WHOIS Proxy/Privacy Abuse

Dr Richard Clayton
National Physical Laboratory
richard.clayton AT cl.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

NPL
National Physical Laboratory

ICANN Whois Studies

- ICANN doing a number of studies on the domain whois system:
  - NORC [in Chicago] has examined validity of whois details (most have some detail wrong!); the overall usage of privacy and proxy services (20%) and classifications of registrants
  - Carnegie Mellon University is investigating the extent to which Whois contact details are being misused
  - Interisle Consulting Group assessed feasibility of studying message relay and identity reveal by privacy/proxy services
  - Whois Service Requirements Survey by a GNSO Working Group
  - The present study by NPL into usage of privacy and proxy services when domains are maliciously registered
- Full (and more precise) details at
  - http://gnso.icann.org/en/group-activities/other/whois/studies

This Study

- National Physical Laboratory (NPL) in the UK commissioned to do a study into use of privacy and proxy services when domains are registered for harmful or illegal Internet activities
  - Main Author
    - Dr Richard Clayton — University of Cambridge
  - Project Team
    - Prof. Tyler Moore — SMU — typosquatting data
    - Dr Nicolas Christin — CMU — fake pharmacy data
    - Dr Tony Mansfield — NPL — experimental design
    - David Hindley — NPL — project management
- Contract started: April 2012
- Draft report issued: 24 Sep 2013
- Public comment period ends: 22 Oct 2013

Privacy and Proxy Services

- Normal Whois data
  - when a domain name is registered the registrant supplies their name and contact details
  - other fields give admin/billing/technical/etc. contacts. One can often learn phone numbers if the registrant is also admin/billing/etc.
  - this data is public (and available on the port 43 Whois service)
- Privacy Service
  - registrant name is provided, but contact details are generic (although sometimes the local part of the email address is specific to the domain name – to allow automated forwarding of email)
- Proxy Service
  - the domain is registered in the name of the proxy service and all contact details are generic (although sometimes the local part of the email address is specific to the registrant – to allow automated forwarding of email)

Research Hypotheses

- Original ICANN objective: assess truth of the hypothesis
  - *"A significant percentage of the domain names used to conduct illegal or harmful Internet activities are registered via privacy or proxy services to obscure the perpetrator's identity*"
- We thought it useful to also consider the hypothesis
  - *"The percentage of domain names used to conduct illegal or harmful Internet activities that are registered via privacy or proxy services to obscure identity is significantly greater than the equivalent percentage of domain names used for entirely lawful Internet activities.*"
- AND since malicious registrants might hide their identity in many different ways, we extended the study to assess whether valid contact information is given in Whois – in particular, can we make contact with the registrant using the phone number that they have provided ?

Spoiler!

- Original ICANN objective: assess truth of the hypothesis
  - *"A significant percentage of the domain names used to conduct illegal or harmful Internet activities are registered via Privacy or Proxy services to obscure the perpetrator's identity"*
- We thought it useful to also consider the hypothesis
  - *"The percentage of domain names used to conduct illegal or harmful Internet activities registered via Privacy or Proxy services to obscure identity is significantly greater than the equivalent percentage of domain names used for entirely lawful Internet activities."*

Yes!

Only partly correct.

- When contact details are maliciously hidden, many different ways we explored the study to assess whether valid contact information is given in Whois – in particular, can we make contact with the registrant using the phone number that they have provided

When domains are maliciously registered contact details are hidden, one way or another.

BUT many other domain registrants don't provide working phone numbers either

Summary of Methodology

- Basic approach:
  - obtain various lists of URLs being used in harmful ways
  - pick out the domain names
  - fetch Whois data for the biz/com/info/net/org domains
  - assess whether registrant is using privacy or proxy service
  - or, if registrant details available, determine contact phone number
- Precise statistics recorded for privacy/proxy/no_phone_number
- Random sample taken from registrants with phone number
  - phone call made (to the phone number provided in the Whois)
  - if answered then 1 question survey (in registrant's native language)
    - "did you register example.com" (for appropriate value of example.com)
  - if not answered then retried on different days/times
- Full details of methodology in the report

Phone Survey Results

- Phone number needed to be “apparently valid” (i.e. have sufficient digits, not be 9999999 or 0000000, and not have an invalid North American area code)
- BUT could turn out to be invalid when we dialled it
- OR the number was valid but just rang and rang
- OR we reached voicemail, or someone answered who could not help us reach the registrant, or registrant wasn’t ever available
- OR phone answered and we were told that the registrant was unknown to them and/or they had never heard of the domain
- OR we spoke to the registrant (or someone speaking for a company) and they agreed they had registered the domain

Phone Survey Results

- Phone number needed to be "apparently valid" (i.e. have enough digits, not be 9999999 or 0000000, or have an invalid North American area code)
- BUT could turn out to be a wrong number we called it
- OR the number was still in use and no one answered
- OR we reached voicemail, or someone answered who could not help us reach the registrant, or registrant wasn't ever available
- OR phone answered and we were told that the registrant was unknown to them and/or they had never heard of the domain
- OR we spoke to the registrant (or someone speaking for a company) and they agreed they had registered the domain

NOPHONE unless "apparently valid"

Treated as failure

Neither success nor failure

Neither success nor failure

Treated as failure

Treated as success

Phishing (Work Package 1)

- Phishing is the creation of fake websites for the purpose of stealing security credentials
  - not just banks, but email services, online games etc. etc.
- Source data was 32806 URLs (one week's worth), using 5105 domains – 56.9% of these were in biz/com/info/net/org
  - details of sources and processing in the report
- Used specialist knowledge to split these into three groups:
  - compromised machines (i.e. phishing pages added to innocent site)
    - 2121 domains
  - third parties (free webhosting domains, cloud services, etc.)
    - 263 domains (plus 1 had no Whois available, so ignored)
  - maliciously registered domain names
    - 449 domains (plus 5 had no Whois data available)

Phishing Analysis Results

- Privacy and proxy usage
  - third parties 13.7% low
  - compromised machines 24.7% average
  - maliciously registered domains 31.2% high
- Able to reach registrant by phone
  - third parties 32.3%
  - compromised machines 23.7%
  - maliciously registered domains 1.8%
- No hope of reaching registrant by phone (i.e.: no phone, or calls failed, or reached person who denied any knowledge of domain)
  - third parties 49.6%
  - compromised machines 61.7%
  - maliciously registered domains 92.5%

Other Types of Malicious Registration

- WP2: Data from aa419.org (Advanced Fee Frauds etc.)
  - 46.5% of registrants using privacy or proxy services
  - 88.9% impossible to contact by phone
- WP3: Unlicensed pharmacies
  - 54.8% of registrants using privacy or proxy services
  - 91.8% impossible to contact by phone
- WP5: Child sexual abuse image websites
  - 29.5% of registrants using privacy or proxy services
  - it is believed that 100% are impossible to contact by phone
- So there is a wide range of rates of usage of privacy or proxy services, but even if they are not used when domains are maliciously registered, we still find that in practice only a small minority of these registrants can be contacted by phone

Legal and Harmless Categories

Category	Privacy or proxy usage	Impossible to reach by phone	Did reach by phone [*]
Legal pharmacies	8.8%	24.2%	23.6%
Law firms	13.4%	33.6%	24.7%
Executive search Consultants	22.4%	36.7%	33.4%
Banks	28.2%	44.6%	15.2%
Alexa top 3500 (which were typo-squatted)	19.2%	47.1%	29.0%
Adult websites	44.2%	55.1%	5.7%

* CAVEAT: small samples mean quite large error bounds for this column

The Story So Far…

- Average usage of privacy or proxy services:
  - 20% NORC measurement across all domains
  - 25% our measure of compromised websites
- Privacy/proxy services used more often than average for maliciously registered domains
  - ranges from 29.5% to 54.8%
- BUT some legal and harmless activities are above average too
  - banks 28.2%, adult websites 44.2%
- If privacy or proxy services not used, phone number may be missing or invalid or reach someone other than registrant. Hence we consider the *a priori* "impossible to contact" rates:
  - malicious registrations: 88% – 92% (perhaps as much as 100%)
  - legal and harmless: 24% – 62%

More Complex Datasets

- WP8: StopBadware (domains where malware may be present)
  - mainly compromised sites, but some malicious registrations
  - 20.4% of registrants use privacy or proxy services
  - but 51.4% not possible to reach by phone
- WP8: SURBL (domains which indicate email is “spammy”)
  - mainly maliciously registered, but by no means all
  - 44.1% of registrants use privacy or proxy services
  - but only 58.5% not possible to reach by phone
  - CAUTION: high error bounds with this dataset because many domains had the same contact phone number
  - ALSO: some evidence of report inflation, i.e. all possible domains listed when multiple domains can be resolved to same location

Typosquatting

- Have already mentioned "typosquattED domains" : major sites from Alexa 3500 where small variants of domain name have been registered in the hope of being visited by sloppy tpyers
- WP4: typoquattING domains
  - privacy or proxy services used by 48.2% of registrants
  - 10.6% reached by phone (c.f. adult websites 5.7%)
    - BUT very high error bounds (small number of registrants involved)
- WP9: domains subject to UDRP (many "similar" names occur)
  - privacy or proxy services used by 39.7% of registrants
- Clearly some typosquatters are attempting to avoid being identified, whereas others have no qualms about this
  - a possible explanation is that typosquatting doesn't usually involve breaches of criminal law and police investigations – but civil action is more likely if the brand owner can identify "economies of scale"

Statistical Significance

- Measurements of privacy and proxy services are exact and for many work packages the samples are large – so expectation is that the results are robust
- Most variations >3% are statistically significant at 90% or better (see report for full details)
- Phone calls to registrants were done on a sampled basis
  - selection was random, but we avoided calling the same number more than once, so see report for (complex) statistical analysis
  - some small sample sizes and presence of large groups of domains with same contact number means that error bounds on the various categories of call outcome are sometimes quite large
  - results of "it is impossible to consider making a phone call to this registrant" have distinctly lower error bounds. These figures give a clear and robust indication that malicious registrants choose from a range of different methods of being uncontactable

Summary of Numerical Results of Study

Work package	Privacy or proxy usage	Not possible to call registrant	Maliciously registered?
Legal pharmacies	8.8%	24.2%	no
Law firms	13.4%	33.6%	no
Executive search consultants	22.4%	36.7%	no
Banks	28.2%	44.6%	no
Typosquatted domains	19.2%	47.1%	no
Phishing: third parties	13.7%	49.6%	no
StopBadware domains	20.4%	51.4%	some
Adult websites	44.2%	55.1%	no
SURBL domains	44.1%	58.5%	mostly
Phishing: compromised sites	24.7%	61.7%	no
Typosquatting	48.2%	67.7%	yes
Advanced Fee Fraud	46.5%	88.9%	yes
Unlicensed pharmacies	54.8%	91.8%	yes
Phishing: malicious registration	31.2%	92.5%	yes

Conclusions

- When domains are maliciously registered, privacy or proxy services ARE used more than average
  - BUT some legal and harmless activities also use privacy or proxy services significantly more than average
- When privacy or proxy services are not used in a malicious domain registration then valid contact phone numbers are very seldom provided – so overall the effect is that at least nine of ten of the registrants cannot be reached directly by phone
  - BUT the Whois details for the domains used for many lawful and harmless activities fail to contain valid contact numbers either, with anything between a quarter and two thirds of these registrants being inherently unreachable
  - BUT one wouldn't necessarily use the phone number on the Whois record as the only way to reach legitimate registrants…

And Now a Live Q&A Session…

Links to report, public comment page etc.:

http://www.icann.org/en/news/public-comment/whois-pp-abuse-study-24sep13-en.htm

Written public comments on the methodology and findings of the report are most welcome; they will need to be submitted before 23:59 UTC on 22 Oct 2013

- please take note that this isn't the place to submit philosophical essays on privacy, eCrime or the general nature of the Whois system – we wish you to address matters you find to be incorrect or unclear in our draft.