Discovering Phishing Dropboxes Using Email Metadata

Tyler Moore
Department of Computer Science and Engineering
Southern Methodist University
Dallas, Texas, 75275, USA
Email: tylerm@smu.edu

Richard Clayton
Computer Laboratory
University of Cambridge
Cambridge, CB3 0FD, UK
Email: richard.clayton@cl.cam.ac.uk

***Abstract*—The criminals who operate phishing scams often deliver harvested credentials to email accounts under their control – but it is difficult, in the general case, to identify these so-called 'dropboxes'. We devise three techniques to identify dropboxes and associated phishing websites by leveraging lists of known phishing websites and metadata maintained by email providers. We demonstrate the techniques' effectiveness using data held by anti-phishing organizations and an email provider. To directly identify dropboxes, we posted fake but distinctive credentials into 170 PayPal phishing pages and inspected an email provider's anti-spam metadata. This metadata recorded the presence of our credentials matching 28 of the phishing pages sending credentials to 17 distinct dropboxes at this particular email provider. We indirectly identified 24 additional dropboxes by searching for email subjects similar to previously-uncovered dropboxes. Based on these findings, we estimate an upper bound of 120 – 160 criminals ran phishing attacks against PayPal in July 2012, a smaller figure than might be expected from the 26 900 PayPal distinct phishing URLs they are known to have employed, spread across 13 018 different hostnames. Finally, in some cases we could extend our metadata processing by running an 'intersection attack'. Whenever victims receive the same URLs as other victims, it is likely that the common URL is for a phishing page. Preliminary evidence suggests that the false positive rate for intersection attacks is low. Furthermore, it can be used to notify impersonated brands immediately after victims disclose their credentials and identify more phishing sites faster than traditional methods currently achieve.**

I. Introduction

Phishing attacks entice victims into disclosing passwords and other credentials to criminal impersonations of genuine websites. The criminals minimize their workload by creating phishing 'kits' which can be used, day after day, in different locations. The contents of the kit may be uploaded onto 'free' webspace or a genuine website that has been compromised by exploiting a security problem.

The kits vary considerably, but there is usually an executable file written in the PHP language, which collects the data entered into the webpage by the victim and packages it up for the criminal. Some kits write the data out into a text file or database on the website. In our experience, though, the most common approach is for the PHP program to send an email containing the data to the criminal. The email address used for this purpose is commonly known as a 'dropbox'.

Since criminals perceive it as risky to use their normal email accounts as dropboxes, they will set up a different account for this purpose – generally, we have found from our inspection of phishing kits, using one of the major 'free' email systems. These email providers will of course shut down any dropboxes they learn about, but in practice they receive few reports. The use of PHP (which is executed 'server side', so the source code cannot be inspected by website visitors) makes it almost impossible for anyone other than the website owner to learn the identity of the dropbox address. Unfortunately, those website owners rarely have the necessary skills and motivation to identify the dropbox and send a report.

The email providers do not want criminals as customers, but the sheer scale of operations – the largest provide service to hundreds of millions of users – necessitates that any abuse detection systems must be efficient, scalable and exhibit vanishingly small false positive rates. In this paper we describe a technique for automatically identifying dropboxes that appears to satisfy these criteria. We apply our technique in experiments undertaken at a particular email provider (which we call WebCo). These experiments demonstrate the feasibility of reliably identifying dropboxes. As an added bonus, we found that we were able to rapidly identify many newly created phishing websites, which the owner of the brand being phished will have an interest in taking down.

Our approach leverages the existing metadata that WebCo collects as part of its day-to-day effort to combat email spam. WebCo operates spam filters that consider the entire content of email, but the metadata does not, for legal reasons, record the content of the email, but is restricted to describing the email and the URLs it contains. The metadata is used by WebCo to tune their spam filtering system and is then deleted. Although the email sent to dropboxes is not spam (the criminal is keen to see it delivered) the metadata describes it sufficiently well to allow dropbox activity to be picked out by some relatively straightforward data processing which, again for legal reasons, had to be carried out for us by a WebCo employee.

The paper is organized as follows. In Section II we discuss the content of phishing kits in more detail and in Section III we describe the metadata collected by WebCo. In Section IV we describe an initial experiment that identifies a small set of dropboxes at WebCo. In Section V we use the metadata to identify more suspect dropboxes. In Section VI we use an innovative backtracking technique to identify the websites

Pre-publication copy, November 2012. To appear in the proceedings of the 7th APWG eCrime Researchers Summit (eCrime)

from which the emails are being sent to the dropboxes. In Section VII we discuss related work before in Section VIII we assess how successful our experiments have been.

II. Phishing Kits

A phishing website must look as much like the site it is impersonating as possible and, when someone is fooled into entering their credentials, it must collect this data and make it available to the criminals. Three main components are therefore required – the HTML page, the graphics (logos, colored bars, images, etc.) and a server-side executable component to handle the credentials. Although some phishing sites use graphics from the original website, savvy brand owners can detect the resulting abnormal traffic patterns and can locate where the phishing website is being hosted. Consequently, most phishing sites use local copies of graphics – meaning that even the display of a single page can require the presence of a dozen or more distinct files on the server.

Criminals rapidly discovered that it was simple to package up all of the files that were needed into an archive format (usually a ZIP file), upload that single file, and then unpack (unzip) the archive to create the phishing site. Kits are sold on the 'underground economy' but even as early as 2004, Sophos was reporting that 'free' versions were available for download [5]. In 2008, Cova et al. investigated kits and found that many of the 'free' kits contained back doors that would deliver credentials to the kit creator as well as to the criminal that had deployed it [1].

A typical PHP executable file will appear, in its most stripped down form, something like this:

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

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

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

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

The first few lines of this PHP code record the login credentials along with the IP address of the victim (which the criminal will use to spot people who enter fake information in bulk). The destination email address and Subject line are then set and the email is dispatched. The error handling is mainly there for the criminal to assess whether or not the environment on the server is conducive to executing this code.

Kits sometimes generate text files on the server, but this has become far less common in the past few years. Instead, the usual modus operandi is to send email to one or more dropbox accounts hosted at 'free' email providers, as illustrated in Figure 1. Occasionally a kit will place credentials into a database, and display the results as a table. But we seldom see this level of sophistication and, since it does not especially improve criminal productivity, it is likely to remain a rarity.

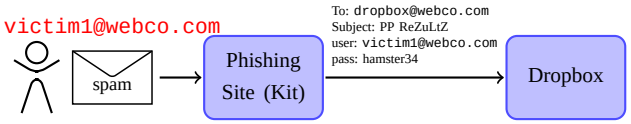


Fig. 1: How phishing kits use dropboxes to harvest credentials.

The executable PHP file is invariably included in the ZIP archive, and it is very seldom edited on the server itself. It is inconvenient for the criminal to alter the Subject line or the dropbox address as they move their phishing activity from server to server. As this paper will demonstrate, this inconvenience is sufficient to ensure that the majority of criminals leave Subject lines unchanged and only move dropbox when the old one is closed down by the email provider. In turn, this allows us to link behavior over time.

III. WebCo's metadata for incoming email

For users of the WebCo email system, legitimate incoming email will appear in their inbox while spam is placed into a separate folder, or in the most egregious cases, rejected outright. As each piece of email is handled, metadata is collected and then stored, with one file entry per email. This data is subsequently used to drive the feedback loops that ensure that email spam is correctly detected and handled. Each metadata entry can have 100 or more fields, chronicling the handling of the email by several different systems, however only a handful of these fields are relevant to this paper:

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

Source IP address
: The machine that sent the email to WebCo.

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

SMTP "mail from"
: The email sender, from the SMTP conversation. This can be forged, but for email sent by phishing kits it usually indicates the true origin.

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

Subject
: This is the 'Subject:' email header field. This is invariably set by the phishing kit.

URLs
: These are the URLs from the body of the email.

The URLs are recorded in the metadata because URLs are a very distinctive way of identifying spam. Of particular relevance to our work is that email addresses (in practice any plausible string including a `@` symbol) are treated as if they were `mailto://` URLs. Consequently, any email address in the message body will be recorded as part of the metadata.

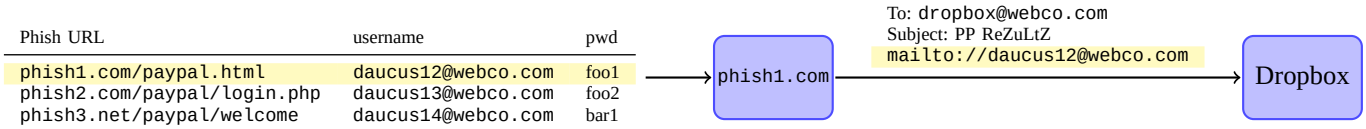


Fig. 2: Technique for directly identifying dropboxes. Bogus submitted credentials are located in email metadata to reveal the dropbox account associated with the phishing website.

IV. Direct Identification of Dropboxes at WebCo

We now explain a number of techniques for identifying dropboxes by combining data on phishing URLs with metadata from WebCo. Essentially, in each case we manually join disparate data sets to uncover additional relevant data.

A. Finding Dropboxes

The treatment of email addresses as URLs and their consequent recording in WebCo's metadata means that there is a simple way to identify dropboxes. If we visit a phishing website and enter a distinctive email address into the form, then we can inspect the metadata held by WebCo to check for a record of our unique email address. Any incoming email that includes our planted email address will be being delivered to a dropbox account.

For the technique to work, we must first identify a reliable source of phishing URLs. As a part of our general work on phishing, we receive 'feeds' of known phishing websites from a brand owner, from commercial brand protection companies, from the public domain repository PhishTank[1] and from the Anti-Phishing Working Group (APWG).[2] We regularly monitor whether these websites are still active.

On 1 June 2012 we compiled a list of 170 different phishing websites that impersonated PayPal (a well-known payment services provider) that were first reported on or after 1 May 2012 and that were still operational. We visited each of these sites and entered a non-existent WebCo email address and a specious password. The email addresses were of the form `daucus999999@webco.com` with the `999999` value different in each case to allow us to correlate results.[3] The process is demonstrated visually in Figure 2.

Inspection of the WebCo metadata logs showed that 28 emails containing our distinctive email addresses had been delivered to 17 distinct `@webco.com` mailboxes (in one case two of these dropboxes were being used in parallel and so the contents of one dropbox was a subset of another).

We were not overly surprised that only find that only 16.4% of the PayPal phishing sites were using WebCo dropboxes. Anecdotal evidence has long suggested that well over three quarters of all dropboxes are hosted by a particular free email provider – although we are unaware of any convincing explanation for why the criminals have this marked preference.

We also found that, as we had expected, the criminals used some very distinctive Subject header fields for their emails and they were not altering these Subjects before deploying their phishing kit in another location.

Some example (albeit slightly anonymized) Subjects were:

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

B. Measuring PayPal Phishing Activity

Having identified the dropboxes, we can examine additional metadata associated with the dropboxes. In particular, we can leverage the metadata to estimate how successful the criminals were in capturing credentials by looking for victim email addresses appearing in dropboxes.

Some mechanics of email dropboxes bear mentioning at this point. First, it is typical for every submission of the form on the phishing page to cause a new email to be sent to the dropbox address. Hence, if the metadata shows that a mailto: URL was included, then the phishing page has claimed another victim. Consequently, we counted the number of emails received by the dropboxes where the Subject header field matched the 'daucus' email and a mailto: URL was present.[4] We discuss below the adjustments that should be made for test emails and for intentionally fake entries, but it is certainly the case that this count of emails serves as an upper bound for the number of victim credentials delivered to each dropbox.

A second relevant aspect of dropboxes is that the email sender is usually unique to each phishing URL. Although some of the criminals customized the From: header field of their email, they did not fully control the mechanics of sending the email and so the sender (the value in the MAIL FROM of the SMTP protocol) will differ for each new phishing URL (except perhaps for multiple pages on the same server, or multiple servers at the same hosting company). Therefore, we also counted the number of distinct email senders, since this gave us a good approximation for the number of distinct phishing sites that were sending credentials to a given dropbox.

Table I shows our results for the three month period 1 May to 31 July 2012. Since our initial experiment was performed on

[1] http://www.phishtank.com

[2] http://www.apwg.org

[3] We wanted our visits to the phishing websites to appear normal, without our tracking email addresses clashing with any real accounts. We used the slightly unusual word daucus, the genus of which carrots are a species, to commemorate the disinformation campaign run in the Second World War by the UK Air Ministry. Not wishing to divulge that the British had developed an airborne radar system, they attributed their night fighters' successes to the consumption of carrots.

[4] We are measuring PayPal phishing here, so we expect the 'login name' to always be an email address.

1 June there is an obvious bias to these results, so it is almost meaningless to divide the counts by the number of days.

TABLE I: Statistics for PayPal related phishing activity in the 17 PayPal phishing dropboxes initially identified at WebCo.

	emails (victims)	senders (sites)	ratio ($\frac{\text{victims}}{\text{sites}}$)
dropbox1	1 470	2	735.0
dropbox2	925	66	14.0
dropbox3	895	16	55.9
dropbox4	695	109	6.4
dropbox5	642	17	37.8
dropbox6	615	72	8.5
dropbox7	384	8	48.0
dropbox8	274	6	45.7
dropbox9	177	44	4.0
dropbox10	158	1	158.0
dropbox11	106	5	21.2
dropbox12	39	3	13.0
dropbox13	37	12	3.1
dropbox14	12	1	12.0
dropbox15	18	5	3.6
dropbox16	9	3	3.0
dropbox17	6	3	2.0
mean	380	22	68.9
median	177	6	13

These figures need to be read with caution – dropbox9 was sent a subset of the email that was delivered to dropbox4; and dropbox1 mainly received email forwarded to WebCo from another email provider – thereby obscuring the true number of senders. With these qualifications in mind, we can make a number of interesting observations. First, we note that the distribution of victims per dropbox is markedly skewed, ranging from a handful of victims to nearly 1 500. The median number of victims per dropbox is 177. Second, we also see wide variation in the number of phishing sites associated with each dropbox. This suggests that there is heterogeneity to the approach taken by criminals.

We also show in the table the ratio of the number of sets of credentials collected per site. Note that the ratio values cannot be directly compared with each other because we have no way of knowing how many emails were sent in each phishing campaign, neither have we factored in the lifetime of the various websites. Nevertheless, the figures do give a feel for the level of activity which is broadly in line with previous results – there are occasional outliers but most sites claim less than 50 victims.

As we indicated above, these figures measure the upper bound of the criminals' success because the metadata logs do not reveal the exact email content. We cannot know for certain whether an email that contains a mailto: URL also contains the corresponding password, or whether the intended victim was aware of the scam and has typed, for example, `die spammer die` into that field.

However, a couple of the kits used by the criminals placed both the email address and the password into the email Subject header field – doubtless this improves their efficiency at sorting the contents of their mailbox. These Subjects are recorded in the metadata, and so we examined them manually to assess whether the email addresses and passwords looked plausible. There were 48 instances of this type of Subject in our dataset, of which 6 looked as if they were the criminals testing out a new phishing installation, and 5 were clearly not valid. The remaining 37 examples of email/password pairs looked entirely real. Acknowledging that we are extrapolating from a very small sample, we tentatively conclude that around three-quarters of the emails we counted contained valid credentials.

C. Measuring Other Phishing Activity

Examining the metadata for the dropboxes we found that some of them were being used to receive credentials for other brands than just PayPal. The Subject header fields indicated that attacks had been mounted on AOL, Gmail, Hotmail, Yahoo!, Alibaba, Bank of America, Bankwest, Barclays, CartaSi, Chase, Nationwide and Visa. We identified the format of the relevant email Subjects and repeated the analysis that we described in Section IV-B above.

However, because many of these attacks do not use email addresses as credentials (which we can detect in the metadata and so exclude email lacking credentials) our results are very likely to be inflated by test traffic, people who type insults into the webpage, or visitors who submit the form without entering any information. We were, however, able to exclude email from phishing sites that generated an email for a mere visit, rather than only sending an email once credentials were typed in.[5] These caveats aside, the results show that further non-trivial levels of activity are being detected in 13 of the 17 dropboxes, as we set out in Table II:

TABLE II: Statistics for non-PayPal related phishing activity in the phishing dropboxes initially identified at WebCo.

	emails (victims)	senders (sites)	ratio ($\frac{\text{victims}}{\text{sites}}$)
dropbox1	1 389	2	694.5
dropbox3	219	16	13.7
dropbox4	1 985	169	11.7
dropbox6	590	85	6.9
dropbox7	194	9	21.6
dropbox8	278	4	69.5
dropbox9	324	65	5.0
dropbox10	44	1	44.0
dropbox11	56	6	9.3
dropbox13	52	3	17.3
dropbox14	5	1	5.0
dropbox15	474	19	24.9
dropbox17	15 254	32	476.7
mean	1 604	32	107.7
median	278	9	17.3

[5]We detected this type of site by observing multiple occurrences of our own IP addresses in the data, caused by our automated system that monitors phishing website longevity – we know that our system never enters credentials, so the emails must have been generated merely because the site was visited.

V. Indirect identification of dropboxes at WebCo

In the previous section we presented a technique for directly identifying dropboxes that relied on active measurement. We first had to acquire a list of known phishing websites, and then we had to transmit fake credentials into the corresponding web forms. Having completed that process, we can use what we have learned from these dropboxes to identify additional dropbox accounts even if we are unaware of the phishing website associated with the dropbox.

We indirectly identify further dropboxes at WebCo by looking to see whether the distinctive Subjects were being received by other mailboxes. These would indicate that the same criminals had set up more than one dropbox, or that other criminals were using the same phishing kit.

The original Subject header fields that appeared in the 28 'daucus' emails boiled down to 15 distinct patterns (i.e. there were at least 15 different kits in use). Looking for other recipients of email where the Subject matched these patterns yielded 81 new dropboxes. This is three times as many dropboxes as we found using the direct approach!

Why do we care about identifying more dropboxes? As well as finding more victims, if we can be comprehensive in collecting dropboxes, then we can use this total to approximate the number of criminals actively engaging in phishing.

It would be wrong to directly extrapolate from the number of dropboxes to the number of criminals – as we have already seen, some of them use multiple dropboxes in parallel. There is also evidence of serial usage. For example we saw `****full123@webco.com` used from 2012-07-09 to 2012-07-10 and `****full121@webco.com` used from 2012-07-13 to 2012-07-30; similarly we saw `*****tat01@webco.com` used from 2012-05-31 to 2012-06-01 and `*****tat001@webco.com` used from 2012-06-01 to 2012-07-29.

Nevertheless, we can roughly estimate an upper bound for the number of criminals attacking PayPal. We identified the dropboxes used for attacks on PayPal (ignoring the attacks on other brands) and found that there were 29 in use during July (ignoring parallel deliveries). Of these 17 were used throughout the month and 12 for shorter periods that overlapped slightly. We therefore estimate that 20 – 29 different criminals were using WebCo for dropboxes. Since our initial experiment described in Section IV showed that 16.4% of dropboxes were at WebCo we can therefore scale our count up and estimate the number of criminals attacking PayPal in July 2012 to be in the region of 122 – 164.

Our estimate is based on several suppositions. We assume that we have identified all the relevant Subject header fields and hence that we have identified all the dropboxes receiving PayPal credentials. If we have omitted any dropboxes then the number of criminals will be higher than our estimate. We assume that we have corrected identified whenever multiple dropboxes are used by the same criminal. If we missed any instances then the number of criminals will be lower than our estimate. We are also making the rather sweeping assumption that there is nothing particularly different about the one sixth of the dropbox traffic we have data for, compared with the five sixths which is going elsewhere – this could affect our estimate of the number of criminals quite substantially.

However, there is another way of analyzing our data which helps to underpin our result. We can count the total number of phishing sites and the number of different sources the dropbox email arrived from.

During July 2012 our phishing feed analysis showed that there were approximately 26 900 PayPal phishing URLs – but this figure is misleadingly high. Many superficially different URLs lead to the same webpage and in some cases multiple phishing pages were set up, presumably by just one criminal, on the same machine. To try and eliminate this over-counting we extracted the 13 018 hostnames from the URLs and resolved them to 2 383 different IP addresses.

Examining the dropbox email for the dropboxes active in July we find that PayPal credentials arrived from 274 different IP addresses – so the ratio when we analyze the data in this manner is that 11.5% of PayPal websites used a dropbox at WebCo. This number is the same order of magnitude as our earlier 16.4% value, which helps give credence to both. We believe that the lower percentage we obtain from this calculation results from groups of machines sending their email through a single email server.

VI. Identifying the source of dropbox email

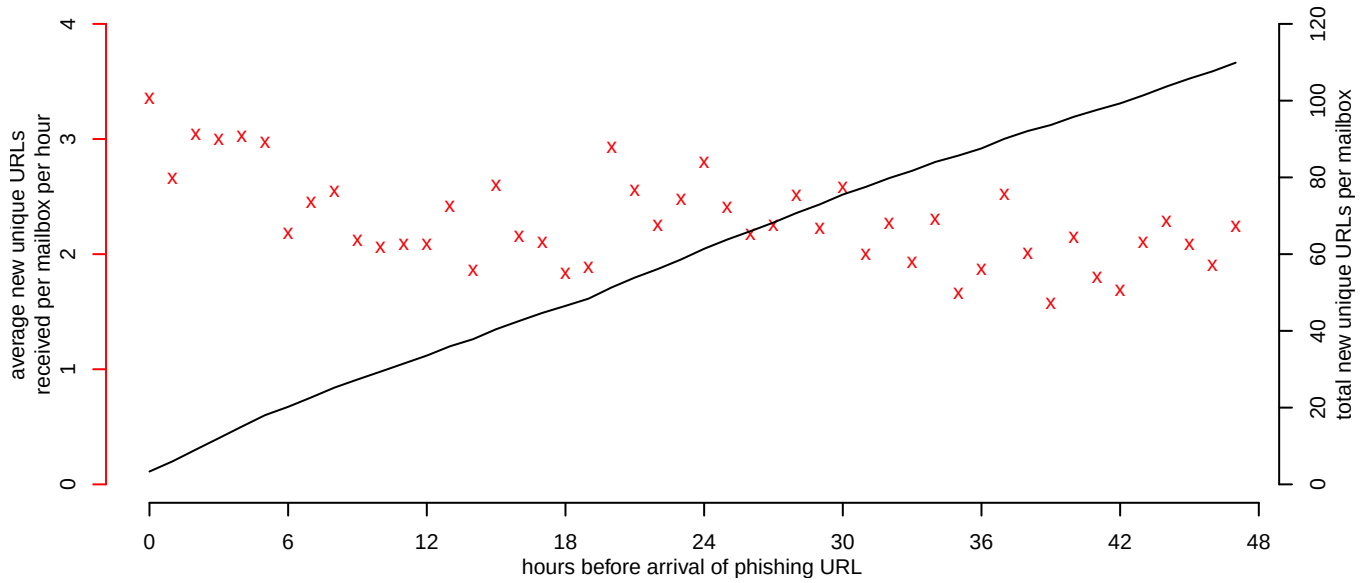
In Section V above we described a technique for using Subject header fields to identify additional dropboxes without knowing anything about the associated phishing sites. Having obtained this extended list of known dropboxes, it would be useful to determine what the corresponding phishing sites are. Unfortunately, it is far from straightforward to identify the phishing site URL even when we have data about the emails that are delivered from it to a dropbox.

Of course we know which IP address the email came from, but that may be a hosting company's mail server rather than the individual machine that created the email. We may sometimes see a distinctive SMTP sender field (such as `username@ns35.hostingcompany.net`) but at best this will identify the hosting account and perhaps then the website; but even if we learn the website's identity we will not obtain a URL for the phishing page itself.

Fortunately, we were able to identify the phishing URLs associated with dropbox emails by using an 'intersection attack'. Intersection attacks can be used to identify users of anonymity systems [3], but we do not need complex statistical tools here. Rather, we will determine the WebCo users whose identity appears in a dropbox (in a mailto: URL) and collate a list of URLs that each has recently received in their email.

The intersection of these lists will contain the phishing URLs. The number of false positives within the intersection, universally received spam URLs for example, will determine how practical this approach will be.

Fig. 3: Count of previously unseen URLs per hour (left hand axis) and the cumulative total (right hand axis). Values are averaged across all 159 phishing victims, with hour 0 being the time at which they visited a phishing website, and time going back into the past before that.



A. The Principle of an Intersection Attack

A simple example will illustrate the method. We look at the incoming email to dropbox8 and find a series of emails around the 8 June which the Subject indicates will have contained PayPal credentials and which came from a specific IP address (209.160.28.xxx, a hosting company in Seattle WA, USA). We select two of these emails which are evidence of when two WebCo users visited the phishing website (the times are in GMT, but the users are US based, so they were visiting the website during the afternoon of succeeding days):

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

We then look at all incoming email to `dest1@webco.com` and `dest2@webco.com` which arrived in the 24 hours leading up to their respective visits to the website and determine which URLs they received in common.

There turned out to be 23 such URLs, but 22 are URLs for well-known sites such as match.com, macys.com, amazon.com etc. which can immediately be discounted. This left URL: http://surses-paypal.com-confirm-cgi.bin.acoount-15f2vb1n.save-data-supportteam1651sd1d45hfdcfgg478521fdsd5ds1d6.dnstour.com/Uid=9863528034/:

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

This rather fishy/phishy looking URL was received by both accounts in an email with the Subject header field: `update your PayPal account information`, and hence we have identified the phishing website that they visited, and could in principle have done so shortly after 9pm on 8 June. In this instance, the URL was reported to the APWG phishing feed shortly after 2am on 10 June – 29 hours after this.

In practice more than a dozen victims were snagged by this URL rather than just the two we have considered. Processing the extra data in the dropbox would not have sped up identification in this case, but it would have reduced the number of false positive URLs – we would not have had to exclude the legitimate sites that were ‘above suspicion’.

Unfortunately, the intersection attack is not universally applicable. In particular, the stolen credentials must include email addresses (mailto: URLs). Consequently, it can be expected to work for attacks on PayPal but will generally fail when a bank is phished because the ‘login’ is typically an account number.

B. Intersection Attacks using a Week of Data

To show how effective intersection attacks might be in practice we considered email that arrived at WebCo during a one week period starting on 15 July 2012. This start date was chosen arbitrarily, but we used a full week’s worth of data to ensure a reasonable sized sample.

We looked for Subject header fields that matched our expanded criteria as discussed in Section V above and then checked for the presence of a mailto: URL. We believe these emails to be phishing emails and the mailto: URL to identify the victim (although the metadata does not allow us to know whether the credentials provided were valid). This process yielded the metadata for 934 emails which had been sent to WebCo from 114 distinct IP addresses.

Closer inspection showed that there are only 159 victims with a WebCo email address (the only company for which we had metadata) and the emails containing their credentials came from just 47 of the IP addresses. Of these 47, there was only a single WebCo victim for 25 of the IP addresses, meaning that

we could run the intersection attack – which requires there to be two or more victims – and learn at most 22 URLs.

Before describing our results, we will make a small digression and consider the risk of false positives.

We measured how many URLs each of the 159 victims received in their incoming email in the time period immediately before they entered their credentials onto the phishing site.[6] If the number of URLs was generally very high then we might expect poor results – because the victims were all receiving a lot of spam, which would have URLs in common.

To perform this analysis we worked backwards in time, hour by hour, counting unique URLs – if the victim received example1.com 30 minutes before they visited the phishing site we counted 1 in hour 0; if they received example2.com 70 minutes earlier we counted 1 in hour 1; but we only counted each URL once, in the hour nearest in time to when the victim visited the phishing site. We plot the results in Figure 3 and see that on average these victims received about 2 new unique URLs per hour and the rate of arrival is fairly constant. That is, these victims regularly see new URLs but not very many of them, so there is considerable hope of avoiding large numbers of false positives in our intersection attack.

C. Results of Intersection Attack

We grouped our data by each of the 22 sending IP addresses and used the intersection attack, generalized for n victims, in an attempt to identify the phishing URLs. For the first URL, P1, the timeline for the URLs received in email by each of six victims (V1 ... V6) was:

2012-07-19 15:16:22	phish arrived at V1
2012-07-19 15:20:02	phish arrived at V3
2012-07-19 15:21:32	V1 becomes a victim
2012-07-19 15:48:30	V6←http://77kids.com etc.
2012-07-19 16:16:18	phish arrived at V5
2012-07-19 16:18:53	phish arrived at V4
2012-07-19 16:23:40	phish arrived at V2
2012-07-19 16:36:11	V2 becomes a victim
2012-07-19 16:37:25	V6←http://www.constantcontact.com
2012-07-19 16:39:16	V3 becomes a victim
2012-07-19 16:46:52	V4 becomes a victim
2012-07-19 17:13:02	phish arrived at V6
2012-07-19 17:32:48	V5 becomes a victim
2012-07-19 18:19:15	V6 becomes a victim

Only V6 received any other email containing URLs during the period of interest making it particularly easy to identify the phish – the URL that all victims receive. In fact, by 16:36:11 when V2 is phished, the phishing URL can be identified with no alternative candidates to consider.

In practice, it is only necessary to consider the URLs seen by the first two victims and discard any URLs that they did not both receive. When we do this for the ten other phish that can be identified there were no other URLs to worry about – the phish is the only URL they receive in common. The full results are in Table III.

[6]In our experience, most currently deployed phishing kits will generate an email as soon as a victim enters their credentials, and that email will be delivered almost immediately, so the metadata timestamp can be considered to accurately reflect the time of the victim's actions.

TABLE III: Timelines of all successful intersection attacks (showing only the events relevant to the first two victims).

PHISH 2	2012-07-14 10:39:07	phish arrived at V1
PHISH 2	2012-07-14 16:42:53	phish arrived at V2
PHISH 2	2012-07-16 21:08:14	V1 becomes a victim
PHISH 2	2012-07-16 23:01:02	V2 becomes a victim
PHISH 3	2012-07-15 15:17:08	phish arrived at V2
PHISH 3	2012-07-15 15:44:01	phish arrived at V1
PHISH 3	2012-07-15 15:51:15	V1 becomes a victim
PHISH 3	2012-07-15 16:35:19	V2 becomes a victim
PHISH 4	2012-07-16 22:57:01	phish arrived at V2
PHISH 4	2012-07-16 23:08:00	phish arrived at V1
PHISH 4	2012-07-16 23:24:04	V1 becomes a victim
PHISH 4	2012-07-17 00:15:27	V2 becomes a victim
PHISH 5	2012-07-16 23:29:02	phish arrived at V2
PHISH 5	2012-07-16 23:29:02	phish arrived at V1
PHISH 5	2012-07-17 00:26:53	V1 becomes a victim
PHISH 5	2012-07-17 01:13:40	V2 becomes a victim
PHISH 6	2012-07-18 01:54:08	phish arrived at V2
PHISH 6	2012-07-18 03:37:46	phish arrived at V1
PHISH 6	2012-07-18 03:53:26	V1 becomes a victim
PHISH 6	2012-07-18 03:58:25	V2 becomes a victim
PHISH 7	2012-07-18 17:36:38	phish arrived at V2
PHISH 7	2012-07-18 17:57:06	phish arrived at V1
PHISH 7	2012-07-18 18:07:48	V1 becomes a victim
PHISH 7	2012-07-18 18:54:24	V2 becomes a victim
PHISH 8	2012-07-18 22:02:32	phish arrived at V1
PHISH 8	2012-07-18 23:11:40	phish arrived at V2
PHISH 8	2012-07-19 02:26:17	V1 becomes a victim
PHISH 8	2012-07-19 04:49:26	V2 becomes a victim
PHISH 9	2012-07-20 13:09:56	phish arrived at V2
PHISH 9	2012-07-20 13:10:55	phish arrived at V1
PHISH 9	2012-07-20 13:21:51	V1 becomes a victim
PHISH 9	2012-07-20 13:35:24	V2 becomes a victim
PHISH10	2012-07-21 02:58:21	phish arrived at V1
PHISH10	2012-07-21 03:01:17	V1 becomes a victim
PHISH10	2012-07-21 13:04:41	phish arrived at V2
PHISH10	2012-07-21 13:17:48	V2 becomes a victim
PHISH11	2012-07-21 23:42:03	phish arrived at V2
PHISH11	2012-07-21 23:44:25	phish arrived at V1
PHISH11	2012-07-22 01:56:49	V1 becomes a victim
PHISH11	2012-07-22 05:20:09	V2 becomes a victim

Unfortunately in 11 cases none of the victims received the same URL as any other and our attack failed. We do not have a definitive explanation for this failure. It may be that the traffic coming into the dropboxes was created when people filling in forms that were embedded into emails as attachments – our metadata processing does not identify these emails because the URL for the POST command embedded into the form is not recorded by WebCo at the present time.

Nevertheless, our intersection attack can successfully identify eleven phishing URLs and in five cases as Table IV shows, this identification occurred before they turned up in any of our phishing feeds (and P1 was never listed there at all):

TABLE IV: Time when phishing URL can first be identified by our intersection attack and in our 'feeds' of phishing URLs. The lag value shows how much earlier our attack detected some of the URLs.

	by intersection attack	in phishing feed	lag
P3	2012-07-15 16:35:19	2012-07-02 21:27:12	–
P2	2012-07-16 23:01:02	2012-07-17 02:18:15	3.2 hours
P4	2012-07-17 00:15:27	2012-07-21 11:13:06	4.5 days
P5	2012-07-17 01:13:40	2012-07-15 15:10:07	–
P6	2012-07-18 03:58:25	2012-07-18 06:21:28	2.5 hours
P7	2012-07-18 18:54:24	2012-07-23 14:18:38	4.8 days
P8	2012-07-19 04:49:26	2012-05-16 18:37:49	–
P1	2012-07-19 16:36:11	never reported	∞
P9	2012-07-20 13:35:24	2012-07-17 20:11:35	–
P10	2012-07-21 13:17:48	2012-07-18 00:05:03	–
P11	2012-07-22 05:20:09	2012-07-20 14:28:44	–

VII. Related work

There has been considerable empirical research investigating the nature of phishing attacks. In one study, Moore and Clayton estimated that between 280 000 and 560 000 people gave away their credentials to phishing websites each year [8]. They also found extensive concentration in attacks – around half of all phishing scams they studied had carried out by a single gang. In a separate study, Florêncio and Herley studied when passwords were entered at unexpected websites and estimated that 0.4% of the Internet population is phished annually [4].

Some studies have examined multiple data sources in order to get a better handle on the extent of phishing, as we have done in this paper with phishing URLs and email metadata. Moore and Clayton examined multiple feeds of phishing URLs, finding that the lists maintained by different take-down companies are substantially incomplete [8]. Weaver and Collins examined two sources of phishing URLs in order to estimate the true extent of phishing websites using capture-recapture methods borrowed from experimental ecology [10]. Moore, Clayton and Stern linked phishing URLs with a large source of email spam maintained by an anti-spam vendor [7]. They examined temporal correlations between the time that spam was sent and when the URLs were identified by phishing URL feeds.

A few studies have looked at kits in greater detail. Wardman and Warner gathered a number of phishing kits in order to automatically identify phishing websites based on the kit characteristics [9]. Cova et al. investigated kits and found that many of the 'free' kits contained back doors that would deliver credentials to the kit creator as well as to the criminal that had deployed it [1]. McCalley, Wardman and Warner also examined back-doored kits, detailing the obfuscation techniques deployed by criminals [6].

The present work builds on the prior literature by providing a means of automatically identifying dropbox email addresses and reporting on their prevalence.

VIII. Concluding remarks

Dropbox email accounts are a critical but often overlooked component of most successful phishing attacks. They serve as a transient repository of stolen credentials, offering criminals easy access to credentials immediately after they are entered by victims on phishing websites.

In this paper, we describe a series of mechanisms for identifying dropboxes and associated criminal data by combining phishing URL lists with metadata maintained by email operators. In particular, we devise techniques to

- reliably identify dropboxes,
- find additional victims from dropbox contents in a timely fashion that could potentially identify victims early enough to successfully block exploitation,
- find additional dropboxes by searching for conspicuous Subject header fields, and
- identify more phishing webpages by comparing URL metadata for multiple victims clustered in time.

We demonstrate the feasibility of these techniques by applying them to a large, frequent, target of phishing (PayPal) using metadata from a particular email provider. This also gives us the opportunity to report summary statistics about the incidence of dropboxes and the number of victims encountered. We found 29 dropbox addresses that were used to receive PayPal credentials during July 2012, and we estimate that, across all email operators, roughly 120 – 160 criminals were maintaining dropboxes. While still a substantial figure, it gives a much more realistic view of the size of the problem to be tackled than by just considering the 26 900 distinct PayPal phishing URLs observed over the same period.

Consistent with prior empirical research on phishing, we observe skewed distributions in the number of victims attracted to each dropbox. As many as 1 470 victim's credentials were observed to be delivered to a single dropbox address, but the median delivery number was a more modest 177 victims.

We believe that increased attention to dropboxes by those defending against phishing could yield substantial additional insights and strategic advantage over criminals. Of course, applying pressure to any aspect of the criminal infrastructure that has thus far eluded attention might trigger attacker adaption. Nonetheless, we hope that the low-cost techniques presented here might be incorporated into the phishing countermeasures adopted by all email operators.

Acknowledgment

The authors would like to thank 'WebCo' and their employees for their extremely helpful co-operation in allowing their data to be mined so as to throw further light on phishing activity.

REFERENCES

[1] Marco Cova, Christopher Kruegel, and Giovanni Vigna. There is no free phish: an analysis of "free" and live phishing kits. In *Proceedings of the 2nd conference on USENIX Workshop on offensive technologies*, WOOT'08, pages 4:1–4:8, Berkeley, CA, USA, 2008. USENIX Association.

[2] Lorrie Faith Cranor, editor. *Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit 2007, Pittsburgh, Pennsylvania, USA, October 4–5, 2007*, volume 269 of *ACM International Conference Proceeding Series*. ACM, 2007.

[3] George Danezis and Andrei Serjantov. Statistical disclosure or intersection attacks on anonymity systems. In *in Proceedings of 6th Information Hiding Workshop (IH 2004*, pages 293–308, 2004.

[4] Dinei Florêncio and Cormac Herley. Evaluating a Trial Deployment of Password Re-Use for Phishing Prevention. In Cranor [2], pages 26–36.

[5] Sophos Inc. Do-it-yourself phishing kits found on the internet, reveals sophos. http://www.sophos.com/en-us/press-office/press-releases/2004/08/sa_diyphishing.aspx, 2004.

[6] Heather McCalley, Brad Wardman, and Gary Warner. Analysis of backdoored phishing kits. In Gilbert L. Peterson and Sujeet Shenoi, editors, *IFIP Int. Conf. Digital Forensics*, volume 361 of *IFIP Advances in Information and Communication Technology*, pages 155–168. Springer, 2011.

[7] T. Moore, R. Clayton, and H. Stern. Temporal correlations between spam and phishing websites. In *2nd USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET '09)*, Boston, MA, April 2009.

[8] Tyler Moore and Richard Clayton. Examining the impact of website take-down on phishing. In Cranor [2], pages 1–13.

[9] B. Wardman and G. Warner. Automating phishing website identification through deep md5 matching. In *eCrime Researchers Summit, 2008*, pages 1 –7, oct. 2008.

[10] R. Weaver and M.P. Collins. Fishing for phishes: applying capture-recapture methods to estimate phishing populations. In *anti-phishing working groups 2nd annual eCrime researchers summit*, eCrime '07, pages 14–25, New York, NY, 2007. ACM.