SSAC
ICANN Security and Stability Advisory Committee

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

October 2007

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

Objectives

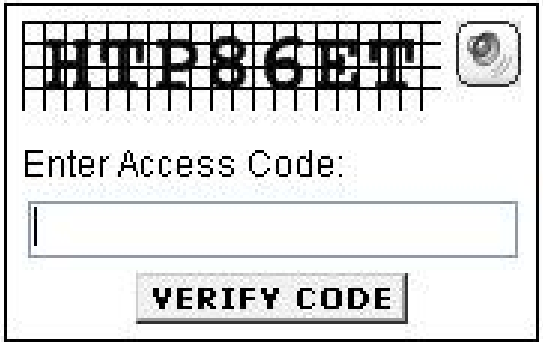
Study the correlation between the publication of WHOIS data and delivery of spam to email addresses accessible via WHOIS services

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?

ICANN, the international governing body for domain names, requires every Registrar to maintain a publicly accessible "WHOIS" database displaying all contact information for all domain names registered.

- 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 3rd party listed as the registrant, other customers obtain a forwarding email address
- SSAC calls such measures *Delegated-WHOIS*

Example: John Smith lives at 1234 Elm Street, Hometown AZ 85000. His home phone is 480-555-5555. He buys "ProxiedDomain.com".

- With a public registration, John's personal information is available for anyone to see.
- With a private registration, John's personal information is shielded from public display, and a private email address allows John to control who reaches him.

Public
Registration WHOIS Listing

Registrant:
John Smith
1234 Elm Street
Hometown, AZ 85000
Registered through: Domains Priced Right™
Domain Name: ProxiedDomain.com
Created on: 15-Oct-02
Expires on: 15-Oct-03
Last Updated on: 17-Oct-02
Administrative Contact:
John Smith
john@ProxiedDomain.com
1234 Elm Street
Hometown, AZ 85000
(480) 555-5555
Technical Contact:
John Smith
john@ProxiedDomain.com
1234 Elm Street
Hometown, AZ 85000
(480) 555-5555

Private
Registration WHOIS Listing

Registrant:
Domains By Proxy, Inc.
DomainsByProxy.com
15111 N. Hayden Road Suite 160/PMB 353
Scottsdale, AZ 85260
Registered through: Domains Priced Right™
Domain Name: ProxiedDomain.com
Created on: 15-Oct-02
Expires on: 15-Oct-03
Last Updated on: 17-Oct-02
Administrative Contact:
Domains By Proxy, Inc.
ProxiedDomain.com@DomainsByProxy.com
DomainsByProxy.com
15111 N. Hayden Road Suite 160/PMB 353
Scottsdale, AZ 85260
(480) 624-2599
Technical Contact:
Domains By Proxy, Inc.
ProxiedDomain.com@DomainsByProxy.com
DomainsByProxy.com
15111 N. Hayden Road Suite 160/PMB 353
Scottsdale, AZ 85260
(480) 624-2599

Close

Objectives

1. Do spammers collect email addresses from domain name registration records using query-based WHOIS services?

2. For an email address that is not published anywhere other than the WHOIS,
   - Do measures to protect query-based WHOIS access from automated collection reduce spam delivery to a registrant
   - Do email substitution and anti-spam services reduce the volume of spam delivered to the end-user/licensee of the domain, who has retained the registrar as his agent to be the public-facing domain name registrant?

3. Does the combination of measures described in (2) result in a decrease in the frequency of spam delivery to a registrant?

Methodology

- Register domain names in 4 TLDs: COM, DE, INFO, ORG
  - Use randomly composed 2nd level labels
- Identify and publish email addresses in WHOIS
  - Use randomly composed <user-ID> for email addresses
- Keep email addresses “off the radar”
  - Do not publish or use addresses in any form or forum
- Monitor email delivered to these addresses under different conditions
  - Addresses are published in WHOIS with no protective measures
  - One or more measures are applied to protect the addresses from disclosure using WHOIS services

Experiments

- Determine the effects on spam delivery when
  - Protected-WHOIS is used
  - Delegated-WHOIS is used
  - Both services are used
- Track email that is delivered to
  - the email address published in the registration record
  - other email recipients @ the registered domain name
- Characterize the kinds of spam delivered to these addresses (please see the report)

Case #1: Neither Protected-WHOIS nor Delegated-WHOIS used

***NO* Protected-WHOIS** ***NO* Delegated-WHOIS**	**# of spam messages delivered**	**Spam delivered to Published Address**	**Spam delivered to all other recipient addresses**
RandomlyChosenName6.info	11700	4446	7254
RandomlyChosenName6.com	57870	10995	46875
RandomlyChosenName7.info	3870	929	2941
RandomlyChosenName7.com	40770	8154	32616
RandomlyChosenName8.info	4590	1561	3029
RandomlyChosenName8.com	28890	12712	16178
RandomlyChosenName9.info	36270	6529	29741
RandomlyChosenName9.com	76500	27540	48960
RandomlyChosenName10.info	1710	1402	308
RandomlyChosenName10.com	16200	8748	7452
Total	278370	83016	195354
Percent of Total		**29.82%**	**70.18%**

Case #2: Protected-WHOIS used but no Delegated-WHOIS

Protected-WHOIS but *NO* Delegated-WHOIS	**# of spam messages delivered**	**Spam delivered to Published Address**	**Spam delivered to all other recipient addresses**
RandomlyChosenName6.org	80	18	62
RandomlyChosenName6.de	38	12	26
RandomlyChosenName7.org	230	41	189
RandomlyChosenName7.de	23	13	10
RandomlyChosenName8.org	322	277	45
RandomlyChosenName8.de	54	12	42
RandomlyChosenName9.org	1220	671	549
RandomlyChosenName9.de	403	161	242
RandomlyChosenName10.org	384	88	296
RandomlyChosenName10.de	125	110	15
Total	2879	1404	1475
Percent of Total		**48.77%**	**51.23%**

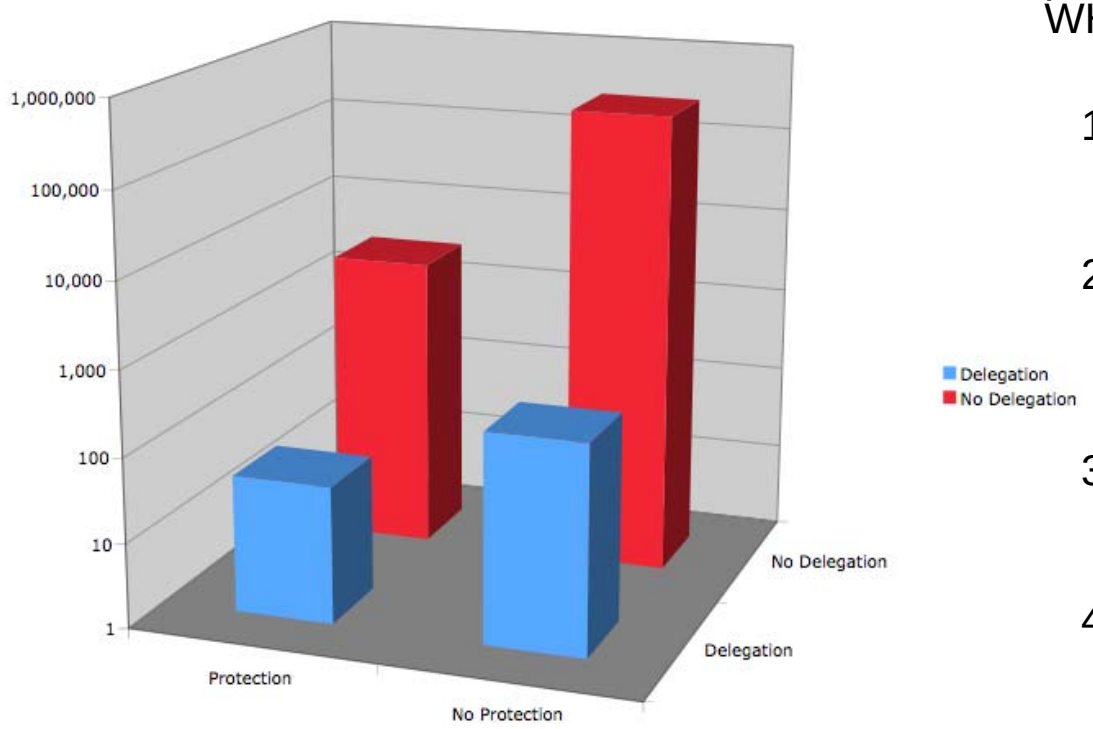
Case #3, Delegated-WHOIS used but no Protected-WHOIS

***NO* Protected-WHOIS but Delegated-WHOIS**	**# of spam messages delivered**	**Spam delivered to Published Address**	**Spam delivered to all other recipient addresses**
RandomlyChosenName1.info	8	1	7
RandomlyChosenName1.com	37	12	25
RandomlyChosenName2.info	39	20	19
RandomlyChosenName2.com	75	16	59
RandomlyChosenName3.info	18	7	11
RandomlyChosenName3.com	54	35	19
RandomlyChosenName4.info	5	1	4
RandomlyChosenName4.com	11	5	6
RandomlyChosenName5.info	14	4	11
RandomlyChosenName5.com	23	17	6
Total	284	118	166
Percent of Total		**41.55%**	**58.45%**

Case #4: Protected-WHOIS used and Delegated-WHOIS used

Protected-WHOIS + Delegated-WHOIS	# of spam messages delivered	Spam delivered to Published Address	Spam delivered to all other recipient addresses
RandomlyChosenName1.org	2	2	0
RandomlyChosenName1.de	0	0	0
RandomlyChosenName2.org	5	4	1
RandomlyChosenName2.de	2	1	1
RandomlyChosenName3.org	7	4	3
RandomlyChosenName3.de	8	4	4
RandomlyChosenName4.org	3	3	0
RandomlyChosenName4.de	3	0	3
RandomlyChosenName5.org	7	0	7
RandomlyChosenName5.de	4	1	3
Total	41	19	22
Percent of Total		**46.34%**	**53.66%**

Comparison of Results



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

1. Unprotected registrant email addresses received significant amounts of spam.
2. Registrant email addresses protected by protected-WHOIS may achieve two orders of magnitude better defense against spam.
3. Registrant email addresses protected by achieve three orders of magnitude better defense against spam.
4. Registrant email addresses protected by Protected-WHOIS *and* Delegated-WHOIS may achieve close to four orders of magnitude better defense against spam.

Findings

1. The appearance of email addresses in responses to WHOIS is a contributor to the receipt of spam, albeit just one of many.

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

Findings (continued)

3. Of the two forms of protective measures registrants can obtain through registries/registrars, the Delegated-WHOIS appears to be somewhat more effective than Protected-WHOIS.

4. Spam messages were delivered to the email address registered as the contact for a domain name and to other (non-existent, non-published) recipient email addresses in the registered domain as well. SSAC draws no conclusions specific to WHOIS services from these deliveries and leaves the matter to the reader to interpret the data.

Conclusions

1. Registries and registrars that implement anti-abuse measures such as rate-limiting, CAPTCHA, non-publication of zone file data 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.

Conclusions (continued)

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

4. The combination of Protected-WHOIS and Delegated-WHOIS services as defined in this report is an effective way to prevent an email address published in the WHOIS service from being used as a source of email addresses for spammers.

Conclusions (continued)

5. Further studies may be needed to investigate whether spammers have preferential targets. Studies might ask such questions as:

   - Are certain TLDs more attractive to spammers?
   - Are large or small registrars more commonly targeted for automated collection?
   - Do spammers favor registrars who have a reseller or retail business model?
   - Does the price of a TLD affect its popularity for use in spam?
   - Can the registries adopt any measures that would reduce the level of spam?
   - Is there any material difference in the spam level for ccTLDs vs. gTLDs?