SSAC
ICANN Security and Stability Advisory Committee

Information Gathering Using Domain Name Registration Records

David M Piscitello

SSAC
ICANN Security and Stability Advisory Committee

Objectives

Approximate the extent to which personal contact information can be extracted from Domain Name Registration Records

SSAC
ICANN Security and Stability Advisory Committee

What is *Personal* Contact Information?

- For this study, personal contact information is *sufficient* attributes to feel confident that
  - The registrant is an individual, or an individual operating a home business, not a "business"
  - It is possible, using the information collected, to speak with or visit the individual at his or her residence, e.g., make personal contact

SSAC
ICANN Security and Stability Advisory Committee

Methodology

- Apply information gathering techniques used by computer network attackers

1. Begin with a set of potential targets
   - ~5000 registration records filtered from over 2 million
   - Filter (search argument) was "Philadelphia PA"
2. Use publicly accessible resources to collect bits and threads of data from registrant and administrative contact information
3. Piece data together until there is ***high confidence*** that a given registration record contains personal contact information

- Similar methods and resources are used by law enforcement agencies

SSAC
ICANN Security and Stability Advisory Committee

Resources used

- Domain name registration records acquired in bulk the using Whois protocol
- Real estate database (trulia.com)
- Internet telephone directory (whitepages.com)
- Search engines (Google, Yahoo!)
- Aerial photographs (GoogleEarth)
- E-maps (Map Quest)
- Companies and Industries directory (hoovers.com)
- Personal familiarity with geographic region
- Web site hosted at registered domain name

SSAC
ICANN Security and Stability Advisory Committee

Classifying results

- Personal contact
  - Individual: the registrant name is an individual's name and other fields contain personal contact information
  - Home-operated business: the registrant name is not personal name but other fields
- Business contact
  - The registrant name identifies a company and other fields indicate this is a business with many employees
- Domain name business
  - Secondary market, tasting, monetization
- Domain name proxy agent
  - Registrant fields contain service provider information
- Inconclusive data
  - Study of registrant data fail to provide convincing number of matches

SSAC
ICANN Security and Stability Advisory Committee

Classifying a record as "containing a personal contact"

Registrant Name is a Personal Name (first, surname)

Registrant Phone# is a residential listing (reverse phone# search)

Registrant's neighbors are individuals (search neighbors in WP)

Registrant Address contains an apartment number

Real estate listings near registrant address are residential

Registrant phone # is a cell phone (reverse phone# search)

Registrant address "looks like" a residence (aerial photograph)

Registrant's neighborhood is known to be residential (familiarity with region)

Registrant's web site reveals additional personal information

The more criteria that are matched, the higher the confidence that the registrant information identifies an individual

SSAC
ICANN Security and Stability Advisory Committee

Classifying a record as "containing a (domain) business contact"

Registrant Name is a Public Corporation or fictitious name (dba)

Registrant Phone# is a business listing (reverse phone# search)

Registrant's neighbors are businesses (search neighbors in WP)

Registrant Address contains an suite number

Real estate listings near registrant address are businesses

Registrant phone # is a toll-free number (reverse phone# search)

Registrant address "looks like" a business (aerial photograph)

Registrant's neighborhood is known to be business (familiarity with region)

Registrant's web site suggests that business is operated from home office

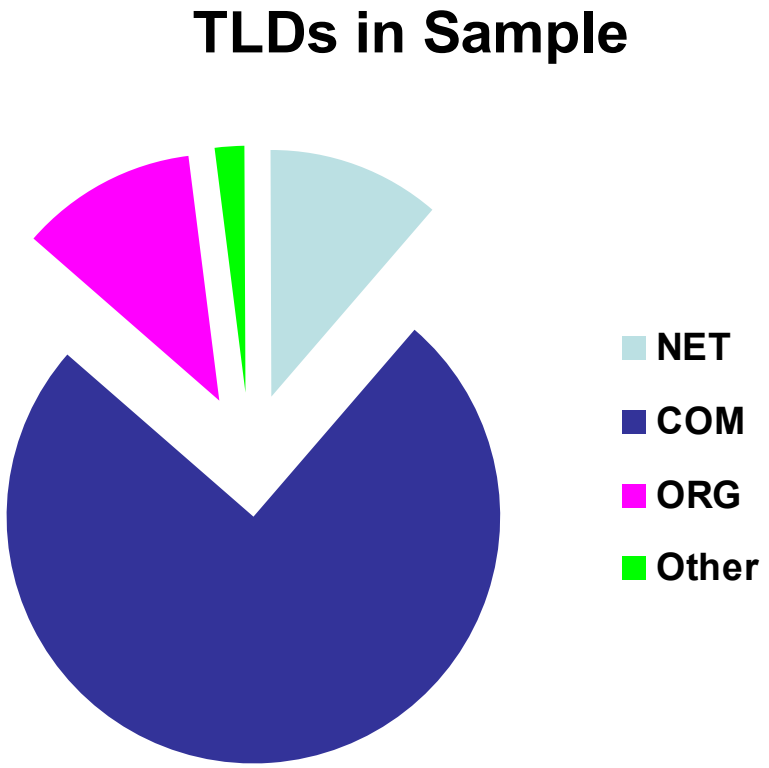
Web site identifies registrant as domain name business, ISP, reseller…

SSAC
ICANN Security and Stability Advisory Committee

TLDs in Sample

NET
- 505 domain names

- COM
  - 3334 domain names
- ORG
  - 520 domain names
- Other
  - 85 domain names

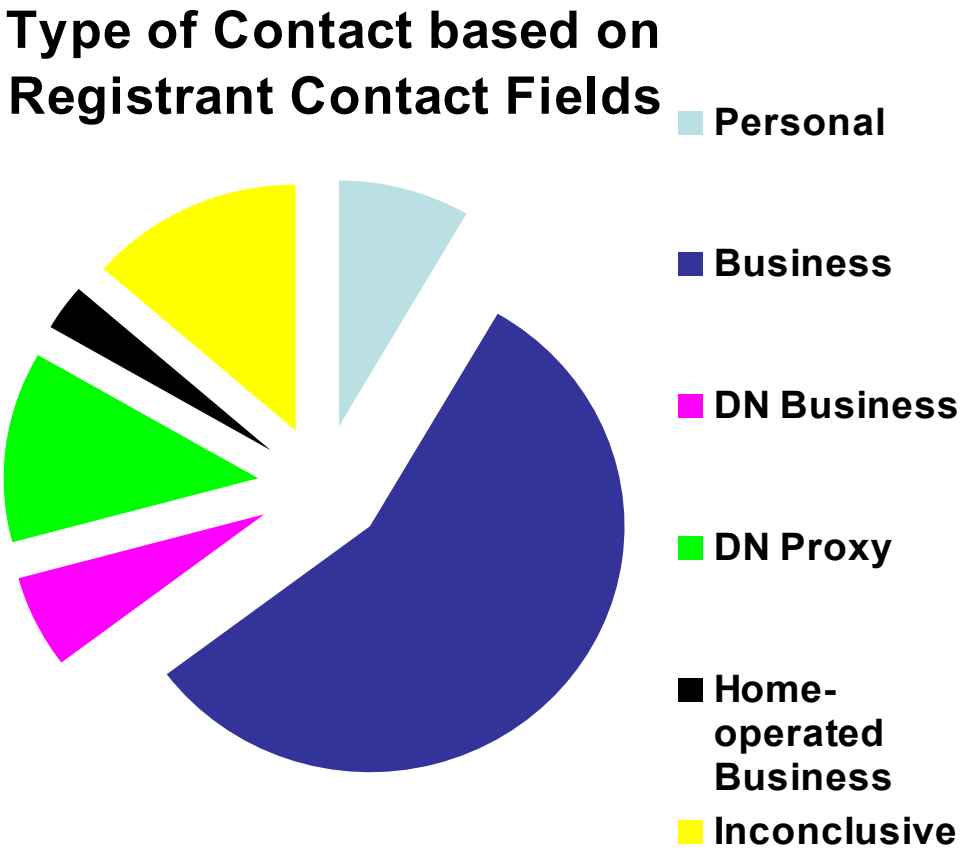


Approximately 4400 of 5000 filtered records had sufficiently accurate data to be useful in the study

SSAC
ICANN Security and Stability Advisory Committee

Findings
(Registrant Contact Fields Only)

- Personal contacts
  - 377 records, 9%
- Business contacts
  - 2501 records, 56%
- Domain name business
  - 269 records, 6%
- Domain name proxy service
  - 562 records, 13%
- Home-operated business
  - 138 records, 3%
- Inconclusive
  - 604 records, 14%



SSAC ICANN Security and Stability Advisory Committee

Simplified Findings (Registration Fields Only)

- Remove inconclusive and proxied domain names
  - *Since one cannot deduce whether the contact is business or individual from available data, these records bias the result*

Classification	Per cent of records
Combine personal contacts and home-operated businesses (515 records)	13.4%
Business contacts (2501 records)	65.1%
Domain name businesses (821 records)	21.6%

SSAC ICANN Security and Stability Advisory Committee

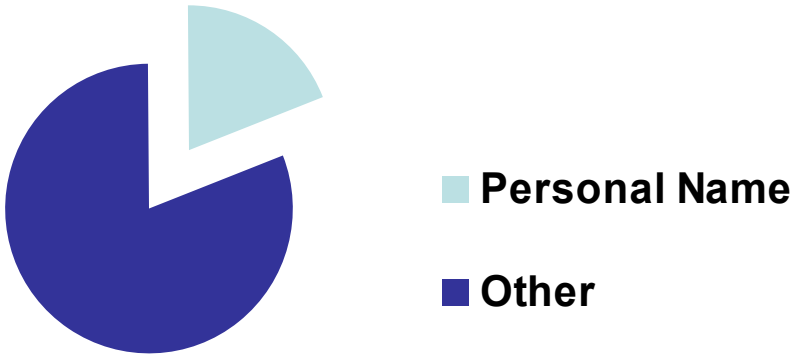
Digging Further

- *If we look at both the registrant contact information and the administrative contact information, what do we find?*
- Of the 377 records that contain personal contacts
  - 347 contain the same contact information in admin contact fields
  - 13 contain information that identify a different individual
  - 8 contain information that identifies a business contact
  - 9 have inconclusive (incomplete) data
- Of the 138 records that contain home-business contacts
  - 125 contain the same contact information in admin contact fields
  - 3 contain information that identify a different individual
  - 4 contain information that identifies a business contact
  - 5 have inconclusive (incomplete) data

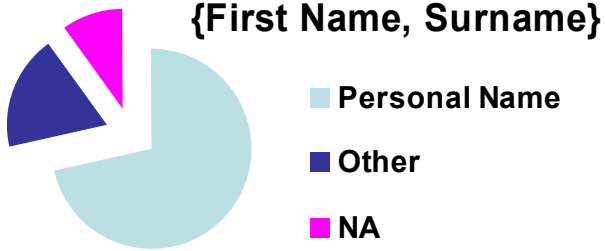
SSAC
ICANN Security and Stability Advisory Committee

Individual Names in Contact Fields

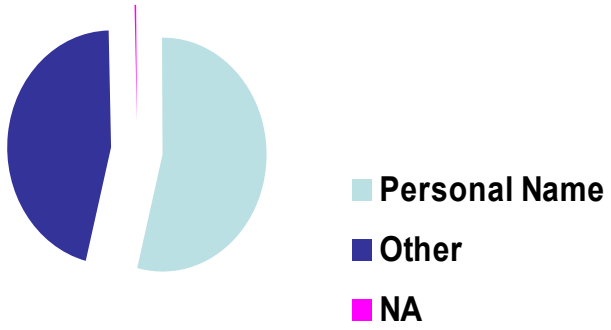
Registrant Name Contains {First Name, Surname}



Admin Contact Name Contains {First Name, Surname}



Tech Contact Name Contains {First Name, Surname}



SSAC
ICANN Security and Stability Advisory Committee

Incomplete records

- Of the 4444 records used in the study
  - 24% are missing registrant phone # (1039 records)
  - 87% are missing registrant fax # (3867 records)
  - 10% are missing admin contact name (439 records)
  - 11% are missing admin contact email (502 records)
  - 12% are missing admin contact address (514 records)
  - 60% are missing admin contact fax (2647 records)

Registrant email addresses were removed from data by seller

SSAC
ICANN Security and Stability Advisory Committee

Conclusions

- The absence of credible statistics on the extent to which personal contact information can be derived from "whois data" instigated this study
  - This study offers one set of findings to hopefully fill that void
- Study shows that
  - Personal contact information can be extracted from approximately 1 in 7 Domain Name Registration Records
  - Approximately 1 in 7 registration record also contain insufficient information to conclusively distinguish whether contacts are businesses or individuals

SSAC
ICANN Security and Stability Advisory Committee

Future Work

- During the examination of the sampling, anecdotal evidence suggests that
  - Causes for - and remedies to reduce - the number of incomplete records merit attention
    - 456 of 5000 originally sampled records were entirely unusable
    - Of the remaining 4444, 600 were missing information used classify a contact
  - Some information collected for registration purposes may not be as useful today as it was in the past