ICANN74 | Policy Forum – NCAP Status Update
Tuesday, June 14, 2022 – 15:00 to 16:00 AMS

JIM GALVIN: Thank you very much. This is the NCAP Update at ICANN74, here on June 14th at 15:00 local time. Once again, thanks everyone for being here. I was a little concerned that we didn't have that many people in the room, but we have quite a few on Zoom. So that's a good thing.

This is an open meeting. This is our opportunity, the discussion group's opportunity to inform the community at large about where we are in our project.

I probably should have started by saying that I'm Jim Galvin, one of your co-chairs for this meeting. And matt Thomas, sitting next to me here, we are the co-chairs for NCAP. I apologize for not introducing ourselves first.

Even if you're in the room, we really would appreciate if you would join the Zoom room because we will use that for the queue management, as others have been doing here along the way. This is an open opportunity for folks to ask questions as we go along. If you want clarifying questions, please do feel free to raise your hand right away. We'll get to those.

But we want to get through all of the slides. There's a lot to take in here. Hopefully, you've had a chance to look at some of the slides ahead of time. And then we'll have some extended time at the end for some open discussion so that we can hear your reactions to all of us.

Even if you're seeing it for the first time, we'd love to get some reactions from the community, from folks other than just the discussion group. It's excellent to get opportunity and get initial reactions from people, especially if you have questions about what stuff really means. It gives us something else to consider as we really get to the homestretch here in the work that we're doing. So thank you for that. Next slide, please.

So this is what we're going to look at today. Very quickly we're going to zip through some background. These slides are here in part for the historical record, just to make it complete so that we always know what we're looking at and where we are. We'll talk a little bit about some completed work. Certainly, you've seen the public comments for a couple of items. And there's been a third report that's actually been produced that we didn't release for public comment, but it will be part of the final work product that comes out. So if you've been tracking the work on the mailing list, you would have seen that third work product.

We have already drawn some key takeaways and findings out of the work that we've seen in our discussions. We'll take some time to go through those and highlight that for you.

And really, I suppose, the most interesting part of the work that the discussion group is doing is the workflow and how we imagine the name collision analysis, the name collision assessment is actually going to occur. And so we're going to talk about a few of the details inside of that workflow so that you've got a sense of what we're looking at and what we're thinking about. And that's the place where it would be especially interesting to get reactions from people—positive or negative, or what questions pop into your mind.

And then we always have the "How to Participate in NCAP" quick slide at the end there. And then whatever time we have left is for open discussion. Next slide, please. And again.

Just a quick reminder, the reason why NCAP exists is because the ICANN Board had produced, essentially, two resolutions that they passed down to SSAC, asking explicitly for SSAC to seek to develop a framework, something to help them answer the question of what to do about .corp, .home, and .mail.

.corp, .home, and .mail were the three TLD applications at the time that were deemed to be of especially high risk, and so they've kind of been put in this deferred status while we figure out what it really means to assess the risk of name collisions.

And then, of course, they also wanted a framework of some sort, some advice on how to deal with name collisions going forward. As the ICANN community seeks to move forward with another round and future rounds of new gTLDs, we know that name collisions are here to stay. So the Board wanted to figure out, see if they could get some ideas about how they could manage that going forward rather than dealing with it as an issue that popped up in the moment in 2012, and they had to take some time to deal with it quickly then. And we did get some specific guidance about the manner in which to do this work, thorough and inclusive. And this is where this [working] comes from.

So, NCAP is not just in an SSAC work party. We sought to create an opportunity for community members to participate and be actively engaged in what we're doing, and that's why we're here Next slide, please.

This is just some links for you. The original Board resolutions, our project charter as was created and accepted by the community, the project proposal that was originally used and, of course, where we are. In the community Wiki, there is a whole project page which has all of our archives. Next slide, please.

They were three studies in the original project proposal—one on Gap Analysis, which has actually long since been completed and was delivered. We had a research analyst who did, essentially, an annotated bibliography of everything that we had learned about

name collisions, everything that had been done in the last decade since the last round.

Study 2 is where we are right now, looking at root cause and impact analysis of name collisions over the past decade. What have we learned since we started delegating TLDs, even though name collisions were present? Can we draw on that experience and somehow create a framework for how to deal with those issues in the future, maybe improve on the controlled interruption mechanism that was used in the 2012 round?

And Study 3 has yet to be done. That's a question that will be examined After we're done with this Study 2 work product, we then need to examine the question of how do you deal with mitigation options—whether there should be mitigation options, what they should look like, and how you evaluate them. Next slide, please.

This is just a look ... We've said this before in other fora, so this is partly just the historical record. This was the changes we made. This project is four years old now, so along the way we've learned some things. We tweaked the way that we worked. We had originally thought we'd build a giant data repository. We've learned that that really wasn't necessary. We didn't do that in Study 1. It got moved into Study 2 when we were doing Study 1. And then in this project, we decided we didn't really need that.

And we also didn't want to build a test harness for checking out mitigation strategies. That was an original topic that we were going to conduct in Study 2, and we decided that we didn't think that we were going to need that. Or we at least did not want to do it in Study 2, but we might come back to that question in Study 3.

And Study 3 has yet to be done. It really has a singular task of analysis of mitigation options, and we'll get to that after we're done here. Next slide.

It seems our computer is being a little bit slow, and that's okay. All right. This is a little more detail on Study 2. I've already covered this, so we can skip over that. And that takes us to Completed Work, and I will hand it off to my co-chair, Matt Thomas.

MATTHEW THOMAS: Thanks, Jim. This is Matt Thomas for the record, co-chair for the NCAP. As Jim previously mentioned, we have several items of completed work the NCAP discussion group has produced over the last several years, some of them being much more recent.

Specifically, one of the first things we published was a case study of collision strings, specifically focusing on six strings. The first three were obviously the .corp, .home, and .mail, as directed by the Board resolutions. We also added .internal, .lan, and .local. It was determined to add those just for a comparative measure

because, at the time, those three strings were receiving more than 100 million queries per day at A and J root servers.

The data for that study primarily came from A and J root servers using a longitudinal view going back five to six years that allowed us to highlight some of the changes of the properties of the DNS queries over time and how those alterations impact DNS risk assessments for name collisions and how they are the byproduct of DNS evolution, things like QNAME minimization, aggressive NSEC, and so forth. This case study was one of the first documents that the work group had produced.

The second is a Perspective Study of DNS Queries for Non-Existent Top-Level Domains. Ultimately, at some point in the future, data needs to be available to do name collision risk assessments, and this study was designed to understand the distribution of DNS name collision traffic throughout the DNS hierarchy.

Specifically, this test or study was designed to really help inform and provide some guidance and guardrails of when looking at name collision data from various points in the DNS hierarchy, what caveats or guardrails should be interpreted around those quantitative measurements that are being assessed. And so those insights are hopefully there to provide us new insights in terms of how DNS data can be used for name collision risk assessments.

The third completed work was the Root Cause Analysis and the New gTLD Collisions. This was work conducted by ICANN OCTO's technical contractor, and they were in responsible for reviewing the 47 name collision reports that ICANN received from 2012 to present for the new gTLDs. These 47 reports span several different TLDs in which the researcher was able to go back, quantify those, and provide some insights looking at DNS data to better understand how controlled interruption was effective and how much of a disruption occurred within those particular name collision reports. Next slide, please.

So what were some of the key takeaways that we had coming out of all of these pieces of work? Well, the first is the case study. The case studies, again, looking at .corp, .home, and .mail clearly indicated the impact has increased. One of the things that we did notice in the longitudinal component of it is that we saw heightened query volumes and diversity for all three strings. You could notice we see those for the COVID impact for those strings during 2020.

But one of the more important things that came out of the case study was something that we're calling critical diagnostic measurements. And these are a set of qualitative and quantitative measurements that we have been using to help predict the impact or assess the risk of name collisions. I'll cover those CDMs ,or critical diagnostic measurements, in just a little bit.

One of the other things we have noticed is that leaking collision strings differ from delegated TLD queries. The actual properties of name collision strings differ from what I'm traditional delegated TLDs look like. And finally, one of the more obvious things that we saw back in 2012 but still seems to be prevalent today is that DNS-SD protocols and suffix search lists are a major problem or root cause of name collisions.

Next, on our perspective study, some of the key takeaways there was that we can see similarities between or differences between RSIs and the public recursive resolvers. And this is kind of expected, since they're very different positions in the DNS hierarchy. But this was important for us to understand for when we're quantitatively assessing maybe potential top-end lists and their completeness and what kind of guard rails need to be provided on those types of lists when being used for name collision risk assessments.

And furthermore, this study also helped inform that maybe existing measurement platforms could be extended to help inform applicants. One of the goals here is to help inform applicants, a priori, of their application to allow them to be aware of name collisions for their particular string before they submit their application.

And finally, on the Root Cause Analysis document and study, three key findings came out of that. And one of them was that

private use of DNS suffixes is widespread. And it seems to continue to be widespread.

The name collision reports also were strongly supported by the measured data used in the study. Meaning that the ones that had higher elevated DNS query data in terms of CDMs also received larger numbers of collision reports. And, finally, that the impact of TLD delegation ranged from no impact to severe impact in some cases.

From all three studies, we have clearly seen that in name collisions continue and will remain an increasing and difficult problem. Next slide, please. So let's talk about some of the findings. Next slide, please.

So some of the current findings that we have coming out of the work the NCAP Discussion Group has conducted so far is that name collisions, like I said, are a problem and will be increasingly difficult. Case studies have also shown increased impact and that the DNS service discovery protocols and suffix search lists are a continuing problem.

We have found that critical diagnostic measurements, the CDMs, are a way to assess name collisions to inform the risk of delegation. Mitigation and mediation is difficult as those CDMs increase in both their volume and the diversity. Next slide, please.

So these are some of the critical diagnostic measurements that I previously described. As we said, one of them is query volume. But query volume is just one of the leading indicators of name collisions. There are other important attributes of DNS queries going up into the name collision context that need to be accounted for. Those are things like query origin diversity. How many different IP addresses are the queries coming from? How many different networks are they coming from? How many different ASNs are they coming from? What is that query origin virgin diversity?

Second is the query type diversity. Are these queries all for type A, for IPv4 addresses? Are they coming for TXT records? Are they coming for SRV records? What is the type diversity and how does that play as a factor into assessing the risk for that particular name collision?

Other things around diversity might include the labels, things to the left of the TLD string. Right? How many different labels? Second-[level] domains do they have. And what other kinds of other characteristics are present in those labels that might signify known types of risks? Things like WPAD queries or DNS service discovery protocols that have been shown to be exploitable in a man-in-the-middle type attack scenario. Next slide, please.

That will hand it back to Jim, who's going to give an overview of the workflow that we're going forward. Thank you.

JIM GALVIN: Thank you, Matt. So now we get into more of the original work, the work that we've created as part of being the discussion group. Everything that we've listened to so far is more on the factual side of things, things that we've learned as we've studied what we know. And so now it's a question of where do we go from here and what do we do.

So here on this slide, this is just sort of taking a bit of a step back and asking ourselves the question what problem are we trying to solve? The Board resolutions are certainly one specific characterization of what we're looking at, but as we think about this we need a methodology for evaluating and reducing the risk of delegation of a new TLD.

What we've come to understand and come to recognize is that name collisions really are always going to be there. There is an extremely long tail of name collisions. There is a certain challenge in coming up with strings that don't have any kind of collision metric, even if it's just a few. Now I mean maybe most of them don't hit any list on the high end, but there's always something there.

So you come up with this question of what do we do about that. And so we've been thinking about that, and we have very recently been characterizing the problem space that we're looking for in the following way. And that is that our objective here is to be

identifying high-risk labels, identifying strings that, in the Board resolutions, they call them collision strings which is those strings that probably should not be delegated. Those strings which are ... Can we find and come up with a methodology that will help to identify those strings that are likely to have very high impact and a high probability of potentially being debilitating?

.corp, .home, and .mail. sort of fell into that category based on the data that was at hand back in 2012. That was what they did. But the Board didn't really have a way to deal with this concept of a collision string at the time, so they're really just in the deferred status right now. So we're trying to come up with some kind of methodology that allows us to pull those kinds of strings out and identify what they are.

So we recognize that name collision analysis, in and of itself, is a risk management problem. It's not an absolute space. They're always going to be there, so it's about what can you do to help understand and manage and reduce the risk of delegating and trying to find the best that you can for pulling those things out.

So we're left with these two questions. Right? One is, is it possible to objectively identify a high-risk label? And if not, can you at least provide some guidance to identify them? Now is it possible that there is a particular number of queries like we had with .corp, .home, and .mail? We know how many DNS queries were there.

Can we pick that number and then decide? Or do we have to allow some flexibility there and some judgment.

And then from the point of view of Subsequent Procedures and that work and the recommendations that came out of there, there is some guidance and some particular ask of being able to identify Do Not Apply labels. Is it possible to tell a potential applicant that, "Well, you can't have this string and you're not going to have it"? How do you manage that process? And even if you can't objectively identify it just like a high-risk label, can you provide some guidance that helps move you down that path so that, as an applicant, you've got a sense of what's in front of you as you wait to see what the analysis begins to show?

So that's where we are in terms of the ... That's the way we've been characterizing the problem we're looking to solve lately. It's a risk management problem, and so how do we wrap some guidance and process around that to get us there? Next slide, please.

So our particular goals and our methodology are two things. One is to ensure that we can actually assess name collisions. We need to find a way and ensure that we can make them visible. And then once they're visible, you want to be able to look at them and form some assessment about what you can see.

And then of course the second thing is, even if they're visible, knowing that they're always going to be here, is there an

opportunity for mitigation or remediation? And what should that look like?

Now to be fair, right now that second bullet there is not a primary purpose of our work at the moment because that's a Study 3 work item. So that first bullet about ensuring that they can be assessed is where we are, and we're trying to make sure we leave some space in our process for the development of a mitigation or remediation plan. But the details of that plan are a Study 3 work item.

And something just occurred to me. I'm sorry. If you could go back one slide. There was a critical point there that I didn't make. And that's the second sub-bullet under the first item there. In looking for high-risk strings, rather than trying to characterize all strings, we're beginning to characterize our work in the form of let's find the high-risk strings. And those will be the strings that will become a special case that are likely to not be delegated and no other string would be blocked. Right? So you would simply choose, as we did in the 2012 round, recognizing that there were name collisions.

We've already made that choice to delegate in the face of name collisions, so what we're trying to do is find the lowest-risk option for doing that, for getting us to that path where you can delegate strings. So we're looking for a methodology to block high-risk strings and let everything else just continue on. Next slide, please.

Now as part of trying to figure out how to assess name collisions, we've come to realize that there are two operating roles that we need that have to come into existence in order to assess named collisions and assess where they are. So we're going to take a moment here to walk through those two roles as we describe them. Next slide, please.

One of the things it's important to know is that we've created this concept of a Technical Review Team and I want to be careful to not say that we don't have any consensus on exactly how these might be implemented. This is a functional role that we're declaring here. We're describing a role that should come into existence.

There are a variety of ways in which something like this could be done, and the ICANN community and ICANN Org has already done things like this in a variety of different ways. It might be something that can be a third party. It could be done in-house. It might be something in between, a little bit of both.

So want to be careful that we are talking about a functional role here. We need a set of independent and neutral experts to actually look at the data that's going to be collected. They need to have an understanding of DNS specifications and the Internet infrastructure.

One of the things that we've learned about the last 10 years is that the DNS infrastructure has changed quite substantially. So, that's

important. The decisions and choices that were made in 2012 were fine for that time, but by themselves they don't apply directly at this point in time.

And you have to also then acknowledge that the DNS infrastructure is likely to change over time. It's going to continue to evolve and change. That's what the Internet does. So it's important to have a set of people who can not only look at the data, but they understand the data that they're looking at. These are going to be people who are going to have to stay in touch with what's happening in the community at large and in the Internet at large.

And they have four explicit responsibilities. One, of course, is to assess the visibility of name collisions. And you'll see in a minute here as we get into some of this that name collisions are ... Because of the DNS infrastructure changing, you have to pull them into existence. You don't see them just by looking at the root server identifiers. And that is something which was drawn out of the perspective on the DNS queries, that second work item that was talked about before.

One of the things that was drawn out is that you can't just depend on the root server operators to tell you about the existence of name collisions. The second, of course—and this might seem somewhat obvious, but it's important to keep that in mind—they have to document what they discover. So as they do the

assessment on the data at hand, they've got to be able to document data, findings, and recommendations.

Bearing in mind that this Technical Review Team is not going to make a decision about what is or is not going to be delegated and what is or is not going to be granted to an applicant. Those decisions rest with the Board, ultimately. Their job is to identify the risks associated with that in a way that the Board members can evaluate and make a judgment as to what they want to do at that risk. Do they want to absorb the risk? Or do they want to say no? And that's the path that we're headed down here.

They'll also have a responsibility to assess a mitigation and remediation plan. Again, the development of that plan is not really part of the scope of this work, but it's important that that's a responsibility we imagine they will have as we get into talking about those.

And an emergency response. It was originally, even in the 2012 round, they stated that there needed to be a mechanism to turn off controlled interruption, for example, and come back. But one of the things that we want to call out is being very clear that these are the people who will make that assessment. They'll do that monitoring and they'll keep that in mind. And they will have a responsibility to act if something should become an emergency situation that has to be dealt with. If you delegate and something really bad happens, you've got to be able to pull back.

At this point, it's probably worth reminding people that we have not granted the TLD at this point. An applicant has applied for their application, but when these assessments are being done, the TLD has not been granted. It's still something which is part of the ICANN due diligence process, all of this. That's the way we're imagining this happen. Next slide, please.

This the other key role that has to happen. And just as with the other slide, I'll say again. We're not making any specific recommendation about how this must be done. This becomes another role which could be part of the Technical Review Team. It could be a different set of people. It could be a third party or not. This is a functional declaration of something which has to be present. And there's an opportunity to think about the most cost-effective way to implement something like this.

But just as there was controlled interruption in the 2012 round, we are describing a system whereby you will need to have servers of some sort. You will need to have a system which is deployed for a period of time so that you can collect data. That's essentially what controlled interruption, as defined in the 2012 round, did. You deployed a server. At that time, the registry operator deployed the server. And that was done in part because the TLD had been granted at the time, subject to whatever would happen during controlled interruption.

So at this point, again, a reminder. We're talking about a step that has to happen but the TLD hasn't been granted yet. So there needs to be the existence of some kind of neutral service provider who can operate the intervention that we're talking about here and using it.

And they have four responsibilities. This particular team would have its own set of four responsibilities to follow through on. One, of course, is to operate the authoritative DNS server for the Passive Collision Assessment. And we'll get to what that is in a moment.

They also have to operate the Active Collision Assessment environment. And we'll talk specifically about what those elements are in upcoming slides. But they actually have to do the technology. They have to roll that out, put it up, stand it up, run it, monitor it, and as the third bullet there says, do the log processing and analysis. So it's just the log processing and analysis of that log processing that they're getting. They're not doing the name collision analysis. They're just watching the logs.

Keep in mind that since they're running the server, they're also going to have the emergency response. So they're going to be monitoring all of that. Another part of the analysis which has not been definitively ... We don't have consensus on this yet, but it certainly is a question that has to be examined at some point here. These logs may or may not have a lot of data in them. And

in particular, as we are so fond of talking about in ICANN, we worry about PII. You worry about unnecessary private data that might be in some of these log files.

So part of this analysis that's going on here is, maybe at this point what you're doing is anonymizing the data at some level or you are aggregating the data in order to reduce the disclosure risk. So these are some of the details that still have to be examined and settled on, but the principle here of maybe the raw data is with this neutral service provider, the TRT—the Technical Review Team—may not see the raw data. They just see enough data so that they can do the analysis that they need to do. And those are details that we don't yet have consensus on, but we're aware of that problem space and we're considering how the best way is to approach that. Next slide, please.

This slide is actually pretty stable. We had started with a model quite some time ago, nine months or more, where the workflow itself looked like it was going to be these five steps. They're sort of obvious and sort of natural. The details of the steps have evolved quite a bit over the last six to nine months, what's kind of inside of them.

We're going to talk a bit about each of the steps. The applicant selects the label. It seems like an obvious kind of step. They have to prepare themselves for what they're doing. And then they submit their application. They're going to go through two

collision assessments—one that's passive, one that's active. Those are our labels for them. And then, of course, the Board gets the final package to make their assessment about whether or not to grant the TLD. Next slide, please.

So looking at the first step of Applicants Selects the Label. This is where the applicant gets their first indication of the presence of name collisions. So this will be the first of three opportunities to identify high-risk strings. This risk assessment will be done via a static list. Even today, ICANN does publish under the ITHI data, a list of the top 1,000, I think. Right? Yeah, it's the top 1,000 NXDOMAIN queries. So those names that are currently non-existent are present in that system.

Now that's just a peek at what it looks like to root server identifiers. It's just a peek at what it looks like on the L root server. So it doesn't tell you about everything that you need to know, but it is just a leading indicator. If you exist on the high side of that list, then the only thing that you know at that point is that you are at risk of higher scrutiny.

So it should give you some pause to think about that because your presence on that list just means that your number of collisions may or may not change much. But the way it will change as it can only go up. Your ranking might go down based on the integration of other data, but the number of collisions you have can only go up as we collect more data. And it's important

for you to keep that in mind and see what you're doing. So that'll be the applicant's first indication in the first risk assessment.

There is an open question here of whether an applicant might have the opportunity to ask the TRT for its thoughts about the high-risk conditions. We don't have consensus on that point yet. It's still an open question in this step. But in any case, the applicant will have that data to simply look at themselves, make the decision, and then follow the rest of the process.

And as it says at the bottom there, then we jump to Step 2, Applicant Submits the Application. There's nothing in that step which is in scope for NCAP, so we'll go to the next slide and jump right to Step 3.

Now in Step 3, this is the Passive Collision Assessment. What we've done here in Step 3 and Step 4 in these passive versus Active Collision Assessments, we have learned about the controlled interruption that was done in 2012—two things. One thing is that it's possible to do something like a controlled interruption step but to do it in a less risky way. The interesting characteristic about the controlled interruption that was done in 2012 is that it was a fairly disruptive interruption. It actually changed the behavior that was visible to the client. It changed the behavior that was visible to the majority of clients when you delegated the TLD.

What a Passive Collision Assessment does is, you do still delegate the TLD, but you delegate an empty zone. The zone has no content in it. And so in principle, most cases will result in no change at the client. The client will still ultimately see an NXDOMAIN response. So the client behavior is the same. There's an extra DNS query for them to get there. But in principle, for the majority of cases, the expectation is that they'll still get their NXDOMAIN.

And for the moment, we have not found any objection to that. We've been asking around in various technical communities, “Is anyone aware of a high-risk edge case in doing this?” We have not seen any objections to it. People have identified issues, as we know, with public suffix list and the way certain things are done and certain kinds of enterprise configurations. But as compared to the original controlled interruption, it's still a lower risk.

You still have those lower category of things. Some number of people are affected, but this is something that where we're settling on recommending at this point. And it is after this ...

So this passive collision assessment would be deployed and would run for some period of time which we have not yet decided. The 2012 round of collision assessment ran for 90 days. Just speaking personally, I suspect we don't have to run this for 90 days. But the details of that have not actually been discussed in the discussion group yet.

And you'll collect CDMs on that data that you have there. And then the Technical Review Team will do an assessment of this new data that it has. So the advantage of this data is that it's actually data that you get at the authoritative server. You will get the advantage of high-quality second-level domain data which you would not get from the root server operators. Root server operators are increasingly seeing less and less of the second-level domain data.

So you'll get better quality second-level domain data from the authoritative server. And in addition, a clear thing that you'll get is that you will see data out of public recursive revolvers. The public recursive resolver data, you won't see. You won't see those queries just looking at root server operators. But if you deploy an authoritative server, then you get to see that data. That data gets pulled because the query has to be made to the authoritative server. So it pulls that data out and makes it visible.

So that's why we say that the numbers that you saw on the static list can only go up because all that's going to happen here is you're going to see more data. And the Technical Review Team is going to have all of that data to look at with respect to DNS queries. So the rank might change, but the numbers will be what the numbers are, and they'll potentially go up.

Let me just give that a moment here. See if any hands come up about that. Looking around the room here. Let's move to the next slide, please.

So what is the difference with Active Collision Assessment? Well, Active Collision Assessment is very much like the controlled interruption that was done in the 2012 round. It is essentially something which is disruptive to client behavior because you're actually going to provide a response to second-level domain queries. You are going to put up a wildcard response to those queries, as was done in the 2012 round. We are going to tweak that a little bit, though, because we're looking to make that experience a bit more valuable.

One of the things the root cause assessment document did for us was to clearly indicate that a low-controlled interruption had a goal of notification to the client. It really did not do that as effectively as was hoped. It didn't really inform the client any more than the fact that things weren't working.

So one of the things that we're proposing here is that we actually respond with an actual IP address, both IPv6 and IPv4, in the response. So, a live one. So again, you're going to set up an authoritative server. And in fact, now this time instead of only gathering, again, DNS data, which is what was done in 2012, we're going to recommend adding other protocols and gathering data about them.

The web is sort of an obvious choice. So in addition to Port 43 for DNS queries, you would listen for Port 80 and 443 for web queries. And you would then be able to respond by displaying an informative page to the user. So that presumes, of course, that someone's making a web query as part of what they're doing. So again, you don't know if that's what they're using. But you would have a more direct notification right to the user about name collisions if you apply that for those who are using web protocols.

We have a continuing discussion at the moment about what other kinds of protocols we might listen to. There's a bit of a list of them that are being considered and thought about. But the idea here is to add additional notification mechanisms. So the notification mechanism in the 2012 round was just the "magic IP address," returning 127.0.53.53. Here in this case, we will choose a set of responses, a different response PR protocol that is being examined. Just as we said with the web, you could put up a nice informative page about name collisions. Ideally, we'll be able to provide a useful and helpful pointer error message regarding other protocols. And that's where we want to go with that.

And this is where the Technical Review Team ... After that, [you run] that for a while. You collect the CDMs. And then the TRT does their third assessment. Again, the idea is to identify high-risk labels. So looking at the CDMs and based on greater diversity and greater volume in all of the CDMs, that increases the risk.

And it's up to the Technical Review Team to characterize that risk in a way that's meaningful to the Board so that you can translate the technical issues that are going on into something they can make a decision about, whether or not they want to absorb that risk or not. This is also where a mitigation and mediation plan would fit into the workflow.

Again, most of that work is really going to happen in Study 3, but it seems that if you're going to end up being put into the high-risk category, which means you're in the collision string category, which means you have a high probability of not allowing for delegation, there comes the question of, well, you should have the opportunity to make a proposal as part of your application to indicate what you're going to do about those name collisions that are happening that have become visible. And this would be the place where that would happen, if that were to occur.

The only thing left after that is that the Board, in Step 5 there, gets the final package. And the final package is the application and all the rest of the due diligence that, of course, the whole process creates for the Board to evaluate. And then, in addition, it includes these assessment—the two of them that are written by the TRT and sent up. And also, if there's a mitigation and mediation plan, that would be included. And that would be the end of that.

I'm seeing a question in the chat room here. "Which criteria will the Board be using to make/take their decision?" It'll be about ... The risks will be identified by the Technical Review Team, and the Board will simply have to decide whether or not it agrees with that risk or not. It's just part of an ordinary risk management process and risk management program.

So, will there be hard criteria? No. In general, there really isn't even in a risk management program. That's what makes it risk management. It tends to be a subjective kind of evaluation. So you're quite dependent on your Technical Review Team to be thorough in what they describe about what's possible to happen so that the Board can make that decision.

And then the next question was, "Knowing all of this, does it not allow for gaming?" So, gaming is interesting. It's come up a few times in the discussion group. We haven't really dug into it in detail. Gaming is certainly on the list of issues that SSAC had identified early on as something that it wants to be concerned about from a technical point of view in evaluating all of this.

We don't have consensus yet in the discussion group about if we have any guidance to deal with gaming. At the moment it feels like we're really dependent on the TRT. Part of its job in assessing the risk and seeing that name collisions are visible is to have available for itself trending data. It needs to be able to see a history of some of this data, the data that gets at ITHI. Maybe

there's other data that it can use over time. It, of course, will develop a history over time of applications and application data. And all of that will be available to it.

It's possible that gaming may be something that the TRT is going to keep in mind and address. It's going to have to be aware for itself of what's been present historically, what's present now, and have some judgment about whether or not this looks suspicious and call that out if it is.

So the rest is just discussion. Next slide, please.

Actually, we're moving really into discussion here, so next slide. This is just for the record here. Do you want to join the discussion group? There's a way to get to it. You can get to it from the slides. And moving on down. Next slide, please.

We are into Q&A. So it really is an open floor at this point to ask questions. Reactions? Any immediate reactions about what you've heard or seen? Looking for hands and watching the chat room some more. I do see a hand from Peter Thomassen. Peter, are you in the room here? Well, go ahead, please. Open your mic and speak.

Okay. His mic fell apart for him. Oh, yeah. Any other hand? Any other questions? Do you want to type in the chat room? I'll give him a minute to get his microphone on the table plugged back in.

PETER THOMASSEN: Okay, it works. Can I speak now?

JIM GALVIN: Yes, you can. Go ahead, please.

PETER THOMASSEN: [inaudible]. Another person had a question. I have three questions, in fact. Separately or [inaudible]?

JIM GALVIN: Well, let's take the first one and then we'll see how easy or hard it is.

PETER THOMASSEN: Okay. I think they're all easy, probably. The first question is if there is evidence that the less-intrusive approach where the local IP address was provisioned in the child zone, in the TLD zone—127 or 53, for example—if there's evidence that that was insufficient and whether collision victims in the past would have preferred the more active approach where an actual web services is deployed, for example, with a warning page.

JIM GALVIN: Are you asking about the applicant making a choice of what they want to do?

PETER THOMASSEN: No. I'm asking about the Active Collision Assessment phase. And the argument was that in the past, in previous rounds where new TLDs were introduced, the approach was to create a zone that has A records that are local—127 or 53, for example. And my question is whether victims of collision at that time did find that insufficient and would have preferred if a web service had been deployed as is now proposed.

MATTHEW THOMAS: Thanks for the question. Let me try to answer that in two ways. First of all, I think the answer that we have so far primarily comes from the Root Cause Analysis from the name collision reports. The technical contractor also conducted some outreach to network operators around name collision experiences that they have occurred, specifically around controlled interruption and how effective and informative that was for them.

So inside the root cause report, there was some initial data—[albeit] the sample size is relatively small—that indicates that it wasn't as effective as an indicator for knowing that you had a name collision problem. Right?

Effectively, 1.70.53.53 was a mnemonic that required you to observe that in your logs and be proactive to go out and actually do a search to find information about that. So doing alternatives like web pages and stuff like that might be a little bit more direct report to the end [inaudible] end user, if that makes sense.

PETER THOMASSEN: Okay, thanks. It does make sense. Thank you. I just wanted to know if the motivation came from the past victims or not, but it comes from the Root Cause Analysis. Okay.

For the record, I'm Peter Thomassen from the SSAC. I forgot to say. The second question I have is how long would the Active Collision Assessment period be?

JIM GALVIN: Obviously, applicants are going to care about that. And we haven't talked about that in the discussion group. I did mention briefly. Right now, controlled interruption is set at 90 days, the original 2012 round. It is interesting to think about, what we're really talking about here with two different assessments, is doubling that period of time. So it is a good question to ask, as to whether ... Does it really need to be 90 times two, so it's 180 days? Or can we pull back a little bit?

We have not really examined that question yet. So that's an open question.

PETER THOMASSEN: Okay. [So it's open to me]. I have no opinion there. I just wondered if there was any arguments. And the last question I have is if a web service or [inaudible] serviced or whatever is

deployed, then those connections would be terminated at an ICANN-provided service. And I can imagine that some people could have privacy concerns that they end up accidentally sending authorization headers or whatever to such services. And you can argue that if the TLD was instead delegated and somebody else was registering a domain under it, the same thing would happen with the service that is then deployed there.

And my question is why would it be admissible to terminate such connections in the pre-delegation phase at an ICANN-provided service and at the same time later disallow the delegation of the TLD? Because you could argue that if it's okay to send such traffic to an ICANN service, it could also be okay to just delegate the TLD and send the traffic to wherever it would go then. I mean, it goes to a party that it's not intended for. What's the difference?

JIM GALVIN: I'll say two things. One is, we are very sensitive to the question of the existence of private data and how that needs to be handled. The original 2012 round of controlled interruption went in that direction versus any kind of honey pot-like solution, especially because they really did not want to take in any private data.

The place that we're at is ... We're in a place where we believe that we have to take that risk. We have to learn more about the name collisions that are present. And so we have to allow for the risk that we might collect some data.

Now we can mitigate that to some extent from a technical point of view by, yes, I'm going to open up a web connection in particular. The HTTP protocol is the perfect example of something which you could get a lot of data in that first blob of things coming at you that really is PII.

So the moment, we know that we have to call this out as an issue, it might be that the best that we can do is the neutral service provider has to scrub the data on the way in so they only deal with what they have, and so you have to take the requirement that part of your implementation has to make sure you scrub out anything you're not using. You don't have to make an assessment about whether it's PII or not, but if it's not interesting to you, make sure that you scrub it right away. So you don't even put it in the logs and take it in.

That really is more of a legal question, though, not a technical question. We're in a place where we are trying to provide data for use by a technical team, and we want to make sure that we can collect enough data for them to do what they need. And we will have to find ways as best we can to mitigate that risk of PII.

Now, the difference between treat pre-delegation and after delegation. One of the benefits of doing it after delegation is you've got a registry operator. And, sure, you could just have them do that. Then you could do all of this then.

I think that our goal here, as we understood it, was to be able to assess name collisions and that risk of them so that the Board could decide in advance whether or not he wanted to delegate, rather than deciding to delegate the string based on criteria other than name collisions and then saying, “Gee, we might take it back if the name collisions show something.” They want to be in a place where they've made a decision and you give it to them.

That's the way that we have interpreted the question and the scope of what we're doing. So it is fair to say that at some point here, somebody could decide to do it afterwards if they wanted.

The other key thing that you lose by doing it afterwards is, you lose control of the data. If you want the TRT to have a role in evaluating the data and being able to do a risk assessment, then you have to make sure that it has access to the high-quality data. And if you just have the registry operator do it, then you have a third party who's got the data. And now you have the TRT team ...

Now you've got to have some kind of relationship there, and you risk ... It's a different kind of relationship. And ensuring that they're going to have the right data to do is a problem, too. So again, we're looking at it from a technical point of view.

And before you get to your third question, if you want to respond to that, that's fine. But I also know one of our committee members wants to respond to you also.

PETER THOMASSEN: Yes. So in fact, those were all of my questions. I have a very quick response. You pointed out that the privacy issues is legal. That's true, but it's also a security issue. And my point is precisely that it's difficult, in my opinion, to weigh that risk against the risk of not reaching everybody by, for example, just announcing a local IP address. Right? Because doing the more passive assessment has a higher risk of not reaching people. Doing the more active one has a higher legal and security risk. It's not clear to me what the balance is. That's just a comment. It's not a question.

JIM GALVIN: Okay. Thank you. Before you get to your third question, we do have again in the back here. I'm wondering, are you available to come to our discussion group meeting in the next session?

PETER THOMASSEN: Yes, but I already did ask all three questions.

JIM GALVIN: Oh, okay. All right. Oh, you asked your third question, too? Okay. We have a hand in the back. Edmon, please.

EDMON CHUNG: Edmon Chung here, speaking individual, I guess. I'm .asia. A couple of questions, actually. One is, especially, the active part ... Do I understand correctly that this is expected to replace the controlled interruption process? Or will control interruption still be required after that? Or how do you envision it?

And the second question would be, hopefully interesting and whether it's in scope or out of scope. Given that this is a process whereby it is possible that an applicant comes in with an application and then finds out that it is rejected, is there a consideration to, at least in the report, to talk about asking the GNSO to reconsider some of the things, whether at that point, the applicant can change the string? Because there's no tool that they could previously figure out whether it will pass this test.

I don't know whether the NCAP would at least touch on this issue or not. If you will not touch on this issue, that might ... Then my question is, the SubPro probably hasn't considered this scenario. How do we get back to it? It might be a question out of scope, but I thought it might be interesting to bring it up.

JIM GALVIN: I'm going to let Anne speak to the second question that you asked, since we are at the top of the hour here. So we're kind of time constrained. But go ahead, Anne.

ANNE AIKMAN-SCALESE: Thanks, Edmon. I believe that would end up being governed by the final report from SubPro with respect to Application Change Requests. And although I don't think that a name collision issue is specified in Application Change Request Policy. I do you think that, in any case, it's the Application Change Request Policy that's in the final report that would apply. So I volunteer to look it up, and hopefully I can get back to you on that one.

JIM GALVIN: The shorter answer to all of that is that it's not really in scope for us how the actual process works in the large—those business kinds of issues. Although, we would want to try to document pointers to things that have to be dealt with.

And then very quickly on your first question, yes, this whole process here would replace controlled interruption as it was before. That is the intent.

And I think with that ... I don't see any other hands, which is a good thing. I appreciate that. So we're going to end this meeting here.

A bit of logistical business. If you're going to come to the discussion group meeting ... And we would welcome folks to come and attend. I know that on the agenda it shows as closed, but you will be welcomed into the room. You are allowed to come.

If you're in the Zoom room, you'll certainly be welcomed into the Zoom room.

As it turns out, the physical room really is smaller. It only has room for 20 people, and it turns out there's just less than 20 people physically here right now. So those who are here could come down to the other room. It's Europe. It's one of the continents. It's on the other side of this building, just walking straight in. But others can join from the Zoom room. And, of course, you could come to the door and once we're all seated, if there are spare seats, first come first serve, you can come sit down with us.

So thanks very much, everyone, for joining us here today. Really appreciate the feedback and the questions. Please do come to the discussion group. It also is an open forum. You can feel free to ask additional questions there and talk more to us at that time.

And with that, we are adjourned.

[END OF TRANSCRIPTION]