</td>
</tr>
<tr>
<td>2012-07-19 16:23:40</td>
<td>Phish arrived at V2</td>
</tr>
<tr>
<td>2012-07-19 16:36:11</td>
<td>V2 becomes a victim</td>
</tr>
<tr>
<td>2012-07-19 16:39:16</td>
<td>V3 becomes a victim</td>
</tr>
<tr>
<td>2012-07-19 16:46:52</td>
<td>V4 becomes a victim</td>
</tr>
<tr>
<td>2012-07-19 17:13:02</td>
<td>Phish arrived at V6</td>
</tr>
<tr>
<td>2012-07-19 17:32:48</td>
<td>V5 becomes a victim</td>
</tr>
<tr>
<td>2012-07-19 18:19:15</td>
<td>V6 becomes a victim</td>
</tr>
<tr>
<td>Phish</td>
<td>Time found by intersection</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>PHISH 3</td>
<td>2012-07-15 16:35:19</td>
</tr>
<tr>
<td>PHISH 2</td>
<td>2012-07-16 23:01:02</td>
</tr>
<tr>
<td>PHISH 4</td>
<td>2012-07-17 00:15:27</td>
</tr>
<tr>
<td>PHISH 5</td>
<td>2012-07-17 01:13:40</td>
</tr>
<tr>
<td>PHISH 6</td>
<td>2012-07-18 03:58:25</td>
</tr>
<tr>
<td>PHISH 7</td>
<td>2012-07-18 18:54:24</td>
</tr>
<tr>
<td>PHISH 1</td>
<td>2012-07-19 16:36:11</td>
</tr>
<tr>
<td>PHISH 9</td>
<td>2012-07-20 13:35:24</td>
</tr>
<tr>
<td>PHISH10</td>
<td>2012-07-21 13:17:48</td>
</tr>
</tbody>
</table>
Conclusion

- Dropbox email accounts are a critical but often overlooked component to most successful phishing attacks.
- We have presented low-cost mechanisms to identify dropboxes by combining phishing URL lists with email metadata.
- The techniques could be used to protect users and identify more phishing sites faster.
- We estimate that 120–160 criminals targeted PayPal in July 2012 using 26,900 distinct URLs.
- Increased attention to dropboxes could have a disruptive effect.
- For more: http://lyle.smu.edu/~tylerrm/