20 June 2008

SAC 033: Domain Name Registration Information and Directory Services
(Complements SAC 027, SSAC Comment to GNSO regarding WHOIS Services)

In SAC027, Comment to GNSO regarding WHOIS studies, SSAC asserts that the limitations of the WHOIS protocol and variability among WHOIS implementations and services contribute to the poor quality of domain name registration data currently available. SSAC further suggests that ICANN community should adopt *an* Internet standard directory service as an initial step toward deprecating the use of the WHOIS protocol in favor of a more complete directory service. This companion document to SAC027 provides complementary rationale and to adds clarity to SSAC's prior comment to the GNSO.

Improving Accountability and Access control

WHOIS services in several forms have served the Internet community for many years and in more ways than the designers of the original NICNAME protocol and even early WHOIS services envisioned. Access to and applications of domain name registration information have changed considerably as well. SAC 023[1] enumerates several ways WHOIS services are used today (reproduced here for the reader's convenience):

- 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

[1] SAC023, Is the WHOIS Service a Source for email Addresses for Spammers? http://www.icann.org/committees/security/sac023.pdf

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[2].

- To contact a domain name registrant on matters related to the protection and enforcement of intellectual property rights[3].

- 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[4]

- 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 enforcement agents use WHOIS to identify points of contact[5]

- 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[6].

- To investigate spam, law enforcement agents look to the WHOIS database to collect information on the website advertised in the spam[7].

- To collect or "farm" email addresses for the purpose of delivering unsolicited electronic mail[8].

---

[2] WHOIS Data Reminder Policy
http://www.icann.org/registrars/wdrp.htm

[3] 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

[4] *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*.

[5] *Incident Response: Investigating Computer crime*, Mandia & Procise, Osborne Press, ISBN 0-07-213182-9, pp 435-439.

[6] *How the FTC uses WHOIS Data,*
http://www.icann.org/presentations/mithal-WHOIS-workshop-24jun03.pdf

[7] *The Importance of WHOIS data bases for spam enforcement*,
http://www.icann.org/presentations/opta-mar-26jun06.pdf

[8] FAQ: How do spammer's get people's email addresses?
http://www.faqs.org/faqs/net-abuse-faq/harvest/

This list contains acknowledged uses and abuses of domain name registration information and is not exhaustive. However, *the list contains appropriate uses and abuses, illustrating that accountability and access control should be improved.*

Improving Accuracy

Various studies have been conducted to assess the quality and accuracy of domain name registration information and ICANN is currently engaged in several compliance projects to improve data accuracy and WHOIS service availability[9]. SSAC's study of ways that WHOIS could be used for information gathering by Internet attackers[10] provides one data point regarding incomplete and inaccurate data: of the 4444 registration records used in the study:

- 10% were missing admin contact name (439 records)
- 11% were missing admin contact email (502 records)
- 12% were missing admin contact address (514 records)
- 24% were missing registrant phone # (1039 records)†
- 60% were missing admin contact fax (2647 records) †
- 87% were missing registrant fax # (3867 records) †

† (Optional field)

Other studies demonstrate that miscreants intentionally populate registration records with invalid data on a large scale. Edelman's study[11], while conducted in 2002, contains statistical findings that are representative of large-scale intentional submission of invalid WHOIS data. In the study, a single entity registered over 2700 domains through multiple registrars using a variety of contact locations. A 2005 U.S. Government Accounting Office (GAO) study[12] found that among the sampling studied, 2.31 million (5.14 percent) of domains had patently false WHOIS contact data, and 1.64 million (3.65 percent) of domains had incomplete information in one or more required fields. Extrapolating these percentages to a current estimate of registered domains suggest that over 8 million registration records (see table extracted from report, below):

---

9 Update: ICANN projects underway to improve Whois accuracy http://www.icann.org/announcements/announcement-2-21dec07.htm

10 SAC 014, Information Gathering Using Domain Name Registration Records, http://www.icann.org/committees/security/information-gathering-28Sep2006.pdf

11 Large-Scale Intentional Invalid WHOIS Data, http://cyber.law.harvard.edu/archived_content/people/edelman/invalid-whois/

12 US GAO Report: Internet Management: Prevalence of False Contact Information for Registered Domain Names, http://www.gao.gov/new.items/d06165.pdf

Prevalence of Patently False Contact Information (in millions; percentages in parentheses)

	Registrant			Administrative contact			Technical contact		
Data	**.COM**	**.ORG**	**.NET**	**.COM**	**.ORG**	**.NET**	**.COM**	**.ORG**	**.NET**
Not patently false	33.13 (92.65)	3.29 (93.69)	5.34 (94.26)	31.90 (89.20)	3.15 (89.77)	5.21 (91.88)	32.18 (89.98)	3.18 (90.63)	5.29 (93.37)
Patently false	1.18 (3.30)	0.10 (2.97)	0.05 (0.89)	1.86 (5.20)	0.22 (6.25)	0.18 (3.13)	1.50 (4.18)	0.19 (5.51)	0.16 (2.76)
Incomplete	0.27 (0.76)	0.07 (2.09)	0.17 (2.98)	0.83 (2.31)	0.11 (3.09)	0.18 (3.13)	0.91 (2.54)	0.10 (2.97)	0.11 (2.01)
Unable to access Whois data	1.18 (3.30)	0.04 (1.25)	0.11 (1.86)	1.18 (3.30)	0.04 (1.25)	0.13 (2.24)	1.18 (3.30)	0.04 (1.25)	0.13 (2.24)

Source: GAO analysis of test results.

Note: Margin of error is ±5 percent or less at the 95 percent confidence level. Some domain names contained both patently false and incomplete information and so percentages do not add up to 100.

Previous SSAC studies explain that domain records often contain "stale" contact information and that this information can cause difficulties when registrants seek to renew domain names or modify DNS information[13]. Domain contact information may remain unchanged from the original party who registered the name long after that party has ceased to work for or on behalf of the business or organization for which he registered the domain. Stale information may prevent registrars from notifying a registrant that a domain registration is about to expire or that changes (possibly unauthorized) have been made to his DNS infrastructure, and may also result in hijacking or a dispute over the "ownership" of a domain.

Recently, APWG's Global Phishing Survey: Domain Use and Trends in 2007 mentions that "registration information is often faked or obscured by proxy services to abet phishing"[14]. Lastly, reluctance to include personal information in public databases and a strong desire to avoid publishing email addresses in a potential source for spam harvesting are cited as motives for individuals and business operators to submit incorrect information[15].

These examples illustrate *that data that are submitted as registration information are incomplete and inaccurate. This underscores the need for improvements in accuracy.*

[13] SAC010: Renewal Considerations for Domain Name Registrants http://www.icann.org/committees/security/renewal-advisory-29jun06.pdf

[14] Global Phishing Survey: Domain Use and Trends in 2007, http://www.apwg.org/reports/APWG_GlobalPhishingSurvey2007.pdf

[15] False domain info may mean jail, http://www.wired.com/politics/law/news/2004/02/62198

Improving WHOIS

Features and characteristics that are common to many proprietary and public directory service (DS) applications may help to reconcile the deficiencies SSAC have identified in WHOIS services. In this section, we discuss these features with the goal of stimulating discussion. Specifically, we anticipate that readers of this paper will ask:

- Is this feature present (uniformly) in WHOIS services?
- Does this feature remedy an acknowledged deficiency in WHOIS services?
- Does this feature enhance the WHOIS user experience?
- Does this feature improve the security, stability and reliability of WHOIS services?
- Does this feature enhance the quality of registration information?
- Can this feature assist ongoing compliance and accuracy activities?
- What communities benefit from inclusion of this feature?

Many organizations consider these and similar, additional questions when they develop requirements for DS applications. The resulting requirements statement is then used to assess cost, choose products, and estimate development and deployment timeframes.

It should be noted that the directory service features and characteristics SSAC believes would prove beneficial to the community are not exclusive to the IETF Internet Registry Information Service (IRIS), nor do they require the implementation of the Cross Registry Information Service Protocol (CRISP). SAC 027 recommended IRIS/CRISP because they are completed, available for review and deserve careful consideration. However, SSAC observes that the features IRIS/CRISP offer are common to these proprietary and open source directory applications, including:

- OSI's X.500 Directory Service
- Various proprietary (Apple, CA, IBM/Tivoli, Oracle) and open source (OpenLDAP, ApacheDS) implementations of the Lightweight Directory Access Protocol (LDAP), and
- Microsoft Active Directory

We mention these for the benefit of readers who may be interested in comparing various directory services applications only and not as candidates for an Internet Directory per se. The following features are commonly associated with a directory service (DS) and found in many of the aforementioned DS applications:

Directory database. Each entry in the database is distinguished using an identity (e.g., a domain name) as its root. Additional data objects are associated with that domain name (registration information). [Note: SSAC observes that multiple autonomously operated domain name registration databases exist in both "thick" and "thin" forms, but Internet users tend to view the WHOIS conceptually as a single database.]

Data schema. A conceptual model for data (a schema) is defined for the database. The schema identifies object classes, object attributes, name bindings and knowledge or namespaces for data. Typically, schemas for data are enforced: it is not possible to inject invalid or malicious data into an object or record. White and yellow pages models may be defined for data in the DS. [Note: One can argue that a schema for WHOIS exists but its definition is not as rigorous as one commonly finds in a DS. Some object classes are defined externally, through RFCs and IANA assigned numbering, such as IP addresses and domain names themselves. Other data objects can be claimed to fall into familiar object classes, e.g. telephony and fax numbers. SSAC notes that WHOIS services do not uniformly detect and reject invalid data submitted as registration information elements.]

Authentication framework. An authentication framework accommodates a diverse set of authentication methods (single and multi-factor). DS applications are typically able to communicate with external authentication servers using such protocols as RADIUS. The framework can often allow an organization to associate an authentication method to a group of users. This allows organizations to employ stronger authentication methods to sensitive data and simpler authentication methods for access to public or less sensitive data. The value of authenticating users, even when they access public data, is to allow an organization to audit user activity.

Authorization framework. DS applications control access to information by granting permissions (privileges) to approved, and commonly, authenticated users or groups. Mechanisms for controlling access permissions are often called access controls. Many DS applications provide *fine granularity*, meaning that a per-object access permissions are applied on a per-object basis (as opposed to a any record or any object basis).

Auditing framework. An auditing function records data object access activities. Many DS applications provide auditing with data object granularity (can audit not only access to a record but given elements of a record).

Accuracy framework. Accuracy processes assure that data are not corrupted, lost or altered without permission or detection. They are also used to synchronize data across distributed databases (e.g., thin registry models) are used. Accuracy frameworks may also provide archival (escrow) facilities.

Availability framework. This aspect of a DS has both a definition and monitoring component. The definition component identifies metrics for service availability and service quality, e.g., Mean Time Between Service Outages (MTBSO) and Mean Time to Restore Service (MTTRS).

Conclusions

This informational paper complements SAC027. It attempts to add clarity to recommendations and comments in SAC 027 and makes no additional recommendations. SSAC suggests that this informational paper could serve as the basis for discussions related to future WHOIS features and services.