New gTLD Program Safeguards Against DNS Abuse: Research Design and Method

Brian Aitchison, MRes, PhD | 12 May 2016

Agenda

1. Introduction (2.5 mins)
2. Defining DNS Abuse (2.5 mins)
3. Safeguards (2.5 mins)
4. Research Method (5 mins)
5. Using a Model (5 mins)
6. Conclusion and Q&A (2.5 mins)

Introduction

- Master of Research (MRes)
  - Qualitative and Quantitative Research Design and Application
- PhD, Political Science
- Research lead on project to study impact of Internet of Things on economic growth
- So what does that have to do with DNS abuse?
  - METHOD!
    - Design
      - Organizing and testing BIG, COMPLEX variables (CCT = BIG!)
    - Primary research over secondary
      - Primary = original
      - Secondary = review/interpretation of original (literature review)
    - Demonstrating *causality* through *hypothesis testing*
      - Explanation over description (inferential over descriptive statistics)
      - Asking and answering "why" over "who what when etc"

Defining DNS Abuse

- DNS abuse aspect of “Consumer Trust”
- Working definition: “…intentionally deceptive, conniving, or unsolicited activities that actively make use of the the DNS and/or the procedures used to register domain names.”
- Carried out via:
  - Compromised (“hacked”) domains
  - Malicious registrations
- Intent: to distribute malware and/or steal (money, credentials, etc)

Defining DNS Abuse (continued)

- Registration Abuse Policies Working Group, 2010
  - Registration Abuse
    - Registering domains with intent to engage in abusive or unethical activity
    - Subject to GNSO policy making and ICANN contract enforcement authority
    - e.g. cybersquatting, false affiliation, deceptive domains …
  - Use Abuse
    - Using domains maliciously after they've been registered
    - Not as subject to GNSO policy-making and ICANN contract enforcement
    - e.g. phishing, spam, malware/botnet command and control, denial of service attacks, stealing credentials…

Safeguards

How do we ensure that bad actors do not run registries?

1. **Vet registry operators** through background checks to reduce the risk that a potential registry operator has been party to criminal, malicious, and/or bad faith behavior.

How do we ensure integrity and utility of registry information?

2. **Require Domain Name System Security Extension (DNSSEC) deployment** on the part of all new registries to minimize the potential for spoofed DNS records.
3. **Prohibit "wildcarding"** to prevent DNS redirection and synthesized DNS responses that may result in arrival at malicious sites.
4. **Encourage removal of "orphan glue" records** to minimize use of these remnants of domains previously removed from registry records as "safe haven" name server entries in the TLD's zone file that malicious actors can exploit. The sixth agenda item

How do we ensure more focused efforts on combating identified abuse?

5. **Require "Thick" WHOIS records** to encourage availability and completeness of WHOIS data.
6. **Centralize Zone File access** to create a more efficient means of obtaining updates on new domains as they are created within each TLD zone
7. **Document registry- and registrar-level abuse contacts and policies** to provide a single point of contact to address abuse complaints
8. **Provide an expedited registry security request process** to address security threats that require immediate action by the registry and an expedited response from ICANN.

How do we provide an enhanced control framework for TLDs with intrinsic potential for malicious conduct?

9. **Create a draft framework for a high security zone verification program** to establish a set of criteria to assure trust in TLDs with higher risk of targeting by malicious actors—e.g. banking and pharmaceutical TLDs—through enhanced operational and security controls.

Research Method

1. Identify the research problem clearly. What is the empirical puzzle we're trying to solve?

   - **Research problem**: It is unclear how effective the safeguards to mitigate DNS abuse in new gTLDs have been.
   - **Empirical puzzle**: Some indicators point to reduced amounts of DNS abuse in TLDs in general (legacy and new), while others point to increasing rates in particular TLDs. The extent to which the safeguards to mitigate DNS abuse have played a role in this **variation** remains unclear.

2. Review and synthesize previously published literature associated with the problem.

3. Clearly and explicitly specify research questions and/or hypotheses central to the research problem.

   - **Research question(s)**: What explains the **variation** in the rates of abuse in different TLDs? To what extent have the safeguards put in place to mitigate them been effective? What explains **variation** in abuse rates in new vs legacy TLDs? What about **variation** within new gTLDs?
   - Research questions and hypotheses should also indicate how each term is segmented, defined and/or measured. What is "effectiveness" and how do we measure it? What is competition, choice, and trust? Be very clear in justifying proxy measures and remember that no data is perfect!

4. Describe the data necessary to adequately answer the research questions and/or test the hypotheses, and explain how such data will be obtained.

5. Describe the methods of analysis to be applied to the data in determining whether or not the hypotheses are true or false.

Research Method: Hypotheses (Step 3)

Independent (Explanatory) Variable(s) → Dependent (Response) Variables
(Independent variables are hypothesized to have effect on dependent)

New gTLD Program → Competition, Choice, and Trust
(The New gTLD Program has had *an effect* on CCT)

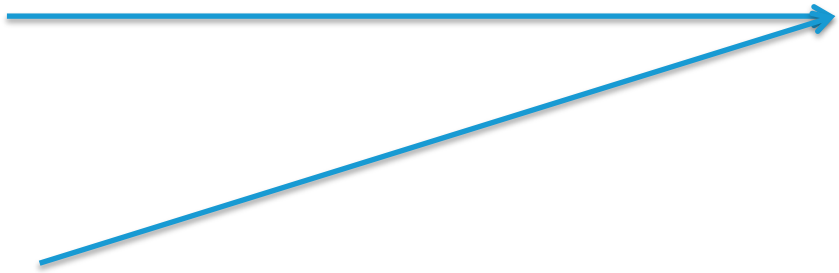
New gTLD Program → Trust
(The New gTLD Program has had *an effect* on Consumer Trust)

Safeguards Included in New gTLD Program → DNS Abuse (proxy for "Trust")
(The safeguards of the New gTLD Program have had *an effect* on DNS abuse)
e.g.
Vetting Registry Operators → DNS abuse
DNSSEC Deployment → DNS abuse
Removal of Orphan Glue Records → DNS abuse
etc…

DNS Abuse: Base Model

Explanatory Variable:
New gTLD Program (DNS Expansion)

- Number of legacy, new, total domains



Response Variable: DNS Abuse Rate (as proxy for "Trust" ie higher abuse = less trust

- Rate of abuse in legacy, new, total domains
- Spam
- Phishing
- Malware
- Cybersquatting
- etc…

"Sub-" Explanatory (Intervening) Variables: Safeguards to Mitigate DNS Abuse

1. Registry Operator Vetting
2. DNSSEC Deployment
3. "Wildcarding"
4. Removal of "orphan glue" records
5. "Thick" WHOIS records
6. Centralized Zone File access
7. Documented registry- and registrar-level abuse contacts policies
8. Expedited registry security request process (ERSR)
9. High Security Zone verification program

Safeguards: Testing Effectiveness (Steps 4 and 5)

Potential Data

1. **Registry Operator Vetting**
   - Formal compliance complaints
   - Termination of RA?
2. **DNSSEC Deployment**
   - # TLDs in root w/ signed keys
   - # of 2nd-level w/ signed keys
   - DNSSEC Compliance, SLA monitoring
3. **"Wildcarding"**
   - Redirection via "error traffic monetization"
   - Compliance complaints (0)
4. **Removal of "orphan glue" records**
   - Zone files to measure OG removal over time
   - Registry anti-abuse policy
5. **"Thick" WHOIS records**
   - WHOIS Accuracy Reporting System?
   - First responder feedback
6. **Centralized Zone File access**
   - CZDS password data (total and per TLD)
   - User feedback
7. **Documented registry- and registrar-level abuse contacts policies**
   - Testing functionality of abuse contacts
   - User feedback
8. **Expedited registry security request process (ERSR)**
   - Speed and ease of process
   - User feedback
9. **High Security Zone verification program**
   - Abuse in private HSZ vs in TLDs without

Methods of Analysis

Qualitative Analysis
- Interviews
- Surveys
- Focus Groups

Quantitative Analysis
- Surveys
- Descriptive Statistics
- Statistical Modelling (Inferential Statistics)

DNS Abuse: Model Applied

High Level Example (Big Picture/Context):

New gTLD Program (DNS Expansion)
- Number of legacy, new, total domains

(Descriptive Stats)

(Inferential Stats)

Proportion of Abusive Domains
- Segmented by legacy, new, total domains; by TLD

(Descriptive Stats)

Sub-Level <u>Quantitative</u> Testing Example (Individual Safeguard“ Effectiveness”)

DNSSEC Deployment
- # of domains w/ DNSSEC

(Descriptive Stats)

(Inferential Stats)

Abuse rate in segment
- in new v legacy, between TLDs, etc

(Descriptive Stats)

Sub-Level <u>Qualitative</u> Testing Example (Individual Safeguard“ Effectiveness”)

Expedited Registry Security Request (ERSR) Process
- Existence vs non

(Process)

(Survey and/or User Feedback Interviews)

<u>Survey</u>: Rate effectiveness of ERSR 1-5 (large pool of respondents)
<u>Interview</u>: Would you describe the ERSR as “effective” or “ineffective”? Why? (smaller pool of respondents)

Specific security threat/issue
- experience with ERSR, ease of use, expediency, etc

(User Feedback)

Conclusion: Points to Remember

1

Research Method = Work Plan!

Having your methods arranged systematically by data type and within a model will give necessary structure to a big research project. Just fill in the gaps!

2

Stick to your hypotheses

You've gathered data. Now which side of the variable relationship does it go on? How will it help you test your hypotheses?

3

Mixed methods are great!

Quantitative work gives breadth, while qualitative gives depth

4

Define terms in a measurable way

Competition? Trust? Choice? Effectiveness? DNS abuse? A good literature review should provide ideas on how nebulous concepts have been measured before. Logically justify proxy measures.

5

The scientific method is universally applicable

These methods and models are modular and can be adapted to structure other parts of the review.

6

An economics consulting/ research firm is money well spent

A small team of well-trained researchers could likely complete "legwork" of entire CCT review in 6 – 12 months (if we're feeding them the data and have a put together model)

Engage with Operations and Policy Research

Thank You and Questions

Reach me at:
Email: brian.aitchison@icann.org
Website: icann.org

twitter.com/icann

gplus.to/icann

facebook.com/icannorg

weibo.com/ICANNorg

linkedin.com/company/icann

flickr.com/photos/icann

youtube.com/user/icannnews

slideshare.net/icannpresentations