SSAC
ICANN Security and Stability Advisory Committee

Is the WHOIS service a source for email addresses for spammers?

SSAC Meeting

25 June 2007

rmohan@afilias.info - Ram Mohan
dave.piscitello@icann.org - Dave Piscitello

Objectives

Study the correlation between the publication of WHOIS data and delivery of spam to email addresses accessible via WHOIS

How Do Spammers obtain email addresses?

- Spammers harvest email addresses from many sources…
  - Web sites (via spambots)
  - Usenet, news groups, social networks, IRCs, and mailing lists
  - Email client Address books (via worms & viruses)
  - Directory Harvest Attacks
  - List Merchants
- Is the WHOIS service *another* source for spammers?

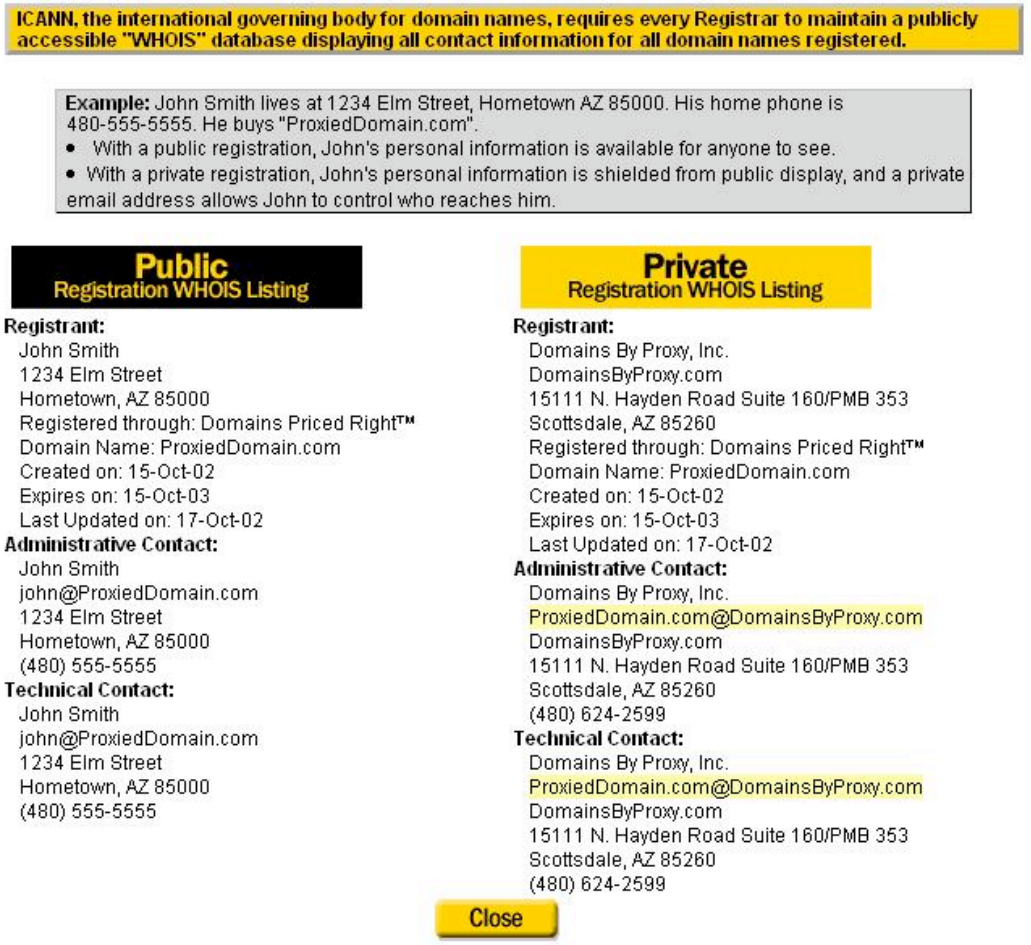
Can registries and registrars help mitigate automated email address collection?

- Registries and registrars offer services to protect registrant email addresses from automated collection via query-based WHOIS services
  - CAPTCHA
  - Rate limiting
  - Anti-scripting techniques
  - Other measures

- SSAC calls these measures *Protected-WHOIS*

Can registries and registrars help mitigate abuses of email addresses?

- Registries and registrars offer services to protect available email addresses from display and abuses
  - Email address substitution
  - Spam and antivirus filtering
- Customer chooses to have a a 3rd party listed as the registrant, other customers obtain a forwarding email address
- SSAC calls such measures *Delegated-WHOIS*



Objectives

1. Do spammers and data harvesters collect email addresses from domain name registration records using query-based WHOIS services?
2. Do measures to protect query-based WHOIS access from automated collection decrease the volume of spam delivered to a registrant?
3. Do email substitution and anti-spam services provided by registrars decrease the volume of spam delivered to a registrant?
4. Does the combination of measures described in (2) and (3) result in a decrease the volume of spam delivered to a registrant?
5. Do spammers favor one Top Level Domain over others?

Methodology

- Register domain names in 4 TLDs:
  - COM, DE, INFO, ORG
- Identify and publish email addresses in WHOIS
  - Avoid name bias in selecting email addresses
- Keep email addresses “off the radar”
  - Do not publish or use addresses in any form or forum
- Experiment! Monitor email received at these addresses under different conditions
  - The only publicly accessible record containing email addresses is the domain name registration records, so any messages delivered to email addresses is assumed to be unsolicited and bulk email (spam)

Experiments

- Determine the effects on spam delivery when Protected-WHOIS or Delegated-WHOIS is used and when both services are used together
- Track email that arrives:
  - Specifically to the email address published in the registration record
  - To any other email addresses under the registered domain name

Case #1: Protected-WHOIS and Delegated-WHOIS used

Delegated-WHOIS Used Protected-WHOIS Used		Feb - May 2007 (90 days)
	RandomlyChosenName1.ORG	2
	RandomlyChosenName1.DE	0
	RandomlyChosenName2.ORG	5
	RandomlyChosenName2.DE	2
	RandomlyChosenName3.ORG	7
	RandomlyChosenName3.DE	8
	RandomlyChosenName4.ORG	3
	RandomlyChosenName4.DE	3
	RandomlyChosenName5.ORG	7
	RandomlyChosenName5.DE	4

Case #2: Protected-WHOIS used, No Delegated-WHOIS

No Delegated-WHOIS Protected-WHOIS Used		Feb - May 2007 (90 days)
	RandomlyChosenName6.ORG	80
	RandomlyChosenName6.DE	38
	RandomlyChosenName7.ORG	230
	RandomlyChosenName7.DE	23
	RandomlyChosenName8.ORG	322
	RandomlyChosenName8.DE	54
	RandomlyChosenName9.ORG	1220
	RandomlyChosenName9.DE	403
	RandomlyChosenName10.ORG	384
	RandomlyChosenName10.DE	125

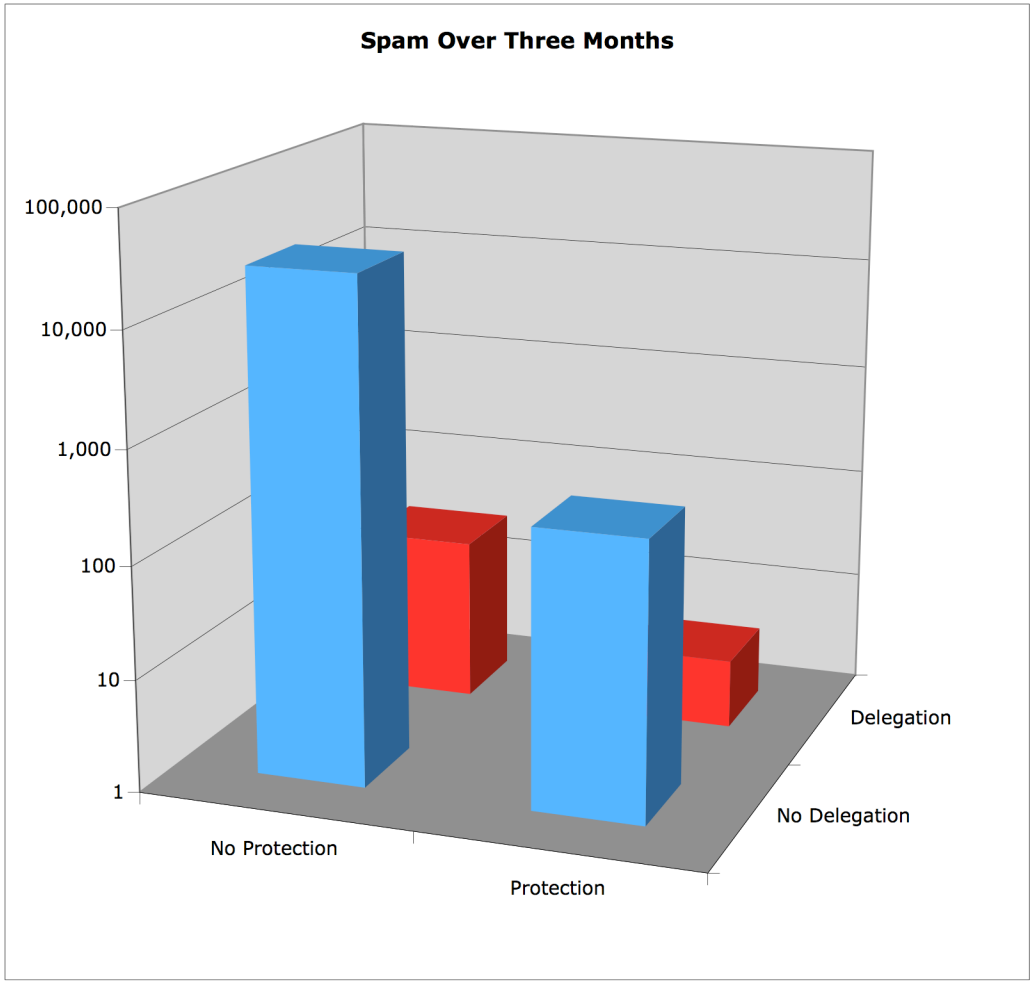
Case #3: Delegated-WHOIS used, No Protected-WHOIS

Delegated-WHOIS used, No Protected-WHOIS		Feb - May 2007 (90 days)
	RandomlyChosenName1.INFO	8
	RandomlyChosenName1.COM	37
	RandomlyChosenName2.INFO	39
	RandomlyChosenName2.COM	75
	RandomlyChosenName3.INFO	18
	RandomlyChosenName3.COM	54
	RandomlyChosenName4.INFO	5
	RandomlyChosenName4.COM	11
	RandomlyChosenName5.INFO	14
	RandomlyChosenName5.COM	23

Case #4: Neither Protected-WHOIS nor Delegated-WHOIS used

No Delegated-WHOIS Used No Protected-WHOIS used		Feb - May 2007 (90 days)
	RandomlyChosenName6.INFO	11700
	RandomlyChosenName6.COM	57870
	RandomlyChosenName7.INFO	3870
	RandomlyChosenName7.COM	40770
	RandomlyChosenName8.INFO	4590
	RandomlyChosenName8.COM	28890
	RandomlyChosenName9.INFO	36270
	RandomlyChosenName9.COM	76500
	RandomlyChosenName10.INFO	1710
	RandomlyChosenName10.COM	16200

Comparison of Results



For an email address that is *not* published anywhere other than the WHOIS

- When protected-WHOIS is used, it is possible to achieve two orders of magnitude reduction in the amount of spam delivered
- When Delegated-WHOIS is used, it is possible to achieve three orders of magnitude reduction in the amount of spam delivered
- When Protected-WHOIS *and* Delegated-WHOIS are used, it is possible to achieve close to four orders of magnitude reduction in the amount of spam delivered

Registrar Experiments

- Is the size and business model of registrars is a factor in registrants receiving spam email?
  - Worth considering?
  - Limited sample size study performed by SSAC
  - Preliminary results not conclusive

Findings

1. The appearance of email addresses in responses to WHOIS queries virtually assures that spam will be delivered to these email addresses.

2. For an email address that is not published anywhere other than the WHOIS, when Protected-WHOIS or Delegated-WHOIS services are used, the volume of spam delivered to recipients in domains is significantly reduced. **The greatest reduction in the delivery of spam is realized when both protective measures are applied.**

3. Of the two forms of preventative measures registrants can obtain through registries and registrars, the Delegated-WHOIS is more effective than Protected-WHOIS.

Findings

With respect to the choice of TLD

4. The TLD itself does not appear to matter
   - Four TLDs were studied
   - .COM names received more spam than other TLDs
   - .DE names received fewer spam than other TLDs
     - SSAC speculates that this could be because .DE zone file is not easily accessible

Conclusions

1. Registries and registrars that implement anti-abuse measures such as rate-limiting, CAPTCHA and similar measures can protect WHOIS data from automated collection.

2. Anti-spam measures provided with domain name registration services are effective in protecting email addresses not published anywhere other than the WHOIS from spam.

3. The appearance of email addresses in responses to WHOIS queries will very likely cause spam to be delivered to these email addresses.

4. The combination of protected-WHOIS and Delegated-WHOIS services as defined in this presentation is the most effective way to prevent the WHOIS service being used as a source of email addresses for spammers.

5. SSAC recommends further studies to investigate whether spammers are more likely to target
   - Certain TLDs over other TLDs
   - Large registrars than small for email address collection
   - Registrars having a reseller or retail business model.