SAC 023: Is the WHOIS Service a Source for email Addresses for Spammers?

A Report from the ICANN
Security and Stability Advisory Committee

SAC023 October 2007

About the Security and Stability Advisory Committee

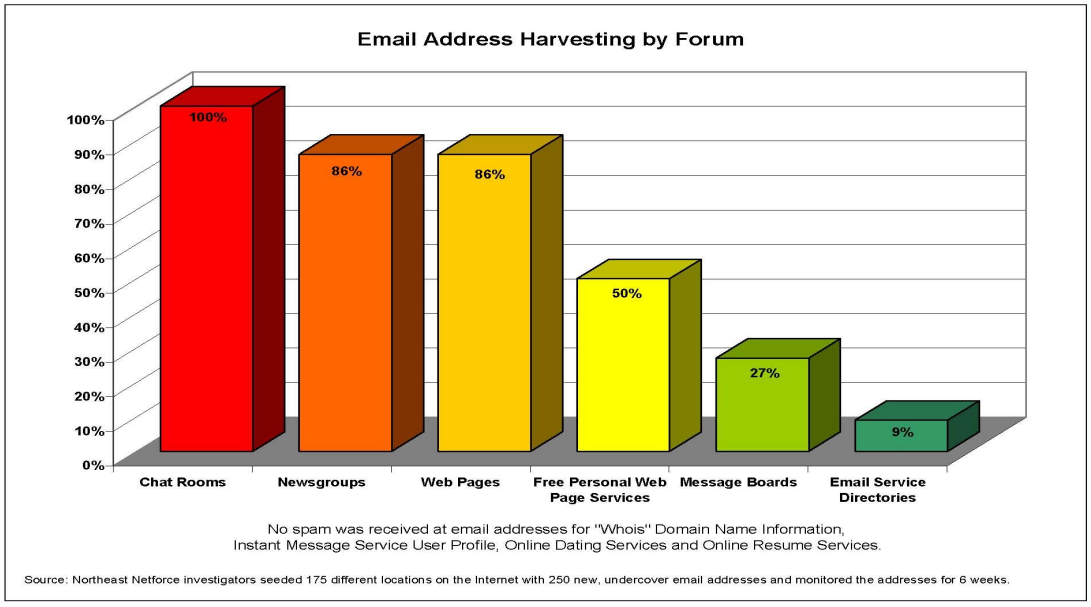
The Security and Stability Advisory Committee (SSAC) is an advisory committee to the Internet Corporation for Assigned Names and Numbers (ICANN). The Committee's purpose is to offer independent advice to the ICANN board, ICANN staff and various ICANN supporting organizations, councils and committees as well as to the community at large on matters relating to the security and integrity of the Internet's naming and address allocation systems. The Committee has no official authority to regulate, enforce or adjudicate. Those functions belong to others. The advice offered by the Committee should be evaluated on its merits, not on the status of the Committee or its members.

About this Report

This report was prepared by the SSAC Fellow, Dave Piscitello, under the direction of Ram Mohan, who designed and executed the study, and the Committee and represents output from the committee as a whole. Appendix A contains the current list of members and contributors to this report.

Executive Summary

In the SSAC's prior work on WHOIS (SAC 003, 2003), the Committee stated that "it is widely believed that WHOIS data is a source of email addresses for the distribution of spam." The US Federal Trade Commission conducted a study at approximately the same time. In *Email Address Harvesting: How Spammers Reap What You Sow*, FTC researchers reported that "email addresses posted in instant message service user profiles, 'WHOIS' domain name registries, online resume services, and online dating services did not receive any spam during the six weeks of [their] investigation."[1] This SSAC study on WHOIS considers again whether the WHOIS service is a source of email addresses for spammers.



Source: http://www.ftc.gov/bcp/conline/edcams/spam/pubs/harvestchart.pdf

To accomplish this task, the SSAC conducted an experiment to see the effects of two services registrars now offer to protect registrant email addresses from publication and abuse. For the sake of brevity, these services are referred to as Protected-WHOIS and Delegated-WHOIS. For the study, SSAC registered and monitored email delivery to randomly composed strings as second-level labels in four Top Level Domains: COM,

[1] The report may be found at http://www.security.iia.net.au/downloads/spamalrt-ftc.pdf. An excerpt of the FTC study is included as Appendix B.

DE, INFO, and ORG. The domain names were registered in February 2007. The recipient chosen for the registrant email address for each of the registration records was also chosen randomly. These were neither used in correspondence nor published electronically in any form (web, IM user, online service…). Thus, the only practical vectors to obtain these specific email addresses other than brute force derivation (or guessing) was via a WHOIS service or through the registrar or reseller in whose database(s) the email address were stored. SSAC collected and analyzed all email messages delivered to these addresses for a period of approximately three months.

Based on the data collected, the Committee finds that the appearance of email addresses in response to WHOIS queries is indeed a contributor to the receipt of spam, albeit just one of many.

This report is narrowly focused on the relationship between WHOIS services and spam, and not on the broad set of issues related to spam. The Committee members involved in the WHOIS study do not believe that the WHOIS service is the dominant source of spam. The Committee did not conduct any work on the proportion of spam received as a result of email addresses appearing in WHOIS responses as compared to other methods of email address discovery.

The Committee offers the following findings for consideration:

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

Finding (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.**

Finding (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.

Finding (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.

On the basis of these Findings, the Committee draws the following conclusions:

Conclusion (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.

Conclusion (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.

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

.

Conclusion (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.

.

Conclusion (5) SSAC concludes that further studies may be needed to investigate whether spammers have preferential targets. Suggested 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?

1. Introduction

Unsolicited bulk email[2] (UBE, or spam) has evolved from an intrusive and productivity-hampering misuse of a critical application to a serious security threat that affects a higher percentage of users than any other form of Internet attack. Spam is a common vector for malicious attacks against computers, scams, deception, fraud, and identity theft. Through the use of a variety of impersonation and deception techniques delivered by email, parties who send spam (spammers) infect computers with viruses and malicious code that turns the infected system into an agent for the spammer. This agent may act as an email relay or spyware. Criminals also use unsolicited email to lure recipients into visiting a web site that impersonates a legitimate site such as an online banking, e-merchant, or e-payment site. The bogus but convincing site often dupes the victim into disclosing personal and financial information which is subsequently fraudulently used for theft and unauthorized purchases. Spam is also used to impersonate network and system administrator-generated email to dupe employees into disclosing organizational account information which can be used to impersonate authorized users and abet attacks against the organization.

The Internet community has invested considerable time, talent and expense to develop numerous spam defenses and countermeasures, governments at local and national levels have enacted laws criminalizing many forms of spam, and law enforcement and activist groups have redoubled efforts to identify and defeat "spam gangs", but spammers continue to evade and confound efforts to bring spam to a halt.

Nearly all Internet email accounts receive some spam. This is an unfortunate consequence of any form of communication where a correspondent's address is made public or can be discovered. Spammers need little sophistication and only a small investment in automated software to collect or "harvest" email addresses and use these to send (tens of) millions of copies of a message containing one or more forms of attack.

Spammers harvest email addresses from many sources. In this report, SSAC considers whether the WHOIS service is one of several widely-perceived sources for collecting email addresses. The report also considers whether measures to thwart automated access to WHOIS and services registrars offer to protect registrants from email abuse are effective methods for mitigating spam. The report begins with background and terminology relevant to the evolution of the protocols, data elements, and services collectively referred to as WHOIS. Readers familiar with this material are encouraged to skip to Section 3.

[2] Unsolicited Bulk Email, or UBE, is Internet mail ("email") that is sent to a group of recipients who have not requested it. A common term for UBE is "spam", although that term encompasses a wider range of intrusive transmissions. Note: The term Unsolicited Commercial Email (UCE) was originally chosen because much of the early debate about UBE was centered in the United States where commercial speech can be regulated by the government but political and religious speech cannot. However, on reflection, because UBE is an international problem, the term "UCE" was changed to "UBE". Source: http://www.imc.org/ube-def.html

2. Background and Terminology

The WHOIS service and protocol were originally developed and deployed in1982 as a transaction based service to provide a registry (directory) for "each individual with a directory on an ARPANET host, who is capable of passing traffic across the ARPANET".[1]

Originally, network operators were asked by the US Defense Communications Agency (DCA) to submit the following information to the registry.

- full name
- middle initial
- U.S. Postal mailing address (including mail stop and full explanation of abbreviations and acronyms)
- ZIP code
- telephone (including Autovon and FTS, if available)
- one network mailbox [1]

The set of Network Information Center names and contacts constituted the first set of what we today call WHOIS service data elements. DCA encouraged network operators to provide users with access to this network service. The query to this service was dubbed "WHOIS" and the contact information was informally referred to as "NICNAMES".

The original service listened to TCP port 43 (NICNAME/WHOIS) for single command-line queries submitted in ASCII and completed using carriage-return and line-feed symbols (ASCII CR and LF).

The WHOIS protocol standard was modified in 1985 (RFC 954,[2]) and again in 2004 (RFC 3912, [3]), in part to remove historical references to protocols (e.g., NCP) and authorities (e.g., US DCA) and to generalize the applicability of WHOIS to the Internet community rather than selected networks (e.g., DDN, ARPANET), but also to acknowledge the range of information services WHOIS had evolved to support[3].

2.1 WHOIS Service and gTLD Registry Agreements

Organizations that have entered into an gTLD Registry Agreement provide a WHOIS information service in accordance with a Public WHOIS Specification. ICANN accredited registrars are obliged by the Registrar Accreditation Agreement (RAA, [4]) to collect and display WHOIS information. These specifications identify the forms of user access registries and their registrars are to provide, the WHOIS service data elements and

[3] From RFC 3912: "While originally used to provide 'white pages' services and information about registered domain names, current deployments cover a much broader range of information services."

output fields (known as Domain Records), and the procedures for providing access and data preparation.[4]

The data elements that comprise a domain name registration record at an ICANN accredited registrar include:

- The name of the domain name registered;
- The IP addresses of the primary name server and secondary name server(s) of the name registered;
- The corresponding names of those name servers;
- The identity of the registrar;
- The original creation date and term of the registration;
- The name and postal address of the Registered Name Holder;
- The name, postal address, email address, voice telephone number, and (where available) fax number of the technical contact for the name registered; and
- The name, postal address, email address, voice telephone number, and (where available) fax number of the administrative contact for the name registered.

This information must be provided by a registrant to a registrar to register a domain name. ICANN has implemented policies and measures to improve the accuracy and availability of domain name registration records, including

- the WHOIS Data Reminder Policy (WDRP, [5]),
- the WHOIS Data Problem Reporting System (WDPRS, [6]), a problem reporting system that allows parties to report allegedly inaccurate WHOIS data and requires that registrars verify the data with the registrant, and
- annual WDRP compliance audits, and will commence a WHOIS data accuracy audit in 2007 [7].

2.2 WHOIS Service and ccTLD Registries

WHOIS services are not covered under accountability frameworks between ICANN and ccTLDs. Readers are encouraged to solicit information regarding WHOIS services directly from individual ccTLD operators.

2.3 WHOIS Access

Domain name registration information is often referred to as "WHOIS data". This loose terminology perpetuates a misconception that all registration records are held in a central repository. In practice, domain name registration information is stored in multiple databases maintained by registries and registrars. These databases can be queried through interfaces provided by registrars and registries. Two forms of access are provided: individual and bulk record access.

[4] Examples of Public WHOIS Specifications can be found in the .BIZ [32], .ORG [33], and .NET [34] agreements.

2.3.1 Query-based WHOIS Access

Registries, registrars, and resellers provide access to individual domain name registration information through one or more forms of query-response applications. Registries and registrars commonly support individual domain name queries via a World Wide Web browser interface. Many commercial and community web portals also provide a web-based WHOIS access by accepting queries from an end user, forwarding these to a registrar or registry, and directing the response from the registrar or registry back to the end user.

A successful query to a "thick" registry (such as .ORG or .INFO) will return the following information, referred to as the **Domain Record:**

- Domain Name
- Domain ID
- Sponsoring Registrar
- Sponsoring Registrar IANA ID
- Domain Status
- Registrant, Administrative, Technical and Billing Contact Information including
  - ID
  - Name
  - Organization
  - Address
  - Geographic Location Code
  - Phone Number
  - Facsimile Number
  - Email
- Name Server(s)
- Created by Registrar
- Last Updated by Registrar
- Domain Registration Date
- Domain Expiration Date
- Domain Last Updated Date

A successful query to a "thin" registry (such as .COM) will return the following information.

Record Type	Summary
domain	domain name
nameserver	nameserver name
registrar	registrar name and whois server

A summary of the matching record is shown and the sub-display follows directly after.

The following keywords restrict a search to a certain TYPE of field in the database:

domain	**Finds a domain record. Find domain name, registrar name, whois server and URL, Name server name and IP Addresses, and updated date. For example, "www.example.com".**
name server	Finds name server records. Find name server name, registrar name, IP addresses, Whois Server name and URL. For example, 'name server NS.EXAMPLE.COM' or 'name server 101.198.1.101'.
registrar	Finds records for "registrar". Find Registrar name, email address, phone number and contact information. For example, "registrar ABC Registrar, Inc."

Command line and graphical user interface (GUI) -based applications available for popular operating systems may also be used to access WHOIS service. These use the WHOIS protocol (RFC 3912) at TCP Port 43/NICNAME. These commercial and freeware applications allow users to compose domain name and IP address queries and to view all or some of the data returned in the responses. WHOIS access is frequently incorporated into network diagnostic and vulnerability assessment utilities, web and security system log analysis applications, and software used by administrators and secondary domain name speculators to monitor and track domain registrations and status.

2.3.2 Bulk WHOIS Access

Section 3.6.6 of ICANN's Registrar Accreditation Agreement (RAA) obliges registrars to provide third-party bulk access upon request to the following data elements (this applies to gTLD registration data):

Data Element	Relevant Section of ICANN's RAA
The name of the Registered Name	§ 3.3.1.1
The names of the primary and secondary domain name server(s) for the Registered Name	§ 3.3.1.2
The identity of registrar	§ 3.3.1.3
The original creation date of the registration	§ 3.3.1.4
The expiration date of the registration	§ 3.3.1.5
The name and postal address of the registered name holder	§ 3.3.1.6
The name, postal address, email address, voice telephone number, and fax number of the technical contact for the registered name	§ 3.3.1.7
The name, postal address, email address, voice telephone number, and fax number of the administrative contact for the registered name	§ 3.3.1.8

§3.3.6.4 - §3.3.6.6 of the RAA identify usage and resale restrictions registrars must impose on third parties who are permitted one form of bulk access (see also the WHOIS Marketing Restriction Policy, WMRP [8]). Any party who requests bulk access must agree to the registrar's terms, which may include an annual fee for this form of access. Registrars are not restricted from offering bulk access under other terms and conditions.

2.3.3 GNSO WHOIS Activities and SPAM

The GNSO and particularly the GNSO WHOIS Task Force have studied a broad set of issues related to the amount of contact information ICANN requires registrars to display. Areas the WHOIS Task Force are actively studying include the protection of personal data, mechanisms for notifying registrants of inaccurate WHOIS data, improving the accuracy of WHOIS data, and dealing with WHOIS data abuse. Issues related to dealing with WHOIS data abuse are referenced in the Final Task Force Report on WHOIS Services 12 March 2007 [9] in a quote from an email by Ross Rader [10]:

> "the amount of data that ICANN requires registrars to display in the WHOIS is facilitating undesirable behaviors like renewal scams, data-mining, phishing, identity theft, ..."

An OPoC (Operational Point of Contact) proposal recommended by the WHOIS Task Force is now being developed by the GNSO. A WHOIS Working Group was created in March 2007 to continue this work. The OPoC proposes that some registrants (such as natural persons) use a new set of contact elements, OPoC, in place of the current administrative and technical contact details in the published WHOIS. This would allow some registrants to only publish the contact details of the OPoC, rather than the administrative and technical contact details. In the case of an issue with the domain name, the OPoC would contact the registrant.

The registrant can opt to have an OPoC displayed instead of the registrant's contact information, including the registrant's email address. Note that registrars are not required to publish the registrant's email address currently. The registrant's name and jurisdiction would still be displayed. Note: It is envisioned that such services as anti-spam or other email filtering features would be provided at the discretion of the registrars. The OPoC proposal can be read in its entirety in [9].

3. Uses of Domain Records

In this section, we attempt to list the known and speculated uses and abuses of WHOIS services.

- To contact network administrators for resolution of technical matters related to networks associated with a domain name (e.g., DNS or routing matter, origin and path analysis of DoS and other network-based attacks).

- To diagnose registration difficulties. WHOIS queries provide information that is often useful in resolving a registration ownership issue, such as the creation and expiration dates and the identity of the registrar.

- To contact web administrators for resolution of technical matters related to web associated with a domain name.

- To obtain the real world identity, business location and contact information of an online merchant or business, or generally, any organization that has an online presence..

- To associate a company, organization, or individual with a domain name, and to identify the party that is operating a web or other publicly accessible service using a domain name, for commercial or other purposes.
.
- To contact a domain name registrant for the purpose of discussing and negotiating a secondary market transaction related to a registered domain name.

- To notify a domain name registrant of the registrant's obligation to maintain accurate registration information[5].

- To contact a domain name registrant on matters related to the protection and enforcement of intellectual property rights[6].

- To gather information about a company, organization, or individual as part of the *footprinting* and target acquisition phase of an Internet attack. Internet footprinting involves searches and queries of available publicly accessible databases, including web pages, the U.S. Securities Exchange Commission's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database, WHOIS, and DNS[7]

- To establish or look into an identity in Cyberspace, and as part of an incident response following an Internet or computer attack, security professionals and law

[5] WHOIS Data Reminder Policy [5]

[6] Comments from the American Intellectual Property Law Assocation, regarding the preliminary reports of the WHOIS Task Forces [35]

[7] *Hacking Exposed*, by McClure, Scambray, & Kurtz, Osborne Press, ISBN 0-07-212127-0; in particular, see Chapter 1, Footprinting – Target Acquisition, pp 7-14. This phase of an Internet attack is sometimes called *reconaissance*.

enforcement agents use WHOIS to identify points of contact[8]

- To gather investigative leads (i.e., to identify parties from whom additional information might be obtained). Law enforcement agents use WHOIS to find email addresses and attempt to identify the location of an alleged perpetrator of a crime involving fraud[9].

- To investigate spam, law enforcement agents look to the WHOIS database to collect information on the website advertised in the spam[10].

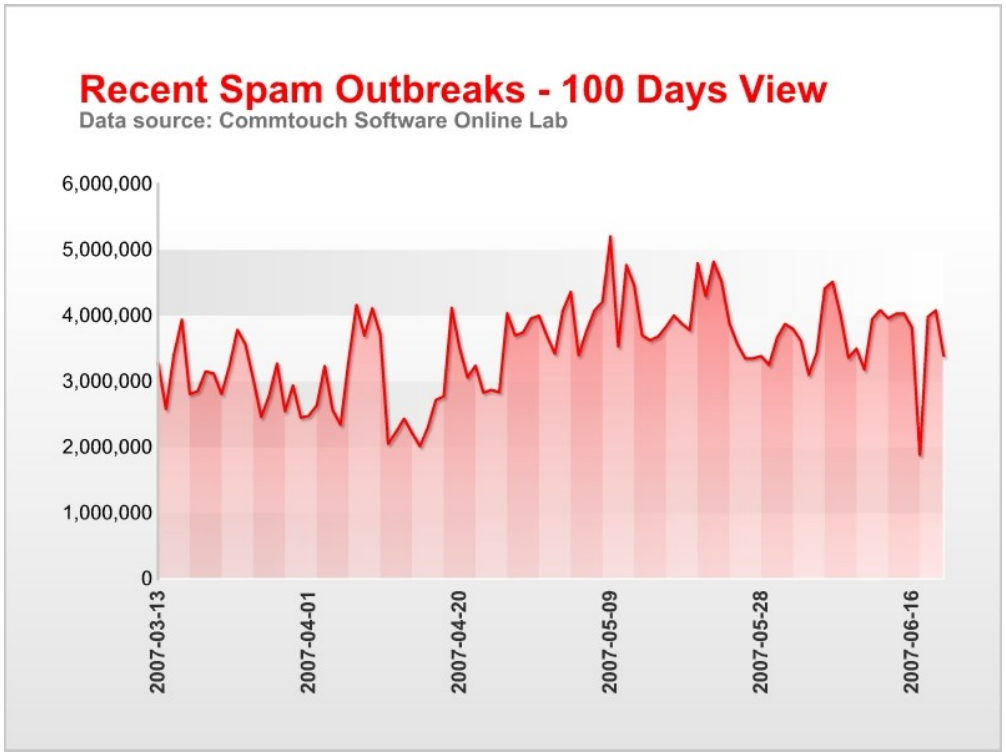
- To collect or "farm" email addresses for the purpose of delivering unsolicited electronic mail[11].

This list is not exhaustive. The Committee makes no claims here except that the sources identified claim that domain records have been used in the manners described.

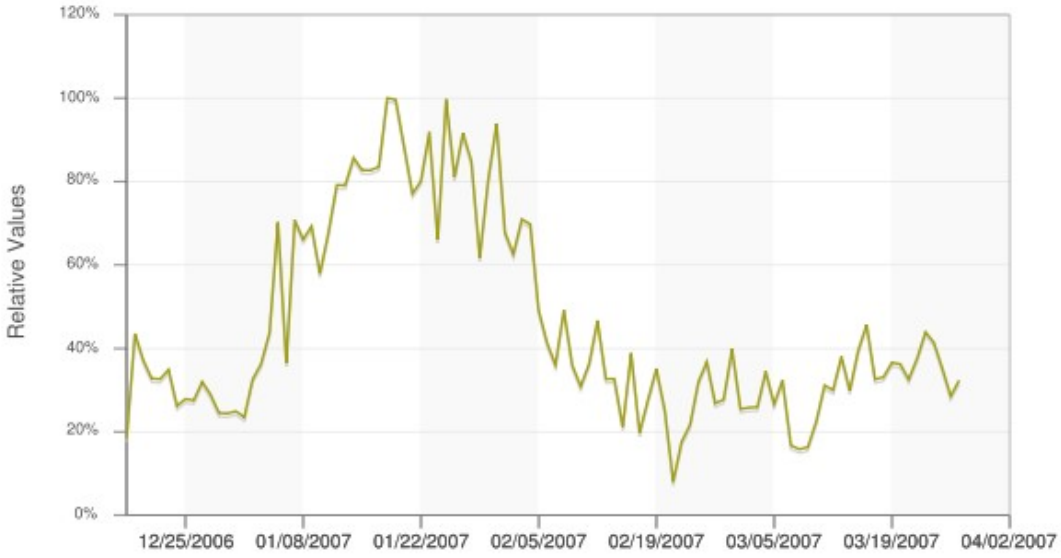
[8] *Incident Response: Investigating Computer crime*, Mandia & Procise, Osborne Press, ISBN 0-07-213182-9, pp 435-439.
[9] *How the FTC uses WHOIS Data* [37]
[10] *The Importance of WHOIS data bases for spam enforcement* [38]
[11] FAQ: How do spammer's get people's email addresses? [39]

4. WHOIS and SPAM

Spam is an Internet pandemic. Depending on the sources of data, between 40 and 90 percent of email that is delivered can be classified as spam by the recipient [11, 12, 13, 14]. Estimates vary in part due to phenomena called spam outbreaks that introduce dramatic fluctuations in spam delivery, as illustrated below:



Effects of spam outbreaks on spam volume



Percent of email considered spam (data: CommTouch, graph: Swivel.com [15])

Spam is the commonly adopted term for Unsolicited Bulk Email, or UBE. The Internet Mail Consortium defines UBE as Internet mail ("email") that is sent to a group of recipients who have not requested it. In practice, the term spam encompasses a wider range of intrusive transmissions.

Estimates also vary depending on who and how spam statistics are collected, how stringently spam enforcement policies are set (i.e., what constitutes spam at a detection point). Anecdotal comparison of statistics published by commercial anti-spam vendors suggests that estimating that 80 per cent of email delivered is spam.

Legal and technical definitions of spam vary, but generally (according to the Electronic Frontier Foundation and anti-spam organizations such as The Spamhaus Project) two characteristics can be used to distinguish spam from legitimately transmitted email. First, spam is unsolicited. The email recipient has not granted (verifiable) permission to the originator to send email. This characteristic alone is insufficient to classify an email message as spam, as it encompasses such legitimate email purposes as a business or personal inquiry, an electronic introduction, and generally other initial forms of contact where the sender is not known to the recipient.

Spam email is also bulk delivered, i.e., it is delivered to large numbers of recipients. However, bulk delivery alone is also insufficient to classify email as spam. Email messages that are delivered to large lists of recipients who subscribe to a newsletter or electronic mailing list are bulk-delivered, but these are not spam. The community generally regards email that is both unsolicited *and* bulk delivered as spam. The technical definition of spam offered by The Spamhaus Project summarizes this description effectively:

An electronic message is "spam" IF:

(1) *the recipient's personal identity and context are irrelevant because the message is equally applicable to many other potential recipients;*

AND

(2) *the recipient has not verifiably granted deliberate, explicit, and still-revocable permission for it to be sent.* [16]

The definition of spam can be further defined by the relationship between the sender and the recipient. If the sender has no consideration or care for the recipient, then the email message is spam.

A considerable portion of spam email serves as a snare for fraudulent activity. Spam is used to elicit user accounts and passwords as well as personal, financial, and credit card information from recipients; to entice recipients into purchasing bogus health products; to lure recipients to invest in falsely represented stocks and commodities; and to convince recipients to participate in (scam) lotteries.

The cost of sending spam to large numbers of recipients (per message sent) is extremely small compared to bulk postal delivery. Much of spam originates from programs that have been installed without authorization on inadequately protected computers. The programs are able to send email through open email relay systems throughout the Internet. Open email relay systems will forward (relay) email from any sender email address without restriction or filtering. While open email relays are widely discouraged, the number available remains more than sufficient to support the spam industry.

Email users are more aware of the dangers of spam today. Awareness combined with more widespread use of anti-spam measures in email client software and at security gateways operated by service providers and private organizations improves users' email experience by decreasing the amount of spam that is delivered to recipients. A side effect of more effective anti-spam measures is that spammers resort to sending email to more recipients. To do so, spammers aggressively search for email addresses.

4.1 How Do Spammers obtain email addresses?

Spammers obtain email addresses from a variety of sources, using many automated techniques. Some known and speculated techniques are briefly introduced here.

Spambots. Spambots are automated software designed to search web sites and harvest email addresses. Spambots vary in sophistication. Some spambots will search for HTML "mailto" tags whereas others will grab any character string containing the @ symbol.

Usenet, news groups, social networks, IRCs, and mailing list scanners. Some spammers subscribe to Usenet, news groups, chat rooms, social networks, and electronic mailing lists, then use automated software to collect email addresses from the {From:, Reply-To:, CC:} headers of email delivered by those list servers or to spam the news group or social network.

Spammer Viruses. Many viruses are programmed to access the address book on an infected computer and use the email addresses found there to propagate and infect other computers. Similar programming techniques are included in viruses (Sobig, Mimail) to collect the contents of address books from infected computers.

Directory Harvest Attacks. Using automated programming, the spammer will establish a Simple Mail Transfer Protocol (SMTP) session to an organization's email servers and attempt to construct an organization's email directory, based on positive responses to attempts to send email to recipients at that domain. Spammers use simple brute force (all possible alphanumeric combinations) or dictionary techniques (individual and concatenated common given and surnames) to generate the user element of a standard user@domain email address. The "harvest" is the list of user elements for which the SMTP server returns a positive acknowledgement when queried.

List Merchants. Parties who have accumulated millions of legitimate email addresses sell their lists to spammers.

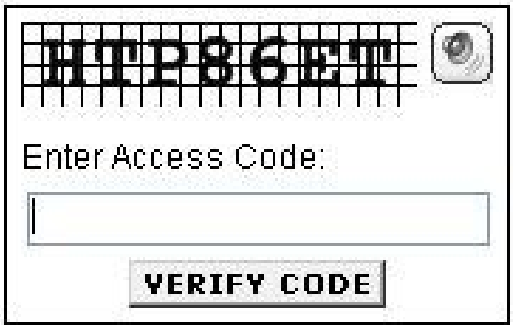
ENUM harvesting. ENUM is an application of the Dynamic Delegation Discovery System using Telephone Numbers to look up Uniform Resource Identifiers in the Domain Name System (RFC 3245, RFC 3761). ENUM is still regarded as an emerging service but industry experts have speculated that URNs could be harvested for contact information such as email addresses by a new generation of spambots.

WHOIS service. Registrants are required to provide email addresses of the registrant as well as technical and administrator contacts for a domain name. These email addresses are routinely used by law enforcement agents, network administrators, and security practitioners to identify spammers and enforce anti-spam laws. Security experts believe WHOIS is commonly used for footprinting and target acquisition as well as a source for collecting email addresses [17].

4.2 How Do Registries and Registrars Protect Against Automated Access?

Registries and registrars employ various countermeasures to thwart automated collection of domain records via query-based WHOIS services. In such cases, web user interfaces challenge the querying party with a visual display and prompt for a response that is not easily automated.

CAPTCHA [18] – Completely Automated PublicTuring Test To Tell Computers and Humans Apart – challenges the querying party with an image (typically, a distorted text) and requires that the querying party type the text in an input form.



ESP-PIX [19] challenges the querying party with a set of images and prompts the party to choose a word that applies to all the images in the set.

Some registries, registrars and resellers may employ anti-scripting and other mechanisms to thwart automated collection of registrant email addresses. Measures as simple as prompting the querying party to explicitly acknowledge having read and accepted a "conditions of use" statement through some web input object method (radio button, checkbox, menu pull down, etc.) can thwart certain automated collection efforts.

Registrars may also rate-limit WHOIS queries based on an identity such as the source IP address. Rate limiting interferes with rapid collection of email addresses. This measure can be applied to applications that access WHOIS service at TCP Port 43/NICNAME as well as web-based WHOIS services.

Some registries do not publish their zone file data to the public. While operators who are under contract with ICANN (gTLD registries) must provide free zone file data, policies concerning publication of zone file data vary by ccTLD. One TLD included in our study, the DE registry (DENIC), does not provide zone file data.

In this report, we generically apply the term **Protected-WHOIS** to these and other forms of protection against automated access.

4.3 Safeguards against email address abuse

Some registrars offer services that allow registrants to protect email addresses and other contact information against public disclosure. The registrar collects and maintains accurate domain records for the registrant who paid for the domain name registration to be registered by the proxy service, who then licenses the use of the name to the end-user. As a service to the original registrant, the registrar substitutes their own address details in

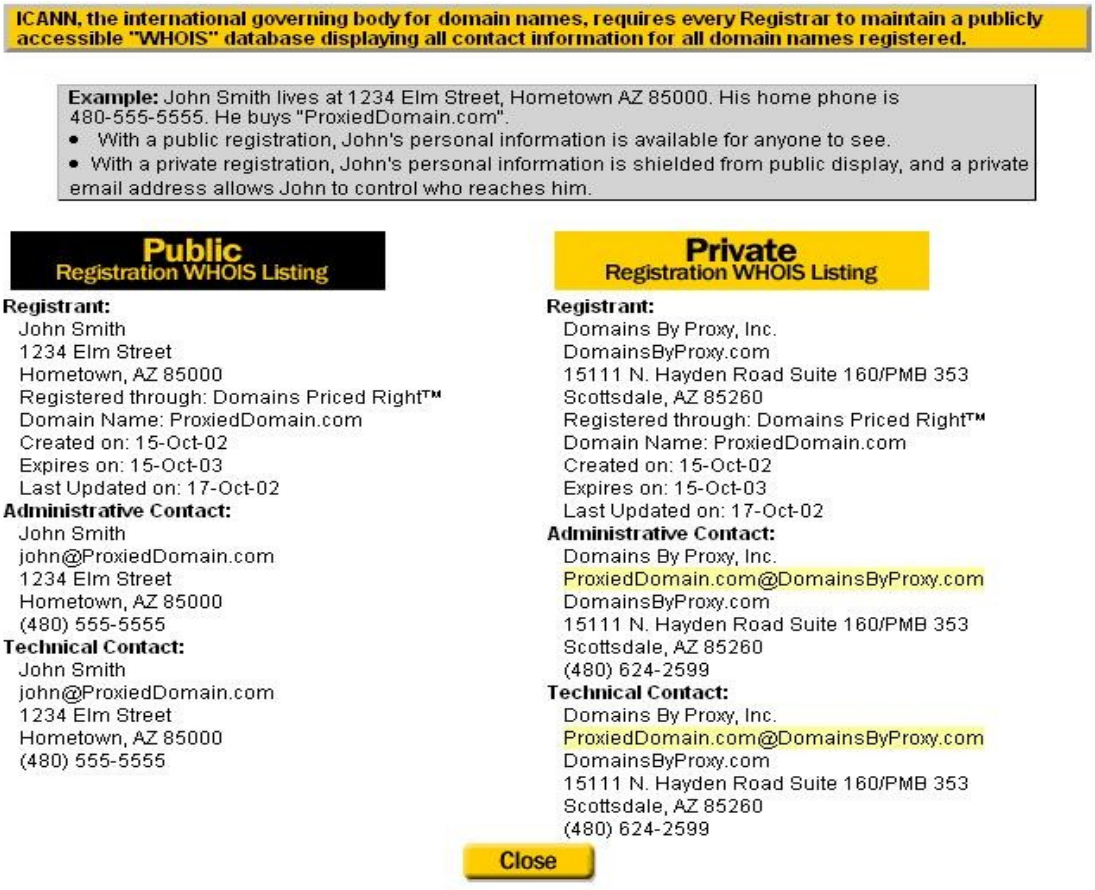
the Registrant fields when the domain name is queried using WHOIS. Spam blocking measures (e.g., spam filtering applications or gateways) are commonly incorporated into such services to further reduce spam delivered to the registrant. Thus, the benefits of this service to a registrant are twofold:

1) The email address returned in response to a WHOIS query is not the registrant's email address. If the registrant is able to prevent his own email address from being published where it is exposed to other harvesting methods, the registrant is less likely to receive spam.

2) Active anti-spam measures applied on the registrar-administered email address will mitigate spam. The effectiveness of such measures, depending on how aggressively the measure is configured, is often between 95-99%. (Note: this percentage periodically drops as spammers learn and apply techniques to evade spam detection, and rises again as anti-spam measures detect such techniques.)

Such services may also protect other registrant contact information and are advertised as methods to mitigate several forms of domain-related attacks (identity theft, fraud, stalking, harassment, data mining) [20, 21, 22, 23].

Certain registrars who offer such services provide a side-by-side comparison illustrating the differences between the contact information displayed in response to a WHOIS query. An example of such side-by-side comparisons is illustrated below [24]:



In this report, we generically apply the term **Delegated-WHOIS** to these services.

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

A US Federal Trade Commission study concluded that WHOIS is not used as a source for collecting email addresses [25]. FTC investigators wanted to determine which sources spammers considered most useful for collecting (harvesting) email addresses. The investigators planted special "undercover" email addresses in different locations on the Internet, including web pages, newsgroups, chat rooms, message boards, online directories for web pages, instant message user profiles, domain names, online resumes and online dating service personal listings.

The FTC investigators reported that very high percentages of email addresses included in web pages in the conventional user@domain format received spam, and that addresses used in email posted to newsgroups and chat rooms received spam as well. The report also made the following assertion:

Addresses posted in instant message service user profiles, "WHOIS" domain name registries, online resume services, and online dating services did not receive any spam during the six weeks of the investigation.

The FTC study is now nearly five years old. SSAC observes that registrars offer a variety of "protection" services including "WHOIS Spam Catcher" service [26], email masking [27], and proxy registration services [28]. Evidently, a market exists for the sale of services that protect email addresses from open publication in various locations, including the WHOIS. Registrars also offer anti-abuse and anti-spam measures to registrants who purchase these services.

SSAC also notes that scripts can be written in common programming and batch languages to automate command-line WHOIS applications to harvest email addresses from the domain records returned in responses to queries, although this behavior is sometimes thwarted by the deployment of rate limiting and/or IP address blacklisting schemes. SSAC also observes that the commercial mass email software market includes products that offer a domain owner email extractor[12] [29, 30].

Given the continued, global interest in defeating spam, SSAC determined that the topic of "WHOIS service and spam" merited additional attention so the committee undertook a study to determine whether spammers use WHOIS services as a means to collect email addresses for spam.

[12] One extraction program [31] is described as being "designed to search through global WHOIS database to extract owners' personal data. Current version of the program is capable of retrieving all contact e-mail addresses, phone and fax numbers, country name and expiration dates."

5. Objectives of the Study

This study attempts to answer the following questions:

1. Do spammers (or data harvesters who sell lists to 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 (Protected-WHOIS) result in a decrease in the quantity of spam delivery to a registrant?

3. For an email address that is not published anywhere other than the WHOIS, do email substitution and anti-spam services provided by registrars (Delegated-WHOIS) result in a decrease in the quantity of spam delivery to the end-user/licensee of the domain, who has retained the registrar as his agent to be the public-facing domain name registrant?

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

5. Do spammers favor one Top Level Domain over others when they attempt to collect email addresses?

This report is narrowly focused on the relationship between WHOIS and spam, and not on the larger aspect of email address harvesting by spammers. **In particular, SSAC makes no claims regarding whether the WHOIS is exclusively or even preferentially used by spammers as a source for email addresses for spam. The Committee members involved in the WHOIS study do not believe that the WHOIS service is the dominant source of spam. The Committee did not conduct any work on the proportion of spam received as a result of email addresses appearing in WHOIS responses as compared to other methods of email address discovery.**

6. Methodology

This SSAC study on WHOIS set out to establish whether the WHOIS service was a source of email addresses for spammers.

For the study, SSAC registered and monitored mail delivery to domains in four Top Level Domains: COM, DE, INFO, and ORG. These domain names were registered during the month of February 2007. SSAC then collected and analyzed all email messages delivered to these addresses for a period of approximately three months. This included the specific email addresses recorded in the domain name registration as well as any recipients to which email was delivered. Spam delivered to email addresses recorded in domain name registration records was counted separately from all other addresses that received email for the purpose of analysis. In each of the cases where a specific email address was used, commonly guessable email addresses such as "admin", "info", "user", "support" were not used. In some cases, the registrant names were common first names or last names, which were used in emails, and could have been "guessed" by a dictionary or name directory attack.

To minimize the possibility of introducing a variable (name bias) to the study sample, SSAC composed second level labels of the domain names using two techniques. We created one set of names by extracting words at random from a newspaper and concatenating several words to create a label of a minimum of ten (10) letters and a second set of names by interleaving letters and numbers to compose second-level labels (e.g., s1a2m3p4l5e). We also used randomly generated strings for the user or recipient component of each registrant email (the string that precedes the "@" sign).

The email domains were hosted on systems operated by registrars. The email addresses recorded in the domain name registration records were not published in any form or forum. In particular, they were neither used in correspondence nor published electronically in any locations on the Internet where FTC investigators planted email addresses in their 2003 study, including web pages, newsgroups, chat rooms, message boards, online directories for web pages, instant message user profiles, domain names, online resumes and online dating service personal listings. Thus, any email delivered to the email addresses recorded in the domain name registration records and not originating from the registrar was considered unsolicited. Further, since it is implausible that any party might be attempting to contact any individual having email addresses assigned in these domains, we assume that email delivered to these specific addresses was a copy of a bulk-addressed message.

This study began on 12 February 2007 and continued through 12 May 2007 (90 days). Email deliveries to recipients at each domain name were collected and counts were accumulated using automated scripts.

The SSAC conducted two sets of experiments.

Experiment 1 attempted to determine the effects on spam delivery when Protected-WHOIS or Delegated-WHOIS services are used. The cases studied in this set of experiments are as follows:

Case #1: Five (5) domain names were registered in the COM and INFO registries with neither **Protected-WHOIS** nor **Delegated-WHOIS**.

Case #2: Five (5) domain names were registered in the DE and ORG registries with **Protected-WHOIS** but not **Delegated -WHOIS.**

Case #3: This case used the same TLD registries as Case #1 with **Delegated-WHOIS** service offered by the registrar but not **Protected-WHOIS**[13]**.**

Case #4: This case used the same TLD registries as Case #2 with both **Protected-WHOIS** and **Delegated-WHOIS** services available to the registrant via the registry or registrar[14].

Experiment 2 attempted to classify the kinds of spam delivered to email addresses at the domain name. For this study, 15 additional domains were included in the analysis to measure the incidence of spam emails arriving at either the email address recorded in the registration record and to any recipient email address at the domain name. For this study, neither **Protected-WHOIS** nor **Delegated-WHOIS** were used. These names were not used in other parts of the study.

[13] INFO rate limits WHOIS queries based on source IP address at the registry web site for port 43 but not for web based queries. COM runs a "thin" registry so WHOIS queries are made directly to the registrar's web site.

[14] ORG rate limits WHOIS queries based on source IP address at the registry web site for both port 43 and web based queries. The Protected-WHOIS service used by the DE registry challenges visitors with a Conditions of Use which requires an explicit (accept) response from the requestor.

7. Effect of Protected & Delegated WHOIS Services

In this section, we summarize the results of the studies in tabular and graphical formats. The actual second-level labels used in the study are not presented here (SSAC may use these for continued testing or for other as-yet-to-be-determined purposes); rather, we use the representative string "RandomlyChosenName" concatenated with a number, e.g., RandomlyChosenName1. We separate spam delivered to the email address recorded in the registration records (denoted in the tables as *Published Address*[15]) from email delivered to all other recipients at the domain name (denoted in the tables as *All other recipient addresses*). Readers should take note that in some cases, the same second-level labels have been registered in multiple TLDs (e.g., RandomlyChosenName1.ORG and RandomlyChosenName1.DE). This was intentional.

7.1 Case #1, Neither Protected-WHOIS nor Delegated-WHOIS used

For this case, SSAC registered domain names with generic TLDs (INFO and COM) and used neither Protected WHOIS nor Delegated-WHOIS services.

***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%**

[15] I.e., randomlychosenusername@randomlychosenname.<tld>

In nearly all cases, the volumes of spam delivered to recipients in these domain names were extraordinarily large compared to all study cases where one or multiple protection services were used.

The number of spam messages delivered to two email addresses is atypical from others included in this case. Our data provide no insight into why the email address RandomlyChosenName10.INFO received a small volume of spam compared to other names in this study. We observe that multiple parties collect email addresses for use in delivering spam and that all or only parts of email lists are sold to multiple parties who send spam messages. It is possible that some spammers use every email address they can purchase, whereas others may be resource-limited (e.g., they may not use very large botnets to send spam), and may send fewer spam messages). This and other variables are outside the control of this study and outside the scope as well.

While the majority of domain names registered under COM did receive more spam than names registered under INFO, RandomlyChosenName9.INFO affects the mean volume of spam delivered to the names registered under INFO and its deviation from the mean is unique in this sample. A larger sample of email addresses and a study across a greater number of TLDs is necessary to determine whether the amount of spam delivered to RandomlyChosenName9.INFO is a statistical anomaly or whether spammers favor one TLD over another. The majority of the results, however, suggest that the TLD itself does matter to spammers as they attempt to harvest email addresses.

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

For this case, SSAC registered domain names with a gTLD (ORG) and a ccTLD (DE). Here, we took advantage of the Protected-WHOIS service offered but did not use a Delegated-WHOIS service.

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%**

On average, two orders of magnitude less spam email messages were delivered to recipients in these domains than those in Case #1; specifically, where domains in Case #1 received thousands or tens of thousands counts of spam, the registrant's email address in the majority of domains in Case #2 received only tens or hundreds.

The results for some email addresses are atypical and unexpected. However, our data provide no insight into why these addresses received a higher volume of spam than other names in this study group. One possibility is that these are examples of situations where a user name was derived by brute-forced or guessed, and once it was used with success, the email address was added to a spam list that was used on more than one occasion and possibly by more than one spammer.

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

For this case, SSAC registered domain names with generic TLDs (INFO and COM) and took advantage of the Delegated-WHOIS service offered but did not use Protected WHOIS services.

***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%**

On average, three orders of magnitude less spam was delivered to recipients in these domains than to recipients in the domains in Case #1, and (on average) the volume of spam delivered to domains in Case #3 was an order of magnitude smaller than the spam volume delivered to domains in Case #2. This suggests that a private registration (and associated anti-spam measures) may be somewhat more effective in combating spam than measures to prevent automated querying of WHOIS for email addresses.

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

SSAC registered domain names with a generic TLD (ORG) and a ccTLD (DE) and took advantage of the Protected-WHOIS and Delegated-WHOIS services offered. As the table illustrates, virtually no spam email messages were delivered to the email address recorded in the registration records from email delivered to all other recipients at the domain name.

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%**

7.5 Comparison of Results across Cases

The results of the four cases are shown in the graph below. Specifically:

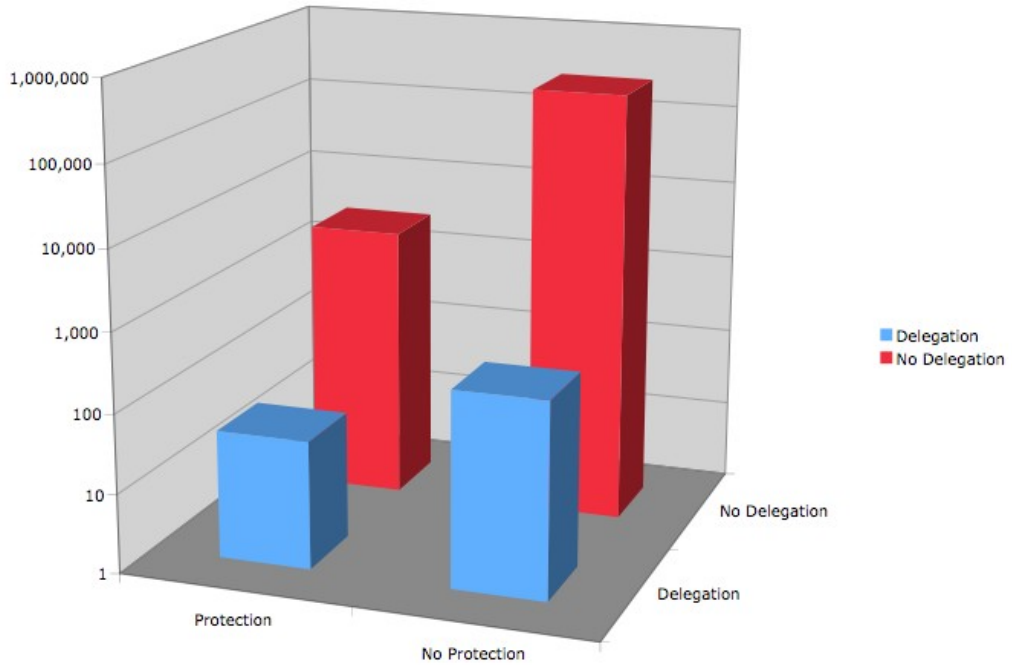
1. Unprotected registrant email addresses received significant amounts of spam.

2. When a domain name is registered at a registry/registrar that offered protected-WHOIS without Delegated-WHOIS, our study indicates it is possible to achieve two orders of magnitude better defense against spam.

3. When a domain name is registered at a registry/registrar that did not offer Protected-WHOIS but offered Delegated-WHOIS, our study indicates it is possible to achieve three orders of magnitude better defense against spam.

4. When a domain name is registered at a registry/registrar that offered Protected-WHOIS *and* Delegated-WHOIS, our study indicates it is possible to achieve close to four orders of magnitude better defense against spam.

Although the data suggests Protected-WHOIS is somewhat more effective than Delegated-WHOIS, our study is not detailed enough to provide a firm basis for such a conclusion.


1,000,000
100,000
10,000
1,000
100
10
1
Protection
No Protection
No Delegation
Delegation
Delegation
No Delegation

8. Analysis of Spam Delivered to Domains Studied

We conducted a second experiment to classify the kinds of spam delivered to email addresses at the domain name.

We grouped spam into categories familiar to many email users, using the following spam assessment criteria:

- Keywords in email headers and message bodies that associate a message with a particular kind of offer or scam
- Hyperlinks that led to redirect pages (interpreted as a phishing site)
- Matches of domains and hyperlinks in messages to known phishing domains

The categories we most frequently encountered in the spam delivered to the addresses used in the study are listed below:

- Direct marketing of discounted products such as watches, printer ink/toner
- Pharmaceuticals and weight loss products
- Discounted commercial software
- Phishing
- Male enhancement and ED products
- Financing offers
- Mortgage offers
- Stock market offers
- Image and other spam

From the spam received, we observe the following:

- Contrary to popular belief, the spam is not limited to sex and pornography. From the spam received at email addresses monitored during the study, we note that approximately 43% of spam messages seek to lure recipients to sites offering illegal pharmaceuticals, bogus products, and unlicensed software.

- While spam associated with known phishing sites accounts for only 9% of overall spam, including spam associated with refinancing, mortgage, and stock scams as possible phishing lures increased the percentage of spam that may be used to obtain credit and financial account information to over 40%.

SSAC offers these observations as complementary information to the studies performed. Simply stated, having collected many samples of unsolicited bulk email, we chose to analyze spam delivered to email addresses published via the WHOIS service to see if any patterns or anomalies might emerge. At this point, we draw no conclusions from our data other than to observe (and corroborate similar claims) that spam is increasingly used as a vehicle to support criminal activities.

	Protected-WHOIS + Delegated-WHOIS	**Protected-WHOIS but *NO* Delegated-WHOIS**	***NO* Protected-WHOIS but Delegated-WHOIS**	***NO* Protected-WHOIS *NO* Delegated-WHOIS**
Category	**# of spam messages delivered**			
Watches, Ink	10	518	45	42194
Pharmacy, Weight Loss	6	605	78	52661
Software	3	173	34	35876
Phishing	3	86	6	12121
Viagra	2	345	28	36391
Finance	7	403	14	25490
Mortgage	5	288	34	31076
Stock Scam	1	29	4	6833
Undetermined	4	432	40	28527
	41	2879	284	271170
Category	**Percent of spam messages delivered per category**			
Watches, Ink	24.4%	18.0%	16.0%	15.6%
Pharmacy, Weight Loss	14.6%	21.0%	27.4%	19.4%
Software	7.3%	6.0%	12.0%	13.2%
Phishing	7.3%	3.0%	2.0%	4.5%
Viagra	4.9%	12.0%	10.0%	13.4%
Finance	17.1%	14.0%	5.0%	9.4%
Mortgage	12.2%	10.0%	12.0%	11.5%
Stock Scam	2.4%	1.0%	1.5%	2.5%
Undetermined	9.8%	15.0%	14.0%	10.5%

9. Findings and Conclusions

The Committee offers the following findings for consideration:

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

Finding (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.**

Finding (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.

Finding (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.

On the basis of these Findings, the Committee draws the following conclusions:

Conclusion (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.

Conclusion (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.

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

.

Conclusion (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.

.

Conclusion (5) SSAC concludes that further studies may be needed to investigate whether spammers have preferential targets. Suggested 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?

References

[1] RFC 812, NICNAME/WHOIS
http://www.faqs.org/rfcs/rfc812.html

[2] RFC 954, NICNAME/WHOIS
http://www.ietf.org/rfc/rfc954.txt

[3] RFC 3912, WHOIS Protocol Specification
http://www.ietf.org/rfc/rfc3912.txt

[4] ICANN Registrar Accreditation Agreement 17 May 2001
http://www.icann.org/registrars/ra-agreement-17may01.htm#3

[5] ICANN WHOIS Data Reminder Policy 16 June 2003
http://www.icann.org/registrars/wdrp.htm

[6] ICANN WHOIS Data Problem Reporting System
http://wdprs.internic.net/

[7] WHOIS Data Accuracy Program 27 April 2007
http://www.icann.org/WHOIS/WHOIS-data-accuracy-program-27apr07.pdf

[8] ICANN WHOIS Marketing Restriction Policy 12 August 2004
http://www.icann.org/tlds/agreements/net/appendix5.html

[9] Final Task Force Report on WHOIS Services 12 Mar 2007 GNSO WHOIS Task Force
http://GNSO.icann.org/issues/WHOIS-privacy/WHOIS-services-final-tf-report-12mar07.htm

[10] Email message from Ross Rader to the registrars mailing list 28 Nov 2005
http://GNSO.icann.org/mailing-lists/archives/registrars/msg03687.html

[11] 90% of E-Mail Will Be Spam By Year's End, *Information Week* 22 Feb 2007
http://www.informationweek.com/news/showArticle.jhtml?articleID=197008209

[12] Spam Volume Hits Record High, Marshall, Ltd. 21 Feb 2007
http://www.marshal.com/pages/newsitem.asp?article=135

[13] CommTouch Spam Lab Online Statistics 22 Jun 2007
http://www.commtouch.com/Site/Resources/statistics.asp

[14] Postini StatTrack 22 Jun 2007
http://www.postini.com/stats/index.php

[15] Spam Volume – Swivel 22 Jun 2007
http://www.swivel.com/graphs/show/9135865

[16] Definition of Spam, SpamHaus.org
http://www.spamhaus.org/definition.html

[17] *Hacking Exposed*, by Stuart McClure, Joel Scambray, & George Kurtz, Osborne Press, ISBN 0-07-212127-0

[18] The CAPTCHA Project
http://www.captcha.net/

[19] ESP-PIX
http://www.captcha.net/cgi-bin/esp-pix

[20] Domains By Proxy: Private Registrations
http://domainsbyproxy.com/

[21] Private Domain Registration
http://www.actnowdomains.com/private-domain-registration.htm
[22] SpamShield/WHOIS Privacy
http://www.mydomain.com/domains_privacypost.php?s_kwcid=private%20domain%20registration|671718391
[23] ID Domain Privacy
http://www.iddp.net/
[24] Domains By Proxy: WHOIS Example
http://www.domainsbyproxy.com/popup/WHOISexample.aspx?app%5Fhdr=0&ci=5165
[25] US Federal Trade Commission Spam Alert
http://www.ftc.gov/bcp/conline/pubs/alerts/spamalrt.htm and
http://www.security.iia.net.au/downloads/spamalrt-ftc.pdf
[26] Jump Domain: WHOIS Spam Catchter
https://domains.jumpdomain.com/index.cgi?page=spam_catcher.tmpl
[27] Network Solutions: Private Registrations
http://www.networksolutions.com/domain-name-registration/private.jsp
[28] ActiveDOMAIN.com: Private WHOIS Protection service
http://www.active-domain.com/WHOIS-proof.htm
[29] Atomic WHOIS Explorer: domain owner email address extractor
http://www.massmailsoftware.com/WHOIS/
[30] Email Spider by EmailSmartz
http://WHOIS-email-extractor.qarchive.org/
[31] WHOIS Extractor by WebExtractor Systems
http://www.programurl.com/WHOIS-extractor.htm
[32] .BIZ Agreement Appendix 5 WHOIS Specifications 8 December 2006
http://www.icann.org/tlds/agreements/biz/appendix-05-08dec06.htm
[33] .ORG Agreement Appendix 5 WHOIS Specifications 8 December 2006
http://www.icann.org/tlds/agreements/org/appendix-05-08dec06.htm
[34] .net Registry Agreement: Appendix 5
http://www.icann.org/tlds/agreements/net/appendix5.html
[35] Comments from the American Intellectual Property Law Assocation, regarding the preliminary reports of the WHOIS Task Forces
http://www.aipla.org/Content/ContentGroups/Issues_and_Advocacy/Comments2/Domain_Name_Comments/WHOISComments.pdf
[36] *Incident Response: Investigating Computer crime*, Kevin Mandia & Chris Procise, Osborne Press, ISBN 0-07-213182-9
[37] How the FTC uses WHOIS Data
http://www.icann.org/presentations/mithal-WHOIS-workshop-24jun03.pdf
[38] The Importance of WHOIS data bases for spam enforcement
http://www.icann.org/presentations/opta-mar-26jun06.pdf
[39] FAQ: How do spammer's get people's email addresses?
http://www.faqs.org/faqs/net-abuse-faq/harvest/

Appendix A. Members of the SSAC

Alain Aina, Consultant

Jaap Akkerhuis, NLnet Labs

KC Claffy, CAIDA

Steve Crocker, Shinkuro (Chairman)

James Galvin, (Exec)

Daniel Karrenberg, RIPE/NCC

Johan Ihrén, Autonomica

Rodney Joffe, Centergate

Mark Kosters, ARIN

Ram Mohan, Afilias

Russ Mundy, SPARTA, Inc

Frederico Neves, Registro Brazil

Jon Peterson, NeuStar

David Piscitello, ICANN SSAC Fellow

Ray Plzak, ARIN, Vice Chairman

Mike St. Johns

Doron Shikmoni, ForeScout, ISOC-IL

Bruce Tonkin, Melbourne IT

Paul Vixie, ISC

Suzanne Woolf, ISC

Acknowledgements

The committee thanks Ram Mohan who led the effort to craft and conduct this study, and to Afilias for providing staff and resources for this study and in particular acknowledges the contributions of Roland LaPlante.

Appendix B. Excerpt from U.S. FTC Commission Study, *Email Address Harvesting: How Spammers Reap What You Sow*

From http://www.security.iia.net.au/downloads/spamalrt-ftc.pdf:

To find out which fields spammers consider most fertile for harvesting, investigators "seeded" 175 different locations on the Internet with 250 new, undercover email addresses. The locations included web pages, newsgroups, chat rooms, message boards, and online directories for web pages, instant message users, domain names, resumes, and dating services. During the six weeks after the postings, the accounts received 3,349 spam emails. The investigators found that:

- 86 percent of the addresses posted to web pages received spam. It didn't matter where the addresses were posted on the page: if the address had the "@" sign in it, it drew spam.

- 86 percent of the addresses posted to newsgroups received spam.

- Chat rooms are virtual magnets for harvesting software. One address posted in a chat room received spam nine minutes after it first was used.

Addresses posted in other areas on the Internet received less spam, the investigators found. Half the addresses posted on free personal web page services received spam, as did 27 percent of addresses posted to message boards and nine percent of addresses listed in email service directories. **Addresses posted in instant message service user profiles, "WHOIS" domain name registries, online resume services, and online dating services did not receive any spam during the six weeks of the investigation.**