An Approach to Measuring Name Collisions Using Online Advertisement

1. Introduction

We propose a method, as an additional step in the name collision analysis flow, that uses online advertisements as a user-based measurement platform to measure the impact of name collisions to end users with minimal disruption.

This document describes the proposal. It is organized the following way. Section 2 briefly describes the advertisement-based measurement as a user-based measurement platform. Section 3 details how the advertisement-based measurement platform can be used to measure name collisions. It also includes an analysis of the benefits and limitations of this approach. Section 4 describes where the ad-based measurement proposal can be added to the overall name collision analysis workflow as currently discussed by the name collision analysis project discussion group, and how it interacts with other components of the workflow.

2. Brief Background on Advertisement-Based Measurement

“User measurement” is based on having end users generate the event or events that we want to observe and measuring the response the network provides to these users. An implementation of this approach treats the system as a “black box” and the analysis exercise is one of correlation of various stimulus inputs to network responses. This approach may be further refined by matching this user-level data up with data obtained from observation points within the network infrastructure. In this case the original event or trigger is known and the network’s response to the event can be observed both within the network infrastructure and in terms of the response delivered back to the user.

An advertisement-based measurement is an example of user measurement. An advertisement is typically a package of one or more images, an HTML 5 script which is executed when the ad is "impressed" on the user device, and some instructions to execute if the user clicks on the ad. The advertisement system typically restricts the code to a subset of the HTML5 functions. The commonly supported function in ad scripts is the retrieval of an image URL.

The process of loading a URL involves two steps: the resolution of the domain name part of the URL, and the use of an HTML session to retrieve the URL. Servers will record the interactions with the Internet infrastructure that are directly triggered by the user's efforts to resolve the DNS name and perform the URL retrieval.

In this respect the exercise is similar to other forms of measurement: data is gathered from the infrastructure elements of the Internet and analyzed to produce the measurement. However, in one respect there is a major difference. Normally the infrastructure measurement process has to infer the reason why the observed behaviors occur. *However, in this case the triggering actions that cause the behavior are known in exact detail. The volume of these trigger events, the timing of these events, even the location of the original events are all known quantities, so in this case the aim of the analysis is to take the original inputs and the observed effects and infer the dynamic behavior of the infrastructure in responding to the trigger events.*

More details of the general approach using this form of active measurement can be found at: https://www.potaroo.net/presentations/2016-02-10-ad-measurement.pdf.

3. Advertisement-Based Measurement For Name Collision Analysis

3.1 How it Works

The proposal would involve the temporary delegation of a candidate TLD into the DNS root zone. The delegation process of inserting a string into the DNS root zone will make the TLD active in the domain name system. The required delegation information in the referral from the root is a complete set of NS records and the minimal set of requisite glue records.

The TLD authoritative name servers will be configured in one of two possible ways:

a. the servers will respond to all queries, including a single label TLD string, with a response of NXDOMAIN, or
b. the servers will contain two additional records for the SOA and name server records and the servers would respond to all other queries with NXDOMAIN.

The level of disruption to existing private use of such labels by this restricted form of name delegation would be reasonably expected to be *minimal*; however, the series of referrals and responses received by resolvers are different from a direct NXDOMAIN response from the root server system. It is possible that this slight difference could impact application resolution processes, such as search list processing. The risk of performing such a temporary delegation for the purpose of name collision assessment must be made appropriately.

After the delegation, one (or more) advertisement-based measurements will be set up for the candidate TLD. The Ad Server will present ad-impressions to end users. Each ad-impression will include a notional '1x1' image that will trigger a script to query the DNS for resolution of a <randomized SLD>.<Candidate TLD>. The time, location and volume of such queries are known in advance. In other words, the cause of the queries seen for the name being studied are *now* known. This will be matched against the queries that are seen at the authoritative name server. For those users whose infrastructure uses <Candidate TLD> in a local context, the query will not be seen at the authoritative nameserver of the Candidate TLD. For those users whose infrastructure does not use <candidate TLD>, the queries will be seen at the authoritative nameserver of the Candidate TLD.

Based on this, for each IP address, we calculate:

> attempted queries to the authoritative nameserver of Candidate TLD / queries sent to each client IP address

We can average this value across all measurements. If there was no local use of this name (a "collision") then the average ratio of this measurement would approach 1. If there was extensive local use of this name the average ratio would approach 0.

The IP addresses can also be geo-located to countries, or located to ASNs and similar averages calculated. That way we can also understand the variance in collision rates between countries and between individual networks.

To protect user privacy, each measurement uses a unique query name profile and would only need to see the queries received by the server set (e.g., the authoritative name server and the HTTP server) that match the name profile associated with the experiment. The query logs of the experiment server should contain only the IP addresses of the recursive resolvers that handle the triggered query and not the IP address of the querier. Any residual Client Subnet information in query logs can be stripped from the retained query data for added privacy protection.

3.2. Benefits of this Approach

This approach has two main benefits. First, advertisement-based measurement provides high fidelity information about name collisions seen at the end user level. Past approaches look for the incidence of queries for non-existent labels at the root zone and attempt to infer the extent to which the label is already being used from these queries. Such an approach requires data from the root servers and recursive resolvers to expose these queries for analysis. Due to privacy considerations, sometimes the data would not be available. Most importantly, recent development of mechanisms such as QNAME Minimization, Aggressive NSEC caching, LocalRoot, etc further occlude DNS queries.[1] These changes to the DNS name resolution process appear to have reached a point where we would no longer be sure the extent of the partial view, and the conclusions that we can draw from it.[2]

User measurement is based on having end users generate the event or events that we want to observe and measuring the response the network provides to these users. This approach may be further refined by matching this user-level data up with data obtained from observation points within the network infrastructure. In this case the original event or trigger is known and the network's response to the event can be observed both within the network infrastructure and in terms of the response delivered back to the user.

It is important to note that this measurement approach is not measuring browsers. It is measuring the DNS infrastructure used by these end users.

Second, this approach does not break systems that currently use non-existent domain names as TLDs. There will be no resource records (RRs) in the delegated zone. The level of disruption to existing private use of such labels by this restricted form of name delegation would be

[1] More discussions on this can be found at https://wkumari.github.io/2022/02/25/losing-visibility-into-dns.html

[2] For example, with the increasing use of LocalRoot and Aggressive NSEC caching, what is the proportion of query names seen at one root server 'letter' for any given label as a portion of the full extent of edge activity of DNS queries being passed into the DNS resolution infrastructure? It is increasingly challenging to accurately estimate the extent of use of a particular top level domain name using only queries gathered from passive observation points.

reasonably expected to be *minimal*; however, the series of referrals and responses received by resolvers are different from a direct NXDOMAIN response from the root server system. It is possible that this slight difference could impact application resolution processes, such as search list processing. The risk of performing such a temporary delegation for the purpose of name collision assessment must be made appropriately.

3.3 Limitations of this Approach

The online advertisement network measurement is focused on individual users, as distinct from device use. So it may miss some devices that are associated with IoT activities when measuring name collisions. It may also under-report TLDs that are specifically used for software development and testing purposes (e.g., .prod). Overall though, this should not be a huge concern in most cases, as even if a TLD is intended to primarily support device activity or devices deployed in the same environments as users would likely share the same resolution environment, and a measurement test directed to browsers that specifically tested such a TLD would expose the same DNS behavior as a DNS request from a device in the same environment. The large sample in the experiment would also mitigate this issue as well.

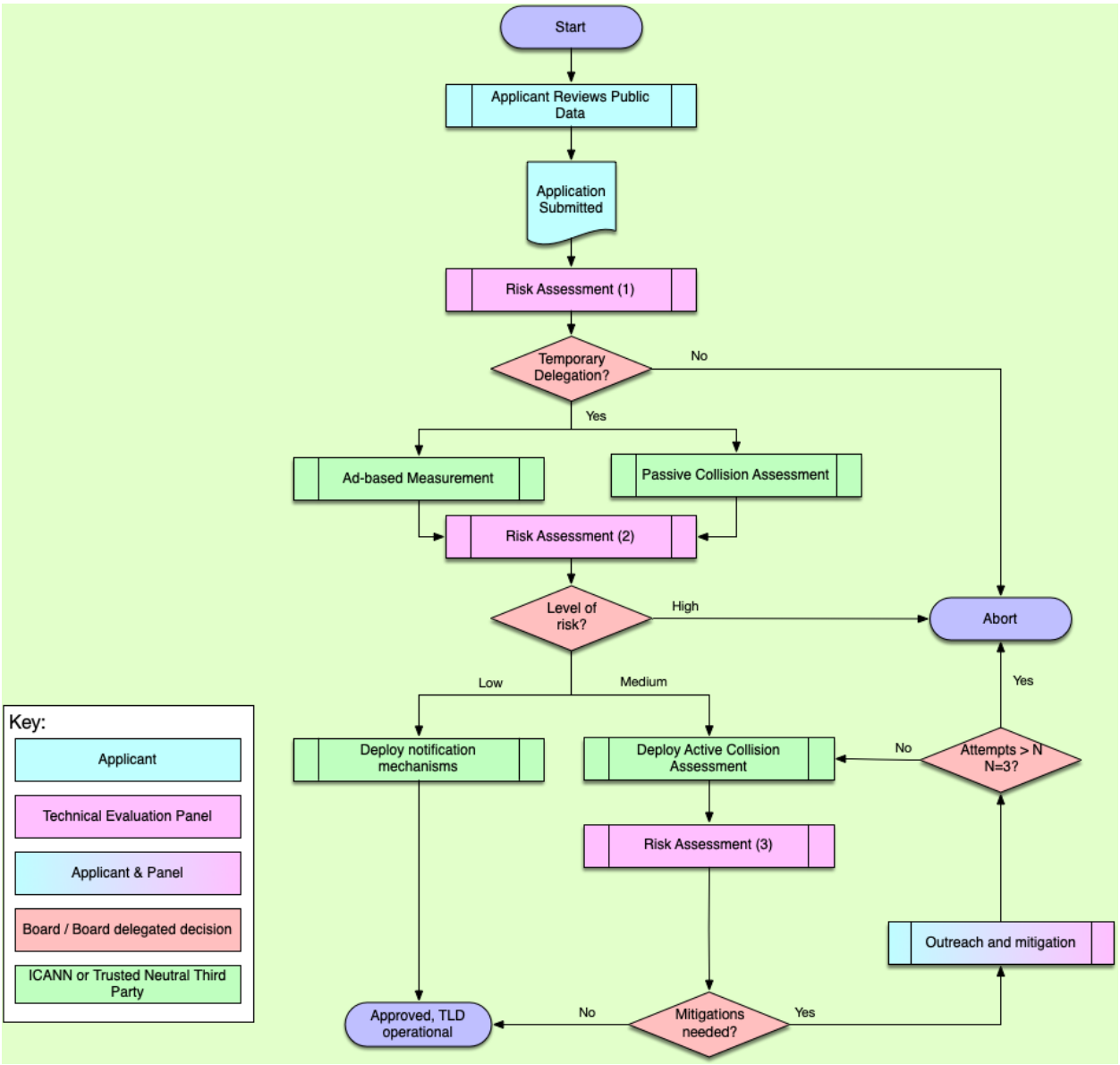
In addition, user-based measurement is challenging and needs careful planning and execution for it to be effective. The issues of enrolment and selection bias are often challenging to measure and compensate for, and even defining a "user" can be a challenging question. A user-based measurement is typically based on sampling, and like all such sampling methods considerable care must be taken to ensure that a statistically significant sample set is enrolled to perform the measurement, and care must be taken to identify various sources of sampling bias and where possible the analysis must include compensation to reduce such the sampling bias. In general, although there are of course exceptions to this, the larger the sample set the better, and if performed over a number of rounds the broader the sample set (i.e. varying the sample set for each round of measurement) the better. When we are considering a user base of billions of human users across some 80,000 components networks using a variety of platforms and applications, then the size and diversity of the sample set are significant factors in setting up a measurement that can provide useful and meaningful results.

This limitation means the experimentation needs to be carried out by researchers who are experienced with internet measurement in general and advertising-based measurement in particular. In addition, there are very few advertisement networks that can accommodate these types of experiments. Thus, this can not be done in an ad hoc manner by an applicant. It would also mean that the policy maker needs to be selective in deciding what strings to delegate to run the experiment.

4. How Advertisement-Based Measurement For Name Collision Analysis Fits into Overall Name Collision Analysis Workflow

In this section, we describe how the advertisement-based measurement for name collision analysis fits into overall name collision analysis workflow. To help the readers understand, we

have created a visual flow diagram based on the workflow slide deck discussed in the NCAP DG. We then added the ad-based measurement in this workflow.



Step 1: Applicant Submits Application

The applicant submits an application for a string (referred to here as the "candidate TLD label").

Step 2: Risk Assessment (1)

After the application is submitted, the technical review team will perform an initial risk assessment based on reviewing existing data on the candidate TLD label, and make a

recommendation to the ICANN Board on whether to delegate the string, to itself or to a trusted neutral third party, for passive collision assessment and advertised-based measurement.

The technical review team may also conclude that the string poses too many risks and should not proceed with temporary delegation. In this case, it would recommend the ICANN Board to stop processing the application.

Step 3: Temporarily delegate TLD

Should the ICANN Board decide to temporarily delegate the candidate TLD label, the applicant or the trusted neutral third party would use a dedicated server environment whose only response for any second level domain queries to the candidate TLD label is NXDOMAIN. There will be no resource records (RRs) in the zone.

The delegation process of inserting a string into the DNS root zone will make the TLD active in the domain name system. The required delegation information in the referral from the root is a complete set of NS records and the minimal set of requisite glue records.

The TLD authoritative name servers will be configured in one of two possible ways:

a. the servers will respond to all queries, including a single label TLD string, with a response of NXDOMAIN, or
b. the servers will contain two additional records for the SOA and name server records and the servers would respond to all other queries with NXDOMAIN.

The level of disruption to existing private use of such labels by this restricted form of name delegation would be reasonably expected to be *minimal*; however, the series of referrals and responses received by resolvers are different from a direct NXDOMAIN response from the root server system. It is possible that this slight difference could impact application resolution processes, such as search list processing. The risk of performing such a temporary delegation for the purpose of name collision assessment must be made appropriately.

The delegation of the TLD and the continued response of NXDOMAIN for the TLD label effectively enables a collection and analysis of DNS queries that would otherwise require a collective root server system measurement. This passive data can be used to assess name collision risks in addition to the active measurements described herein. They are discussed in 2a and 2b below.

Step 3a: Passive Collision Assessment

Upon delegation, instrumentation will be set up to collect the following data:

Critical Diagnostic Measurements (CDMs) of Name Collision

Traffic Properties:
- Network diversity
  - Number of unique ASNs, /24s, etc.
  - Distribution of traffic (e.g. heavily weighted in a few ASNs)
- Geographical diversity
- Qtype distribution
- Query volume
- Longitudinal trends

Qname and Labels:
- Distinct SLDs
  - Distribution of traffic over SLDs
- Amount of "noise" (e.g. Chromium)
- SLDs appear to be delegated TLDs
- First label features
  - DNS-SD
  - Common protocols
- Qname Minimization effect

Other Attributes:
- The string's context
- OSINT of string being used
- Data sensitivity and catchment of data collector

A collection period of at least four weeks should be set up to collect the data. At the conclusion of the collection period the data is compiled and sent to the technical review team along with the advertisement-based measurement data, described in 2b. All data may also be made available to the applicant.

Step 3b: Advertisement-based measurement

In parallel to the passive collision assessment described above in 3a, one (or more) advertisement-based measurements will be set up per TLD. The details of each measurement is explained in section 3.1 above.

Each advertisement-based measurement in our setup will run up to 4 weeks, or when 100M individual samples are collected. This is informed by past advertisement-based measurement at APNIC, which shows that within 7 days one can have a relatively good set of data across all parts of the Internet with around 100M individual samples. Lower ad rates generally means a longer run time.

Step 4: Risk Assessment (2)

The data collected during the passive collision assessment (3a) and ad-based measurement (3b) will be shared with the applicant and technical review team. The technical review team will perform a second risk assessment and will categorize the candidate TLD into one of the three risk categories:

- High risk: The name collision impact for candidate TLDs in this category is very high, and the ability to mitigate effectively is very limited.
- Medium risk: The name collision impact for candidate TLDs in this category is medium to high. However, there exists an opportunity to mitigate some of the collision risks.

- Low risk: The name collision impact for candidate TLDs in this category is low to medium.

The criteria for the risk categorization will need to be defined ahead of time.