Name Collision Analysis Project (NCAP) Update

ICANN74 - 14 June 2022

James Galvin & Matt Thomas, co-Chairs

Agenda

1. Background
    a. NCAP Project Proposal
    b. NCAP Studies One and Two
2. Completed Work
3. Findings
4. Workflow
5. How to Participate in NCAP
6. Q&A

1. Background

Board Request

- ICANN Board tasked SSAC to conduct studies to present data, analysis and points of view, and provide advice to the Board on name collisions
  - Specific advice regarding .home/.corp/.mail
  - General advice regarding name collisions going forward

- Studies to be conducted in a thorough and inclusive manner that includes other technical experts
  - 25 discussion group members, including 14 SSAC work party members
  - 23 community observers

NCAP Project Proposal

- Board Resolutions
- Project Charter
- Project Proposal
- Community Wiki Home

NCAP Studies

- **Study One: Gap Analysis**
  - Properly define name collision
  - Review and analyze past studies and work on name collision and perform a gap analysis

- **Study Two: Root Cause and Impact Analysis**
  - Suggested criteria for determining whether an undelegated string should be considered a string that manifests name collisions, i.e., is a “collision string”
  - Suggested criteria for determining whether a Collision String should not be delegated
  - Suggested criteria for determining how to remove an undelegated string from the list of “Collision Strings” (aka mitigations)

- **Study Three: Analysis of Mitigation Options**
  - Identification and assessment of mitigation options
  - Production of recommendations regarding delegation

First Revision - Study One Proposal

- **Study One**
  - Bibliography of all things name collision related
  - ~~Build a data repository~~
  - *Recommendation regarding future studies*
- **Study Two**
  - Original goals
    - ~~Build a data repository~~
    - Understand the root cause of most name collisions
    - Understand the impact of name collisions
  - Original tasks
    - Conduct root cause analysis
    - ~~Build a test system which can be used for impact analysis and to test possible mitigation strategies~~
    - *Conduct impact analysis*
    - Produce a report on the results of Study Two
    - Undertake a formal public consultation on the results of Study Two
- **Study Three** - yet to be done - analysis of mitigation options

Second Revision - Study Two Proposal

Study Two Goals:

1. ~~Build a data repository~~
2. Understand the root cause of most name collisions
3. Understand the impact of name collisions

Study Two Tasks:

1. Conduct root cause analysis
2. ~~Build a test system which can be used for impact analysis and to test possible mitigation strategies~~
3. Conduct impact analysis
    a. Perform updated case studies of the CORP, MAIL, HOME
    b. Perform a data sensitivity analysis
4. Produce a report on the results of Study Two
5. Undertake a formal public consultation on the results of Study Two

2. Completed Work

Completed Work

- Case Study of Collision Strings
  - Studies of .corp, .home, .mail, .internal, .lan, and .local using DNS query data from A and J root servers
  - Highlight changes over time of the properties of DNS queries and traffic alterations as a result of DNS evolution
- A Perspective Study of DNS Queries for Nonexistent Top-Level Domains
  - Aims to understand the distribution of DNS name collision traffic throughout the DNS hierarchy
  - Provide insights into where and how DNS data can be collected and assessed
- Root Cause Analysis - New gTLD Collisions
  - Seeks to analyze various aspects of name collisions and the 2012 round controlled interruption to identify the root cause of related incidents reported by affected parties

Key Takeaways

- Case Study
  - Case studies of CORP, HOME, and MAIL indicates *impact* has increased
  - Critical Diagnostic Measurements help predict the *impact* of name collisions
  - Leaking collision strings differ from delegated TLD queries
  - DNS-SD protocols and suffix search lists are a major problem
- Perspective of DNS Queries
  - Study shows similarities and differences of RSIs and PRR
  - Existing measurement platforms could be extended to help inform applicants
- Root Cause Analysis
  - Private use of DNS suffixes is widespread
  - Name collision reports are supported strongly by measured data
  - The impact of TLD delegation ranged from no impact to severe impact
- **Name collisions are and will continue to be a increasingly difficult problem**

3. Findings

Current Findings

- Name collisions are and will continue to be an increasingly difficult problem
  - Case study indicates impact has increased
  - DNS service discovery protocols and suffix search lists are a continuing problem
- Critical diagnostic measurements (CDMs) are a way to assess name collisions to inform the assessment of the risk of delegation
- Mitigation and remediation is problematic, increasingly difficult as the volume and diversity of CDMs increases
- Existing measurement platforms could be extended to help inform applicants

Critical Diagnostic Measurements (CDMs)

- Query Volume
- Query Origin Diversity
  - IP address distribution
  - ASN distribution
- Query TYPE Diversity
- Label Diversity
- Other characteristics
  - Open-Source Intelligence (OSINT)
- Case Study and 2012 Round Controlled Interruption focused on DNS queries
  - Queries other than DNS should be considered

4. Workflow

What Problem Are We Trying To Solve?

- ICANN Board needs a methodology for evaluating and reducing the risk of delegation of a new TLD proposed string
  - Propose a methodology for identifying collision strings (“high risk” labels) that should not be delegated
  - No other string would be blocked as a result of name collisions
- **Name collision analysis is a risk management problem**
- Is it possible to objectively identify a “high risk” label?
  - If not, is it possible to provide guidance to identify a “high risk” label?
- Is it possible to objectively identify “do not apply” labels?
  - If not, is it possible to provide guidance to identify “do not apply” labels?

Goals of the Workflow

- To ensure that name collisions can be assessed
  - Requires name collisions to be visible, if they exist
- To ensure there is an opportunity for a mitigation or remediation plan to be developed and assessed
  - Requires understanding the cause of name collisions such that a mitigation or remediation plan (or both) can be developed and assessed

- Two operating roles are needed to conduct the assessments

Technical Review Team

- Need to be independent and neutral experts
- Technical expertise must include:
  - Knowledge and understanding of DNS specifications, provisioning, and operation
  - Knowledge and understanding of Internet infrastructure
    - Where it intersects with the DNS
    - Where it intersects with the usage of the DNS by applications and services
  - Ability to review and understand data collected (e.g., CDMs)
  - Ability to understand and assess risk
- Four responsibilities
  - Assess the visibility of name collisions
  - Document data, findings, and recommendation(s)
  - Assess mitigation and remediation plan
  - Emergency response

Neutral Service Provider

- Responsible for operation of the servers that will collect the CDMs
  - Data privacy concerns are still under discussion
  - Is this part of the Technical Review Team or a separate team?
  - If a separate team, could there be more than one?

- Four responsibilities
  - Operate an authoritative DNS name server for the Passive Collision Assessment
  - Operate Active Collision Assessment environment
  - Log processing and analysis
  - Emergency response

Name Collision Analysis Workflow

1. Applicant selects TLD label
2. Applicant submits application
3. Passive Collision Assessment (PCA)
4. Active Collision Assessment (ACA)
5. Board gets final package for decision

1. Applicant Selects Label

- Objective: Applicant gets an indication of the presence of name collisions
  - This is not definitive of acceptance or rejection of application
  - If collisions are present this is likely indicative of the need for further scrutiny

- Indication of the presence of name collisions?
  - Assumes passive data publicly available
  - ICANN will likely be source of passive, factual data

- Should applicant be able to request TRT Initial Risk Assessment
  - Perhaps only under “high risk” conditions?

Step 2. Applicant submits application

3. Passive Collision Assessment (PCA)

- Goal is to make name collisions visible
  - Pull data from throughout the DNS infrastructure
- Technical Review Team conducts first assessment
  - To identify “high risk” labels - based on public data - if so, becomes “special case”
- Passive provides low risk to clients
  - Minimally disruptive to existing behavior
  - Proposed TLD added to the root zone
  - Deploy a DNS authoritative service with “no content” in the zone
  - Collect CDMs
- Technical Review Team conducts second assessment
  - To identify “high risk” labels - if so, becomes “special case”

4. Active Collision Assessment (ACA)

- Goal is to support preparation of a mitigation or remediation plan (or both)
  - Seek additional data in support of investigating cause of name collision
- Active is a risk to clients because it is disruptive to existing behavior
  - Proposed TLD added to root zone
  - Deploy an TLD authoritative service for DNS and other protocols (e.g., web)
    - Include real wildcard IP addresses (IPv4 and IPv6)
  - Collect CDMs
- Technical Review Team conducts third assessment
  - To identify “high risk” labels - if so, becomes a “special case”
- Mitigation and remediation plan?
  - When to create?
  - Access to data?
  - Will need to be reviewed by Technical Review Team

Step 5. Package is submitted to the Board for review and decision

5. How to Participate

5. NCAP - How to Participate

- Join the discussion group
  - https://docs.google.com/forms/d/1PDIX6sMIdP4vLn1LLuefxsup78mLM0iDb8ybWhlw2T4/edit

6. Q&A