What matters in URLs?

Richard Clayton

SHB
17th June 2011

UNIVERSITY OF
CAMBRIDGE
Computer Laboratory

NPL
National Physical Laboratory

Phishing URLs (barclays is just an e.g.!)

1. www.barklays.com/login.html
2. www.barclays.com.account.1234567.kjakjas.info/login.html
3. www.kjakjas.info/www.barclays.com/login.html
4. www.kjakjas.info/~user/www.barclays.com/login.html
5. www.kjakjas.info/joomla/images/www.barclays.com/login.html
6. www.barclays.com.verysecure.com/login.html

Some special (ancient) cases

- http://www.barclays.com:security@www.kjakjas.info/login.html
  - disallowed by Microsoft (for HTTP) in Feb 2004

- http://www.barclays.com. ⌂⌂kjakjas.info/login.html
  - changes made to browser display c 2005

Does the bank name matter?

- Can be trivially obscured:

```
<a href="http://www.example.com">www.barclays.com</a>
```

- Clearly the continued use of the bankname is thought to be useful – but it's hard to measure, the widespread use of "kits" means that the kit builder makes the decision for the phisher
- One datapoint is that online game phishing is heavily domain name based:

eu-batt1e-gm-wow.com, eu-batt1e-gm-wow.net, eu-batt1e-gmwow.com, eu-batt1e-gmwow.net, eu-batt1e-wow-gm.net, eu-batt1e-wowgm.com, eu-batt1e-wowgm.net, eu-battle-bizzgm.com, eu-battle-bizzgm.net, eu-battle-eugm.net, eu-battle-wowgm.com, eu-battlegm-wow.com, eu-battlegm-wow.net, eu-battlegm-wow.org

Instant messaging worms

- Your IM “buddy” sends you a message

`Foto ☺ http://www.facebookj.com/album.php?richard@example.com`

- You click on message and you are one click on “OK” away from being infected with the malware
- If you are infected then you will IM all of your buddies with the message (and their email addresses – extracted from your IM client) and so the malware spreads

Behind the scenes

- The malware resolves a hostname
- The DNS “A record” gives the IP address of an IRC server
- When the IRC server broadcasts a “channel topic” this causes all infected machines to send out an IM to buddies
- There is a distinct pattern to the hostnames used for the malware:

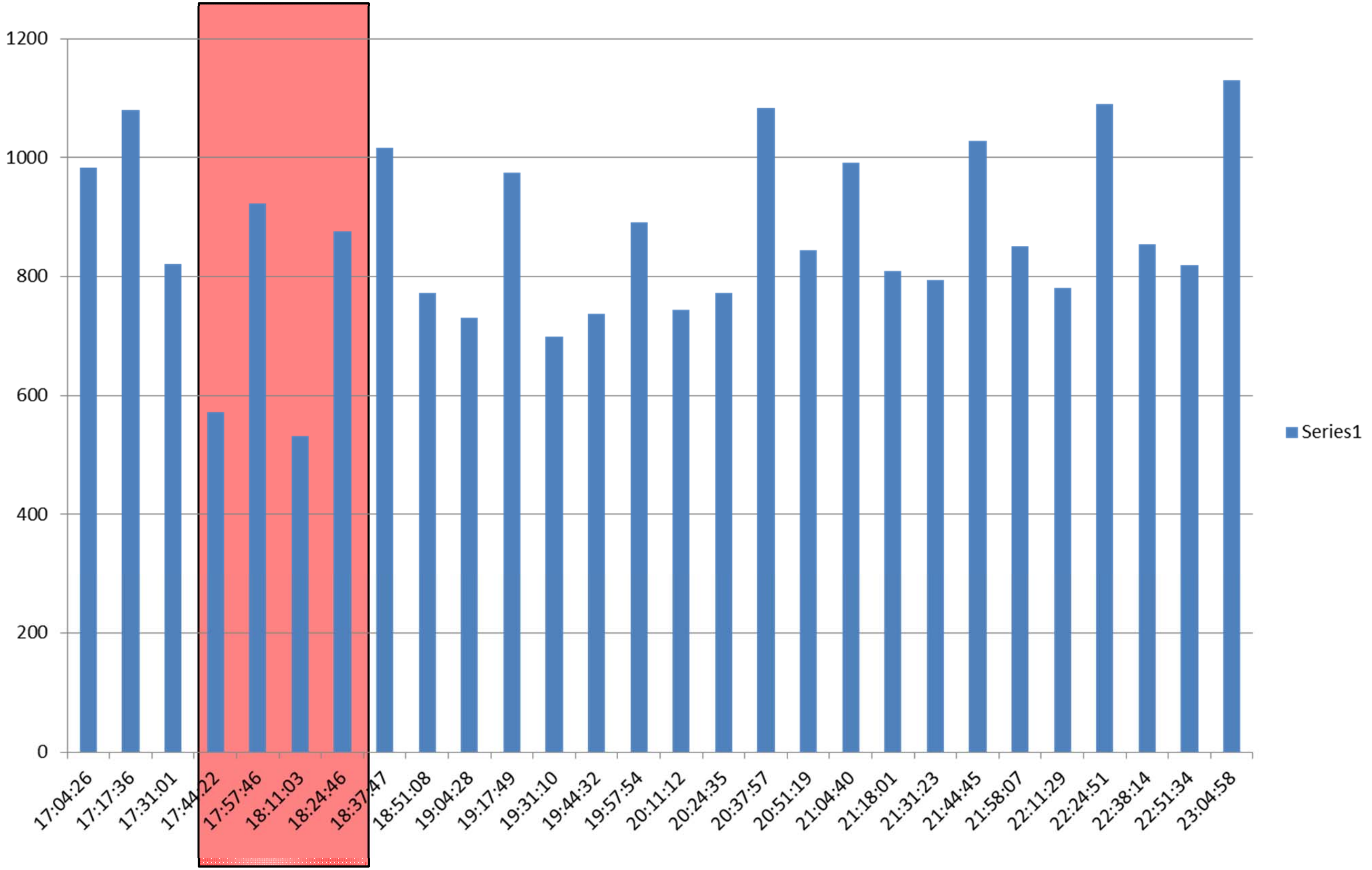
**minefacebok.com, msg-facebook.com, my-faceblook.com,
my-facebookli.com, my-facebookzk.net, my-facefoto.com,
myfacebloofghs.net, myfacebo0k.com, myfacebook12.com**

**myspace81.net, myspacebookghi.com, myspacebooks.net,
myspacebooksx.com, myspacefic.com, myspacegibso.com,
myspacegisp.net, myspacekodegks.net, myspacelootsi.com**

But sometimes URL shorteners are used

```
2011-02-17 17:04:26 is this you on pic? http://kunfacebook.net/album.php?=
2011-02-17 17:17:36 is this you on pic? http://kunfacebook.net/album.php?=
2011-02-17 17:31:01 is this you? http://kunfacebook.net/album.php?=
2011-02-17 17:44:22 is this you? http://linkmenow.org/images555?=
2011-02-17 17:57:46 is this you? http://linkmenow.org/images555?=
2011-02-17 18:11:03 is this you? http://linkmenow.org/images555?=
2011-02-17 18:24:46 is this you? http://linkmenow.org/images555?=
2011-02-17 18:37:47 is this you? http://kunfacebook.net/album.php?=
2011-02-17 18:51:08 is this you? http://kunfacebook.net/album.php?=
2011-02-17 19:04:28 is this you? http://kunfacebook.net/album.php?=
2011-02-17 19:17:49 is this you? http://kunfacebook.net/album.php?=
2011-02-17 19:31:10 is this you? http://kunfacebook.net/album.php?=
2011-02-17 19:44:32 is this you? http://kunfacebook.net/album.php?=
2011-02-17 19:57:54 is this you? http://kunfacebook.net/album.php?=
2011-02-17 20:11:12 is this you? http://kunfacebook.net/album.php?=
```

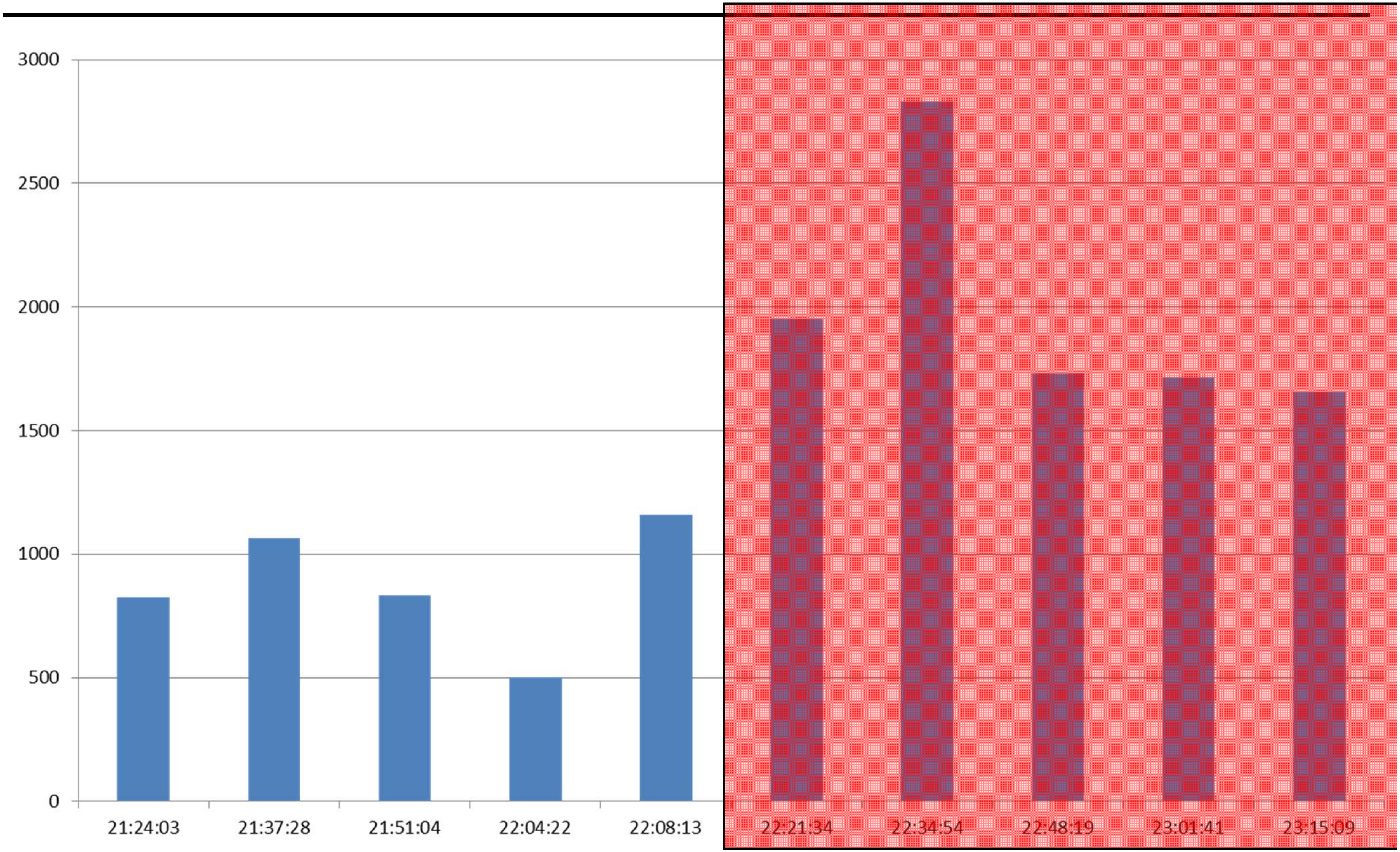
Some impact on clicks



Another example

```
2011-02-14 21:24:03 Foto :D http://fogz.eu/images886?=
2011-02-14 21:37:28 Foto :D http://fogz.eu/images886?=
2011-02-14 21:51:04
2011-02-14 22:04:22
2011-02-14 22:08:13 Foto :D http://fogz.eu/images91?=
2011-02-14 22:21:34 Foto :D http://justinloveis.net/album.php?=
2011-02-14 22:34:54 Foto :D http://justinloveis.net/album.php?=
2011-02-14 22:48:19 Foto :D http://justinloveis.net/album.php?=
2011-02-14 23:01:41
2011-02-14 23:15:09 Foto :D http://justinloveis.net/album.php?=
2011-02-14 23:28:27 Foto :D http://justinloveis.net/album.php?=
```

justinloveis works better than fogz.eu



Conclusions

- Criminals believe that having the bankname within their URLs improves the effectiveness of their phishing
  - This appears to be widely subscribed to
- Most phishing victims cannot parse URLs so the location of the bankname is pretty much irrelevant
  - This is a lesson that is not universally understood
- URLs shorteners appear to reduce the likelihood that people will be fooled
- BUT data is very noisy – so we will have to wait for the criminals to do more experiments before we can be entirely sure that we have the correct analysis

What matters in URLs?

`http://www.lightbluetouchpaper.org`

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

NPL
National Physical Laboratory