ICANN78 | Prep Week – Name Collision Analysis Project Study 2 Update
Tuesday, October 10, 2023 – 10:00 to 11:00 HAM

KATHY SCHNITT: Hello, and welcome to the Name Collision Analysis Project Study 2 Update. My name is Kathy, and I am the remote participation manager for this session.

Please note that this session is being recorded and is governed by the ICANN Expected Standards of Behavior. During the session, questions or comments submitted in the Q&A pod will be read aloud at the time set aside for question and answer. Questions written in chat may not be read aloud.

Interpretation for this session will include English, French, Spanish, Chinese, Russian, Arabic, and Portuguese. Click on the interpretation icon in Zoom and select the language you will listen to during the session.

If you wish to speak, please raise your hand in the Zoom room, and once the session facilitator calls upon your name, kindly unmute your microphone and take the floor. Please state your name for the record and the language you will speak if speaking a language other than English. When speaking, please be sure to mute all other devices and notifications. Please speak clearly and at a reasonable pace to allow for accurate interpretation.

This session includes automated, real-time transcription. Please note, this transcript is not official or authoritative. To view the real-time transcription, click on the Closed Caption button in the Zoom toolbar.

To ensure transparency of participation in ICANN's multistakeholder model, we ask that you sign in the Zoom sessions using your full name.

And with that, I'm happy to hand the floor over to NCAP co-chair, Matthew Thomas.

MATTHEW THOMAS: Thank you, Kathy. This is Matthew Thomas for the record, one of the NCAP co-chairs, along with my other co-chair, Suzanne Wolf. And we're very pleased to be here tonight, or today, to give an update on the Name Collision Analysis Project to all of you.

We hope this session is insightful to illuminate where the Name Collision Project has been going over the last several years and as to where we are and hope to conclude shortly. Next slide, please.

So at a high level today, we will be giving a little bit of a background around name collisions. As you might have heard when Kathy introduced this, this is Study 2 of the name collision. So we're going to give a little bit of background in terms of Study 1, what we're doing in Study 2. And that will bring us into talking about some of our completed work within Study 2, specifically looking at some of the completed studies and reports.

We'll then switch gears looking at the Board questions that the discussion group and SSAC was formally requested by the ICANN Board

to provide advice and guidance to; then into our current findings, which will dovetail into a discussion about a potential workflow for addressing name collisions in a sustainable and repeatable fashion going forward; followed by a general information about how to participate in NCAP as well as time for Q&A at the end. Next slide, please.

So just a little bit of background here. This slide, which will be included in the upcoming Study 2 Report, is a depiction of the historical major milestones related to name collisions over the last 10-plus years. Clearly, as you can see in this graph or chart, there is a large amount of work that has been put into the study analysis, guidance, and guardrails around name collisions.

And I think it's important for everyone to really understand all of the historical context around name collisions to really fully understand where Study 2 is evolving to and the goals that it's trying to achieve.

Clearly, we can't enumerate and talk about all of these major milestones, but when you have an opportunity to look at the Study 2 Report, these annotated figures or callouts on this graph are discussed in a detailed backstory and history of name collisions which will, hopefully, provide sufficient context and insights in terms of how name collisions are to be assessed and what the NCAP Discussion Group has framed as a risk management solution going forward. Next slide, please.

So specifically, the Name Collision Analysis Project has come out of a request from the ICANN Board to the SSAC in which the Board requested SSAC to conduct some studies to present data analysis and

an informed point of view as well as advice to the Board on the topic of name collisions in general. Within that, there was a specific request on advice regarding the strings of .home, .corp, and .mail, as well as a series of questions that were targeting more general advice regarding name collisions going forward.

The NCAP discussion project or group has been ongoing for several years now. I believe we're on year four or five. And we have been working in a thorough and inclusive manner that includes both technical and non-technical experts within the community. There are roughly about 25 discussion group members actively participating, of which 14 are SSAC Work Party members. And we also have roughly about 23 community observers. Next slide, please.

There's quite a substantial amount of NCAP-related information out there. It's too verbose to go into details here, but when these slides are made public, we would refer you all to the links on these four major items here, in which you can see the exact Board resolutions or questions that were asked to the SSAC and, subsequently, the NCAP Discussion Group; as well as the project charters, the project proposals.

And most importantly, probably, is the community Wiki homepage, which includes the last 100 and some recordings of the NCAP Discussion Group calls, presentations, reports, and whatnot, in which you can see all of the work that has been going on within this project. Next slide, please.

So at a top level, the NCAP study was actually broken down into three distinct phases. Study 1 was to look at a Gap Analysis around name

collisions. And that study had two major goals, the first of which was to give a proper definition of name collisions.

And if you recall on the first historical timeline slide that we showed here, when you look at that, one of the things you will notice is that the definition of what a name collision is or is not has evolved over the last decade. The connotation and denotation of name collisions has not always been aligned. So one of the major goals of Study 1 was to propose a proper definition of the name collision.

The second major objective was to conduct an exhaustive study of all things name collisions from all of the previous studies and to perform a potential Gap Analysis in terms of the works produced and the research done on name collisions.

Now, the Study 1 work was completed several years ago. It went out for public comment and has been officially published. But as an outcome of that work, it was identified in conjunction with the work from the NCAP Discussion Group that there did exist some gaps, specifically in the technology and the evolutionary standpoint of the DNS and how the DNS ecosystem has evolved since 2012 until now, that led to an altered and revised proposal for Study 2.

Study 2 is the current workings of the NCAP Discussion Group in which the revised goals and objectives are listed here below. The previous goals and objectives had larger objectives of trying to create data repositories and things that were deemed not to be further possible.

But the current scope was really focused on identifying what criteria makes a collision string, what criteria would allow for a string to be

removed from a list of collision strings if it was deemed a collision string. And it also called for some specific studies looking at various related components of name collisions that will get [inaudible] discussion here in just a little bit.

Now, the final study that has yet to be conducted is the Study 3, which is an analysis of mitigation options. I will note that just earlier, or last week, the NCAP Discussion Group had a two-day workshop in which many of the members came and spent all day hashing out all things name collisions.

And one of the important takeaways out of that workshop was that the discussion group, based off of its current understandings/ findings, and the scope and remit of what was proposed in Study 3, are going to likely recommend that Study 3 not be performed, given what we know about mitigation options.

In a nutshell, mitigation options are likely to be tailored on a string-by-string or name collision-by-name collision basis. There is really no one-solution-fits-all type of name collision remediation or mitigation strategy, and the group has felt that that can be reflected in our notes coming out in Study 2; that Study 3 is likely to be not necessary at this time. Next slide, please.

So where do we stand in Study 2? So as I mentioned, in addition to those high-level objectives in Study 2, there were also three specific studies that were asked to be conducted. The first study was a case study of the collision strings. And as we mentioned earlier, one of the Board

resolutions specifically asked for advice on .corp, .home, and .mail. And this is what that case study was aiming towards.

But instead of just looking at .corp, .home, and .mail, three additional strings, .internal, .lan, and .local, were also added to the study simply because at the time of the study, those three strings were also receiving more than 100 million queries per day at A and J root servers at the time.

That case study was designed to look at the highlighting and the changes over time. It was a longitudinal study of the DNS traffic observed at A and J going back to, roughly, 2015 and looking at how the DNS traffic has changed or altered related to name collisions for those six different strings.

The second study was A Perspective Study of DNS Queries for Non-Existent Top-Level Domains. And this study was really aimed at trying to better understand where and how and to what degree of confidence DNS telemetry data can be used at various vantage points in the DNS hierarchy, specifically at the Root Server System, as well as recursive resolvers, to understand DNS name collision traffic.

And its goals were really to provide some insights into how and where DNS data could be collected and assessed, and what types of appropriate guardrails would need to be placed on that type of data to be used in future assessments of name collisions and their potential risks.

The last study is the Root Cause Analysis. And this is a study that was conducted by a contracted technical investigator from ICANN who was able to utilize the 40-some name collision reports that ICANN received

since 2012 related to various name collision problems stemming from the notification mechanisms, presumably around controlled interruption, and to understand how those incidents were reported, what the effects of name collisions were on those parties, and to also correlate those name collisions with any type of DNS telemetry data to confirm or deny the supporting of identifications of name collisions surrounding those. Next slide, please.

So these three studies were all conducted. The first two have already gone out for public comment. The third has been finalized. And consensus has been called, at least internally, within the NCAP Discussion Group. And there have been some key takeaways out of these three studies.

The first one which was, again, looking at .corp, .home, and .mail. That study clearly indicated that the potential for impact for those strings has increased over the years, especially over the last several years during the pandemic.

The Critical Diagnostic Measurements, which are a type of quantitative measurements that the NCAP Discussion Group has helped frame and form, in conjunction with the excellent work done by previous research from folks like [inaudible] and JAS and their assessments in which name collisions can be quantified through different metrics of the DNS like DNS query value, but also other dimensions like diversity, in which diversity can span multiple different vectors such as name diversity, QTYPE diversity, network diversity, so forth and so on.

The case study also identified that DNS service discovery protocols and suffix search lists are still some of the major contributing problems to name collisions as a whole. Things like WPAD, or the Web Proxy Auto-Discovery, or ISATAP, an IPv6 tunneling service, are some of the major ones that are causing problems out there.

Now meanwhile, on the second study, looking at the Perspective Study of DNS Queries, that study looking at how and where DNS traffic can be assessed provided some insights into the similarities and the differences of each root server identity within, as well as the differences that were observed at the root versus seen at large, public recursive resolvers.

Those insights helped form some potential opportunities going forward in which maybe new measurement platforms could be enhanced or extended to provide additional name collision DNS telemetry to potential applicants a priori their application submission. These are systems like ICANN's ITHI or ICANN's IMRS system which allows, already, some insights to be gleaned into potential name collision risks.

And finally, we have the Root Cause Analysis, which was looking at those 40-some reports. And that report really identified, again, that private use of DNS suffixes is still very widespread; name collision reports are strongly measured by data in which that data was collected from passive DNS data sources; and, most importantly, that the impact of the TLD delegations from the previous 2012 Round in which name collision reports were submitted to ICANN range from no impact to severe impact.

Now, name collisions, I think as a whole what we see from all of these key takeaways, is that name collisions are and will continue to be a difficult problem to identify and to remediate. Next slide, please.

JENNIFER BRYCE: Matt, the interpreters are just asking if you can slow down a little bit, given the technical nature of the presentation.

MATTHEW THOMAS: Will do. I'll try my best. Thanks, Jen. Okay. So hopefully, this is not as technical talk here. But one of the major remits of the original Board resolutions was to answer and to provide advice and/or guidance to several Board questions. And those Board questions actually consisted of nine different questions, the first of which was defining the name collision properly. That has been finalized and went out in Study 1 for public comment. You can see that out already on ICANN resources.

The remaining seven questions, which are not explicitly outlined here but rather put into a rough taxonomy of topics, range on a set of topics looking at the role of, for example, negative answers. Right now, name collision strings expect negative answers from the root servers in the Domain Name System, or the DNS.

So it was important to understand how those negative answers can impact potential end systems, users, or applications. Subsequently then, question three focuses on: what could be the potential harm if those negative answers stop being returned, what things can change in

those applications and systems, and what could be the potential harm from those when the negative answers stop coming back?

On the next point then, questions four and five look at what are potential ways to mitigate that harm. Are there techniques/tools/ procedures that could be used to help prevent that harm related to name collisions?

And then six, we have: what would be the risks of actually delegating the domain? Finally, we have questions seven, eight, and nine. And those are all around the context of undelegated strings and collision strings and the management of those strings in general.

The NCAP Discussion Group has already provided a first rough draft of those answers and advice on the Board questions. A rough consensus has been established within the NCAP Discussion Group.

However, as we'll go into in the next section talking about the workflow, there is a desire for the group to update those questions when the Study 2 Report nears a more final state in which the Board question advice and answers can be more closely cross-referenced with the substantial amount of work material, findings, and recommendations that will go into the overall Study 2 Report. Next slide, please.

So speaking of findings, we currently have a rough set of findings that you can see here. As I mentioned before, we have established, hopefully, a proper and final definition of what name collisions actually are and what name collisions are in certain contexts that are relevant to the ICANN community.

We have also determined that name collisions are and will continue to be an increasingly difficult problem to identify and to remediate. This was supported by our case study in which we've seen impact has increased. We've seen this by the proliferation of DNS service discovery protocols and the growth of suffix search lists. But we've also [noted] that potential bad collisions can still occur in the future.

We've also come to terms around name collision identification and quantification, as I mentioned before, around Critical Diagnostic Measurements, or the CDMs. These are a way that helps quantitatively describe potential DNS traffic-related telemetry that informs potential risk or impact of name collisions. However, the group will also note that there is still a qualitative component to name collisions that need to be assessed for all strings.

And then these ties into limitations around the data and that there are still limitations and growing limitations within the DNS about the ability to assess name collisions as was done in the 2012 round.

As noted, before, one of the major driving forces for a revised Study 2 was a technical Gap Analysis that the discussion group identified that major changes within the DNS ecosystem have either impaired or severely reduced the visibility of name collisions due to technology changes like things such as QNAME minimization, aggressive NSEC caching, the introduction of public recursive resolvers, and other things.

The group has also found that it is impractical, at least in advance of the next round, to create a do-not-apply list of strings. That being said, we

have, as mentioned before, found that there are potential opportunities for existing measurement platforms to be extended to help inform applicants a priori their application submission about potential name collision risks. And that type of information would most likely be reflected in some type of quantitative CDM types of measurements. Next slide, please.

So the workflow. As we mentioned before, the Board formally charged SSAC and NCAP Discussion Group to answer this series of questions and provide advice around name collisions. However, what really is needed is the creation of some type of sustainable, repeatable, and deterministic model of being able to assess name collision impacts and risks going forward. That method or workflow is really to help identify those high-risk labels that should not be delegated.

Outside of those high-risk ones, the underlying assumption is that all other strings would proceed through a typical procurement and onboarding process within the Subsequent Procedures. So what this really frames the Name Collision Analysis Project as is making this a risk management problem.

How can name collisions be assessed in a risk management type of framework? How do we objectively identify and distinguish high-risk labels versus the ones that should, by de facto, proceed normally? And is it possible to identify certain strings in advance or not? Next slide, please.

So in order to make that assessment, one of the things that needs to happen is that data needs to become available to assess the potential

risk or impact for name collisions. That might mean a multitude of different measurements or techniques that can be deployed in order to make that data available to make an informed decision within the context of a risk management problem, as we previously discussed.

Also, the goal of the workflow is also to provide an opportunity to create potential mitigation or remediation plans, if they need, in the case of high-risk streams. As we noted earlier, the intention of the discussion group is to provide advice that Study 3 around mitigation plans should not proceed.

That advice stems from a lot of what is being said here, that from the research and the findings that the discussion group has observed over the last several years, remediation and mitigation plans seem to be purpose-built for one collision at a time. General, broad-stroke solutions don't seem to be a desirable or achievable solution for solving the name collision problem. Next slide, please.

So this, at a high level, is a potential, proposed workflow, and timeline that the NCAP Discussion Group has been working on over the years. It's essentially a five-step process, and these five steps were conducted or designed to be in a logical, risk-adverse manner.

And risk-adverse here means that the most accessible data without disturbing the DNS ecosystem is first assessed for name collisions before proceeding to the next step in which a technology or operational change can incur to gather additional name collision data and make an informed assessment.

So if we walk through the workflow at the far left, we have an applicant potentially making a static assessment. This is nothing more than applicants a priori their application submission being able to go to resources like ICANN's ITHI or IMRS system that publish name collision related data to do a simple analysis to see if their strings are already contained within those published name collision data.

Now, by no means does the inclusion or exclusion of that string on those lists mean a final determination if it's good or bad, but it is at least a leading indicator and gives the applicant some informed decision before submitting their application. That is denoted as a potential offramp opportunity up there on number one. The offramps are still being discussed within the discussion group in terms of how they fit.

But the design and the concept in the overall objective of offramps in this workflow is to provide the applicants, as well as ICANN, the capability of being able to exit this workflow and not end up in a purgatory state where you have strings like .corp, .home, and .mail a decade later. It's hoping that this will allow for a more deterministic and finite period of execution.

So once the application has been submitted, it goes into the application processing procedure. This will commingle with the other processes that go on during subsequent procedures. The timing of when name collision assessments occurs is not exactly determined, but it's somewhat irrelevant based off of what the general advice and guardrails and guidelines were trying to provide ICANN here.

But the next major step will go into this Technical Review Team, or a TRT static assessment. And this Technical Review Team is an entity—a third party, neutral party—who has expertise in being able to assess DNS traffic data, understand the problem of name collisions, and assess potential impact or risk of that data and strings based off of the data that they are able to collect.

The TRT is providing output into the process of the granting of a TLD. But during this static assessment here, the TRT, again, is simply looking at the sources that were mentioned prior to the application being submitted things like ITHI or IMRS, or whatever other data the TRT may have privileged access to.

If it doesn't see an outlier or severe risk, the process of assessment will proceed into the next phase, which is Passive Collision Assessment, or PCA.

This is a new phase that is being proposed in which the applied-for TLD is actually delegated into the root zone, and the delegation points to authoritative top-level domain name servers that serve a wild-carded zone with a special record type that will respond with an answer of “no air and no data” but is effectively the equivalence of an NXDOMAIN response that was previously being received by stub resolvers.

What this delegation and data collection does—or this delegation does—is it provides a much more enhanced and extensive data collection than relying on some of the static data sources such as single root letter instances or other snapshots of DNS data from the root.

This allows, effectively, a Root Server System-wide collection that will help get around some of the technology changes that we mentioned before such as QNAME minimization, aggressive NSEC caching, and so forth to provide richer data, to allow the TRT to provide a more informed decision.

The next phase after Passive Collision Assessment, it goes into something called Active Collision Assessment. And the Active Collision Assessment is still being discussed within the discussion group. There are a set of tools and techniques that are being discussed in terms of what and how each one could be performed.

But at the heart of it, the goals of these tools and techniques are trying to achieve two major goals. One, collecting additional data in which most of this data would end up being actual impacted devices because their IPs are being collected instead of recursive resolvers. And two, trying to provide some kind of notification mechanism to those impacted entities that are incurring name collision problems.

Once the Active Collision Assessment period is completed, the Technical Review Team will complete its analysis, and it will provide its go/no-go assessment out into the general process that will facilitate the rest of the onboarding of a new [TLT] in its ultimate granting procedures within the ICANN process.

Again, this is a high-level overview of that. There is rough working consensus on this. There are some details that are still being worked out within the discussion group. But that's the overall working consensus towards something on this sustainable, repeatable model

that wants to be proposed towards ICANN going forward. Next slide, please.

So this is, again, a little bit more about the Technical Review Team. This is talking about some of their roles and responsibilities. Here, we're just simply saying that they need to be independent and neutral experts. They have to have DNS expertise, name collision expertise, and then also be able to frame all of that within a risk assessment kind of framework. And they really have the four responsibilities that are listed down there below. Next slide, please.

So how to participate? Well, hopefully, fingers crossed, NCAP is nearing its completion of the Study 2 Report. We are targeting a public comment before the end of the year for the Study 2. But we are always looking for new members, new voices, new thoughts. There is a link here to join the discussion group. So we welcome you all to come and participate.

And if not, just to be aware that, hopefully, that Study 2 Report will be coming out for public comment soon, and we encourage you to read it. It will be lengthy. It will be detailed. It will be technical. But hopefully, it will be concise, also, in the points that it needs to convey in terms of name collisions and how it fits into the overall picture of ICANN. Next slide, please.

And with that, I'll just ask my co-chair, Suzanne, if I missed anything or if she wanted to add any other color or commentary. And if not, we can open it up to Q&A.

SUZANNE WOOLF: That was very thorough. I do not have anything written down here that needs to be added, so I'd be happy to hear any questions.

KATHY SCHNITT: Well, Matt and Suzanne, as of right now, we do not have any questions.

SUZANNE WOOLF: See, you covered everything, Matt.

MATTHEW THOMAS: Excellent. Well, hopefully, this will be the last update we give out on this, and the next update will be on the actual report that goes out. Right?

SUZANNE WOOLF: And public comment and all of those good things.

MATTHEW THOMAS: Exactly.

KATHY SCHNITT: We actually got a hand raised. Sebastian, you can ask your question.

SEBASTIEN DUCOS: Hi. This is Sebastian Ducos from the GNSO. Could we go two slides back where you had your timelines? Maybe a bit more, sorry. Yeah, one more. This one.

So in the last round, we have the famous controlled interruption. Upon delegation, TLDs had to be in controlled interruption for 90 days, if I remember well, during which, essentially, the TLD couldn't be used because all traffic was interrupted. And after which, the TLD was usable.

But this is a period where the TLD was delegated, and from a registry operator point of view, was incurring both backend operator costs—and in some cases, that was renegotiated [but any case] ICANN cost.

Are you suggesting now that that period would be essentially six months and under similar conditions? Or is this something that has been discussed within the group or to be discussed with the IRT or whatever other instance?

SUZANNE WOOLF: Hi. That's actually one of the areas that is still under very active discussion: how to reduce the amount of time that we're talking about for this timeline.

Anything to add to that—

SEBASTIEN DUCOS: No.

SUZANNE WOOLF: —Matt?

SEBASTIEN DUCOS: I'll be sure to try to find a link and see if I can participate in the discussion then. Thank you.

KATHY SCHNITT: And no further questions—

SUZANNE WOOLF: Yeah. There's a lot of moving parts, so we're trying to make sure that they mesh rather than collide.

Go ahead, Kathy.

KATHY SCHNITT: Suzanne, just stating there was no further questions.

MATTHEW THOMAS: Excellent. Well, think we can give everyone 12 minutes back if we're done?

KATHY SCHNITT: Beautiful. Thank you. This does conclude today's session, and you may stop the recording.

[END OF TRANSCRIPTION]