KIM CARLSON: Thank you and welcome to today's NCAP discussion group call on January 13th at 19:00 UTC. In the interest of time, there will be no roll call. Attendance will be taken by the Zoom list. Kathy and I will update the Wiki with the names of the participants as quickly as possible. We have one apology from Rod Rasmussen. All calls are recorded and transcribed and the recording and transcripts will be published on the public Wiki. As a reminder to avoid background noise and echoing while others are speaking, please mute your phones and microphones. And with that, I'll turn the call back over to you, Matt.

MATT LARSON: Thank you, Kim. And good afternoon, everyone. Welcome to the weekly NCAP call. It's hard to believe it's another Wednesday already. Not sure where the week goes. Why don't we just go ahead and get started. Hopefully, I have some new and exciting content that no one has probably seen before since—unless you went and looked at the slideshow presentation already.

But with that, does anyone at this time have any updates to their SOI? I see no hands so we'll take that as a no. Just for the sake of being complete, Jim, do you mind giving a two second update on where we are on study two and just a reminder on the dates?

JAMES GALVIN: Yeah. We were expecting the close on Monday here of comments from the SSAC for their review of study two. We do have one comment back

Note: The following is the output resulting from transcribing an audio file into a word/text document. Although the transcription is largely accurate, in some cases may be incomplete or inaccurate due to inaudible passages and grammatical corrections. It is posted as an aid to the original audio file, but should not be treated as an authoritative record.

that we just want to figure out how to fold in. It's just an administrative detail.

We've got to sort that out and we will. We'll let people know exactly what we do in the proposal but hopefully we'll get that done here in a day or so and get the documents sent off to the BTC where its next meeting is the 28th. So, that's the next deadline to wait for a change.

So, we'll let you know when we submit it off to them and what it looks like if we've made any changes and then we'll wait for that date to see what the Board does. Thanks.

MATT LARSON: Thanks, Jim for that update. Going on, the agenda for today has posted that we were going to continue off the questions from the previous NCAP call, specifically, we were going to take a look at the mail QTYPE and then start taking a look at [.internal].

I'll be honest and I started pulling the data for .mail and I went down the rabbit hole and started looking at the data for it in a lot of different aspects. And it made more of this week's call focused on .mail, maybe the first version of the case study for the .mail that we can take a look at.

So, I do apologize that there's maybe a little bit of a discrepancy between what was listed there but we will get a chance today to hopefully dive into .mail in detail. With that, I'll paste the link in there and thanks, Kim, for presenting.

So, hopefully, last week was a good high-level overview. And before you start doing research or you start drilling for oil, it's always good to do a survey of the land to just get an understanding of where things are so hopefully last week's high-level overview level-set everyone.

But now we're going to start actually getting into the strings and the name collision data of interest. And I think just to preface this, I think there are two questions that we should always keep in the back of our mind on this and that is when we're looking at this data, one of the things we need to ask ourselves is where is the harm, right?

And number two is how do we assess it? And so, hopefully, as we get into this presentation, we can start to think about that specifically for .mail. All right. Kim, if we could switch to the first slide, please?

Thank you. So, as we may have mentioned in previous NCAP calls, Verisign is committed to serving or analyzing root data and we've been collecting high fidelity root data for multiple years now going back into 2017 and somewhat before.

But this presentation is going to look at data captured from A and J root servers with an extra—I'm going to call it special old J-Root there. So, just for a little bit of route history, in 1997, J-Root was added as the 10th name server and it was initially co-located with A-Root but it used the IP address 198.41.0.10.

Later in 2002, it was renumbered to a new IP address and it's been that since then. Since 2002, Verisign has continued to run a root instance on that IP address because for whatever reason, it still receives a fair amount of traffic.

So, while I don't anticipate it to be earth shattering in terms of query volume or moving the number in terms of queries for various different strings, I think it might in some cases, potentially be interesting to see if there is some kind of affinity with strings in this old legacy kind of root.

So, just for that reason and the other reason is that it's very easy for me to do that, I've included that in this analysis going forward. The graph on the right is a total daily query volume for .mail going over the last four years.

It kind of as you can see through 2017 to mid-2018 was relatively stable and didn't have very much variance but then suddenly had a pretty significant drop. Since then, it has ramped up and then again here in April of 2020 had a sudden another drop.

And since then, the traffic has been going down and they're approaching near four-year low levels. The three graphs on the right show a breakout of each individual route by query volume as well as by IP version.

As you can see, that significant drop was actually at A-Root back in 2018. And when we move to the next slide, it might become a little bit more evident of what was going on. I have a little bit of a theory there. And this was—the original ask was that we wanted to take a look at the QTYPE distribution for .mail.

I think there was maybe a suspicion that a large percentage of the traffic was actually for MX QTYPES but in fact it's actually mainly A in quad A and the long tail of the QTYPE distribution seemed to be similar to any other kind of TLD that you would see delegated.

There's no special affinity for a particular QTYPE and like a string like .arpa or something like that. But going back into that upper left-hand corner, traffic for mail at A-Root really seemed to be split 50/50 in terms of A and quad A. It's almost like a happy eyeballs instance that you would expect there but then it had a sudden drop.

And while I can't prove it, my gut tells me that this was probably something systematic that there was just not enough variance in there and then something suddenly got shut off and then the growth continued.

As for the drop again back in 2020 in April, I mean, I haven't done a detailed analysis but my only hunch there is given the context of what happened last year was that is around coronavirus and OCTO had made presentations in terms of seeing different changes in the DNS, maybe this is something related to that. I honestly don't know. I'd be curious if anyone else had any theories. If we can continue to the next slide, Kim? So this is—sorry, Jim. Yes, please go ahead.

JAMES GALVIN: Yeah. Before you go ahead if you go back to the slide, I don't know that I have a question about the sides. I just wanted to come back to the question that you started with here that we have in front of us, right? I mean, where is the harm and how do we assess this?

And so, what's interesting to me here is, yes, we're seeing some just sort of spikes or odd drops and you're saying yourself we don't really know what that is. I think the question that we have to decide for

ourselves and figure out is, does this represent harm or is this an Internet vagary?

And I heard you say that you don't really know but I just wanted to highlight the question again. And I don't know if there's any way that we can get more information or really dig into it or not but that's just a question that I have in my mind as we look at these interesting anomalies across the data. And I wonder if you have any additional thoughts and just wanted to reiterate that for people here too to be thinking about. Thanks.

MATT LARSON: Thanks, Jim. And I think that's an excellent point and as we hopefully continue through the analysis and we start looking at labels and source distributions and network affinities, maybe that will help paint or give some color to where is the harm or is it just vagary of noise in the background.

Yes. So, this graph is similar to the previous graph except I removed all of the A in the quad A queries. This was just really to kind of zoom up and see what was the long tail and specifically look at the MX QTYPE in there.

And you have to almost switch back and forth between the graphs but these scales here on the Y-axis are significantly lower than what they were before. These QTYPE queries make a very minimal amount and they follow a standard distribution of what you would see on other TLDs that are delegated.

Although there is some MX, I wouldn't say it's outside of the norm of what you see in the other ones. Yes, Jim, I see your hand up or is that an old hand?

JAMES GALVIN: Yeah. So, you made reference again here to what's ordinary for other TLDs and I'm wondering if at some point here it might be possible to have a representation of what is the affinity, what's typical in other TLDs. Is that something that you think you could do?

MATT LARSON: So, you're suggesting kind of take a baseline measurement from the delegated TLDs and take a look at QTYPE distribution and see if we can compare those distributions for different trends? Is that what you're thinking or?

JAMES GALVIN: Yeah. I mean, you're the one who just made the reference about this looks like other TLDs and I'm like, "Okay, can you show that to me what you mean by that? Can I see that in data as opposed to you just saying it?" Does that make sense?

MATT LARSON: That makes absolute sense. I think I can pull that and normalize that onto a scale and so we can compare apples and apples hopefully.

JAMES GALVIN: Thank you.

MATT LARSON: Thank you. Good suggestion. I think we can probably go forward to the next slide. So, this is taking a look instead of at [track] query volume. This is looking at the number of unique sources that are actually sending queries for .mail over time.

And although we've seen the traffic decrease, we're starting to see more and more source IPs either both IPv4 or IPv6, sending more queries for .mail to A and J. So, that is not exactly what I was expecting. I would've seen that there was somewhat of a correlation between less traffic and fewer sources. But this seems to indicate that .mail is being requested from a larger set of IPs out there. If we can move forward to the next slide, please.

And those IPs seem to be coming from a very long tail of countries. While there's a significant amount of traffic coming from the U.S., a little bit over 35%, the country distribution pretty quickly dips into the long tail.

I mean, you have France and Australia with a little bit of heightened affinity for the string but the traffic otherwise seems to be coming from a very dispersed set of sources. If we can continue to the next slide, please. In fact—sorry, go ahead, Jim.

JAMES GALVIN: Yeah. I know you're going to get into other kinds of looks at across the .mail but I guess maybe this is just a question to put out here and

wondering if there's a way for us to get at this data or not. But what is making .mail queries? Has anyone that you know of done any kind of backwards look in what is making the .mail queries? Or are we going to get into that later when we look more at—I don't know—Jeff, I see that Jeff is here with us too, if he has any thoughts about that from work that he's done in the past so an open question. Thanks.

MATT LARSON: Thank you, Jim. Yes, I think maybe when we get into the label analysis that might give us a little bit more sense of what is issuing these queries. Yes, Jeff Schmidt, thank you for attending today. I know you and I exchanged a couple of emails. I appreciate you making time to join today. So, I don't know if he would like to chime in now or if you want to hold it for a couple more slides when you have the data relevant.

JEFF SCHMIDT: Yeah. Hi, folks. And we found, we JAS, long ago when we looked at this, we found a set of sample Sendmail configuration files. They were published in one of the O'Reilly books, I think. I was actually trying to go find this in preparation for this meeting but I wasn't able to.

But we found a set of Sendmail configuration files that had .mail in them, hard-coded in them that obviously if anybody copies and pastes, we suspected that that was responsible for at least some of the behavior. We have some theories about other systemic issues here and there but the only rock-solid contributor that we found were those Sendmail scripts.

MATT LARSON: Thanks for that, Jeff. That's good to know. I'll make a little note of that. Going back to talking about source distribution and growth, the graph on the left is taken from the last day of 2020 so December 31st. And here I actually aggregated the IPs out to distinct autonomous systems or ASNs.

And so for .mail, A and J on that day, they received approximately 980 distinct ASNs requesting various different mail strings. Now, if you look at the CDF curve where the elbow starts to turn there, you're out at roughly 100 ASNs make up 85% or 87% of this traffic.

And so this is—to me, when I look at something like this compared to other strings that we've seen or I've seen at the root and worked with folks to remediate, this is fairly diverse. This set, to me, tells me that you're going to have to have a pretty large outreach effort to remediate a good amount of traffic, unless the names being queried seem to have commonalities that they're all using the same SLD or some kind of property in which it's more likely associated with some kind of software vendor that you'd be able to have them fix or alter that and change the traffic accordingly.

I can mention other strings that are much higher query volume than .mail where the ASN distribution is over three ASs and 95% of it is coming from one. So, mediating that kind of a name collision scenario is much different than something like this where the traffic starts to get pushed out into the long tail over a wider, more diverse set of sources.

Just anecdotally, I pulled really quickly I don't have this in the slides. As of yesterday, .mail ranks at the 184th most queried, non-existent domain at A and J. So, I know we're tasked to also look at CORP and HOME just remind you that those are both in the top 10 so this is essentially a magnitude lower in terms of popularity or demand for query volume up there.

Now, looking at the specific ASNs on the right, you can see that there's a handful of ASs that you might be able to reach out and you'd still get up to 30 or 40. But the actual networks behind those, a lot of them turned out to be Internet service providers for various countries in Australia or France.

There are a couple of companies that seem to be anchored in or using .mail internally. I'll say the first one, AS1504A is a large American insurance company. The query traffic out of that definitely seems to suggest that they have used .mail as an internal string. But, yes. Maybe we can go forward to the next slide when we actually start taking a look at some of the labels coming in here.

So, on the left, this is over the course of the month of December, all of the queries, looking at the number of labels present in the queries that were received at A and J.

And we've talked about this before about the implications of QNAME minimization but over, what is that? 56%, 57% of the queries coming in, only contain the label mail. So, outside of that, you only have, for context, the QTYPE and the source IP address of it which doesn't give

you a lot of interesting insight into it in terms of being able to figure out what is causing that and is it potentially harmful or is it just vagary.

The rest of the traffic then seems to be mainly using up to two or three labels. But I will say if you look at the middle column which is ranking the most popular SLDs, the second row, the underscore is actually another QNAME implementation instead of just truncating or deleting unnecessary labels, they changed those unnecessary labels to a single underscore.

So, this actually probably puts QNAME minimized traffic for .mail upwards to 63% which means 63% of the signal that we're trying to analyze, we only have essentially the QTYPE and the IP address for helpful clues in my opinion.

But the rest of the list, some of it, it seems obviously tied to various popular mail providers. So, string number one g.mail or number 11, hot.mail. So, there seems to be some association or affinity with that but then you also start to see some general infrastructure names coming in there too like NS1, NS2, Line 6 WPAD which we've talked about before.

And then in the column to the right for the ones that did have three or more labels, these are the most popular labels there. Again, WPAD which we have known and it is known that this is a potentially dangerous protocol in the context of name collisions is number one.

But the rest of the strings that seem to have this WinHex pattern—when I did a little bit of Google-fu on those, they seem to be tied with window exchange servers that are offered by various hosting companies

in Australia and France or Europe and they used those exact strings in terms of their suggested host names.

So, if there's a connection with that, it seems possible but the rest of the names in the third mail, I don't see anything outside of the standard things that you would associate with the mail protocol like labels like mail or POP3 or IMAP, so forth and so on.

Item number 15 on that list on that far right column, MSOID, that's another good indicator that this is all likely linked to Microsoft Exchange. MSOID is another automatically prepended label that is used for configuration management. It's like an ISATAP or WPAD in terms of that way so I think that gives us a little bit more sense that this is probably—a good portion of this might be coming out of Microsoft environments or system. If we can go to the next slide, please.

So, this slide is again taking that SLD list from 2020 and comparing it to—I think, Jaap, mentioned this last call that there was a publication done in 2017 looking at CORP, HOME and MAIL.

So, I grabbed the table from that and tried to put it together. Unfortunately, you can't easily copy and paste a table out of a PDF and put it into Microsoft Excel very easily and do this. I apologize for that. But there are a few interesting things in here I wanted to just point out that, it seems that there are certain things that have definitely changed over time that certain strings, at least in 2017, no longer seem to even be present in the 2020 data.

For instance, system.mail doesn't seem to be up in the top, neither does army or navy. And I don't know, Jeff, if you'd like to re-chime in here

about the work that you've done on those and just give me maybe a little bit more context around those strings.

JEFF SCHMIDT: Yeah. We identified an issue where the DOD had some infrastructure that was dropping .MIL from queries exiting their network. And so that exposed what was previously the second-level domain as the top-level domain which then resulted in an Internet query into that top-level domain.

You can imagine badness associated with that and we brought it up privately with the DOD and worked with them to get it fixed. I have a suspicion, and we were emailing a little bit about this back and forth yesterday or over the weekend or something, I have a sneaking suspicion that this might be related to that issue which I also know now has been fixed for a couple of years so that would also explain the change in behavior. When you see things related to service branch as the SLD, that was a very specific situation that has been fixed.

MATT LARSON: Thanks, Jeff, for that. I appreciate that. The only other thing that I wanted to highlight on this and maybe bring up for discussion was the concept of measuring and ranking domains based off of total observed query volume or observed daily sources.

And I'm going to state that I have a little bit of a bias that I don't believe average daily sources is an efficient metric to give you a true sense of risk around that particular string or subset of strings. That being said, it

is useful, but I think it needs to be in the context with some additional network parameter metrics either number of unique /24s or more properly the number of unique ASNs.

When you have large companies or Internet service providers that own /8s out there, that's a whole lot of IP addresses but if you looked at the data on a per AS basis, that number is much more manageable. So, I think moving forward, if we make recommendations in terms of various different measurements that we should be doing to calculate and start to assess risks, I would suggest that we include expanding the number of unique daily sources into various other network [cuts], either /24s or ASNs specifically.

And I think that might be a little bit more clear as to why, when I start to look at some of the next slides with you all. If we can move to the next slide, please.

And so, outside of just looking at the names and the source diversity for mail which was maybe the starting point in terms of how you start to assess risk for it, I also wanted to cover the other aspect that we've talked so much about and that is data sensitivity.

How do we ensure that when risk assessments in the future are being conducted, that the data collected from whatever entity at that point in time is representative enough to show the actual or give confidence that we were actually measuring and conducting the correct risk assessment?

So, the next four slides are going to take a look at just that. So, again, this is data all from December of 2020 at A- and J-Root but the graph on

the left is only for the last day of the month on the 31st. And this is taking a look at which ASNs sent the queries to which root, either A-Root or J-Root, right?

So, you have this classic Venn diagram that clearly shows that there is some overlap but there's still a pretty significant specific collection point at each root. So, J-Root had 312 ASNs that A-Root didn't see. So, if in the future you're going forward and you're thinking, this one root is just enough, obviously each root has its own catchment and it's going to have its own bias going forward.

So, what I wanted to do was to take a look over time at the cumulative new amount of data or a growing Venn diagram, as you would say, or maybe, over time at A and J. So, the figure on the right is looking at, from December 1st to December 31st, how many unique IPs were seen for .mail queries over that month cumulatively.

And you would expect that eventually, hopefully, this curve would just flatten out, right? But it doesn't exactly seem to flatten very quickly. As time continues, you're still continuing to see more and more and more sources.

And I think this is important because some of the most available data that we've used in the past is DITL data for analyzing root data and those are typically only in two days at a time. So, what I'm seeing here is that, this is over 31 days, I'm still starting to get new information. The entropy in our data is not decreasing at all. This is a lot of new information every day that you're gaining. And so if we can go forward to the next slide, please.

I wanted to better understand for each one of these new source IPs, how much traffic is it really sending us? And so, the graph on the left is looking at the cumulative distribution of traffic so how many queries did a particular IP address send over the course of the month?

And it turns out that roughly 55% to 60% of them are sending less than 10 queries for .mail domains over the entire month. So, using that, I was like, "Okay, well maybe, if we wanted to slice and dice the data a little bit differently to—if that's just pure noise or like Jim quoted earlier, Internet vagary—Steve, I'm sorry. I don't know if you had your hand up for a while. Please, jump in.

STEVE CROCKER: Thank you. Back on the previous slide, you mentioned catchment which is exactly where my mind was going when I focused on your Venn diagram. What do we know about the distribution of the instances of A- and J-Root against ASNs? It seems to me that one could ask the question—never mind, looking at any actual data, just looking at the map of the Internet—what could you predict about how many ASNs would likely go to a J-Root and how many ASNs would likely go to an A-Root and how many would likely go to either or both?

MATT LARSON: That's a really good question. And I think maybe some of the RSSAC002 data would be useful for that as well as just understanding the number of instances each root is actually broadcasting out.

STEVE CROCKER: Because I hadn't ever thought about this before but it now seems to me obvious unless I'm falling into a trap of some sort that each instance of a root server has a pretty natural alignment with nearby ASNs topologically.

And so, one could get a pretty good predictor of what this Venn diagram is likely to look like at least from a percentage point of view, if not the actual numbers just by studying the distribution of the instances and the topological map of ASNs. And if this data lines up with that, then there's nothing more here than what you would predict from the distribution of the root servers.

MATT LARSON: Correct. I agree with that. Not to put you on the spot here but I think ICANN OCTO is doing some research on looking at ASN placement for the IMRS letter. Is there anything that you'd like to chime in on that talking about this theory and if you have any thoughts on catching it in terms of announcements or affinities?

Well, we'll see if he comes back a little bit later. But yes, Steve, I can say anecdotally, back in the previous round, there were some name collision concerns around CBA and Commonwealth Bank of Australia originally had claimed that that traffic was all of theirs but in fact it was all coming out of Chiba, Japan.

And we were able to identify that because Verisign had so many site placements in that geographical area that we were probably more likely to have received that kind of string affinity for that at A-Root rather than opposed to other different roots.

STEVE CROCKER: Good. The thing that I always gravitate toward is how much can we explain why these are not just at the observational level of seeing how stable the data is or what the data is but get an explanatory theory underneath this that tells us why things are happening this way. Now, this is just one piece of that kind of inquiry, the catchment thing. Thank you.

MATT LARSON: You're welcome. And I think if we continue to get—I'll just touch on this real quickly. So, going back, this was just also using as a litmus test to maybe remove some of that vagary in the Internet looking at the CDF. I removed all the unique IP sources that were querying less than 10 times and I regenerated that cumulative map or graph on the right because I was hoping to see that curve flatten much more quickly.

But unfortunately, it seems to have slightly flattened but you still continue to see that growth over time that there are more and more IP addresses regardless of having some filtering criteria. If we can go to the next slide, please.

So, then this is instead of looking at where the traffic's coming from and seeing the catchment for particular ASNs, this is looking at the second-level domains to see if you're getting a different set of names at each different root.

Well, old J-Root clearly has very minimal traffic. You can probably ignore that whole purplish, blue diagram down there. It's pretty clear that A-

and J-Root have very minimal overlap, right? It's only 1500 and 30 second-level domains overlap between the two and the rest of it is unique per root. So, to me, again, this goes back to your catchment theory in terms of each root is going to have its own particular vantage point in the world and whoever is talking to that root at a particular time.

So, going to the graph on the right, this is again just looking at the cumulative number of unique SLDs for mail over time. And here you don't see the curve bend at all. These are just straight lines. You keep getting more and more and more unique, second-level domains under .mail as time continues. Yes, Steve?

STEVE CROCKER: Yeah. To your point about the lack of overlap between A and J, that's qualitatively different as you've pointed out from the previous Venn diagram and that suggests to me something quite different. I don't know exactly what it is but for one reason or another, particular SLDs are pointed to A and other particular SLDs are pointed to J and that does not seem likely to be just because of the catchment.

MATT LARSON: I would completely agree with you and I have one small theory I'd like to share on the next slide as to why that might be the case. So, the graph on the left is looking at the number of queries, a unique second-level domain received over the entire month of December.

97% of these second-level domains are only receiving one query. To me, that says that you're getting all of these random strings, something random .mail and you're never seeing it again. And my hunch to that is this is Chromium queries that might be going through a suffix search list processing where .mail is being attached to the random label being generated. And this is why you're seeing so many unique, non-overlapping domains going forward.

That's my hunch on this. That is also backed by some analysis that Duane Wessels and I did a few months ago when we looked at the Chromium queries, we actually could detect suffix appendages for queries that match the Chromium pattern.

We did not purposely look at .mail but that might be an item that we might want to go back and take a particular look at. And doing just as I did before, looking at most of these only received one queries, I redid the graph on the right looking at cumulative growth suggests saying—but this time saying that the second-level domain has to be queried at least two or more times.

And this is where you get to at least see the curves start to flatten, right? And they're not quite as straight as they were before. But you're still seeing very unique catchments at A and J and the growth pattern continues and continues to grow over time.

I believe that is the end of the deck that I have. The only other random tidbit I want to throw in here is in regarding the Chromium queries as well, was that in mid-November, the Chromium code base was actually

modified and they've changed their behavior to how and when they push out the random NXDOMAIN queries to the root.

And since that deployment in Chromium 87, the total root server system traffic volume has decreased by 40%. So, take that for what it is but maybe that would change if we look at .mail again here in the next few weeks. Yes, Jim, I see your hand. Please go ahead.

JAMES GALVIN: Thanks. I just want to observe, you probably can't see it because I think you're presenting too, aren't you, Matt? Anyway, there's a question from our attendees but it's in the Q&A pod. If you can open up the Q&A pod, you can see it or I can read it out here for you. Would you like me to read it?

MATT LARSON: Yeah. If you don't mind reading it. I'm trying to find it.

JAMES GALVIN: So, if we can go back to the ASN [then] slide which—how far back is that? I guess the question's kind of been there for a while because that's going back.

MATT LARSON: This should be the slide.

JAMES GALVIN: Yeah. Okay. The question is, if they're waiting on IP addresses that is Cloudflare or DNS Google IP addresses as one hit but so with data about a micro ISP but have higher downstream impact.

MATT LARSON: So, there was no weighting of IP done in these graphs. These are just pure—every IP is valued the same. This is just unique instances on the right graphs and then on the left those IPS are aggregated into their proper ASN for the overlap. Does that answer your question?

JAMES GALVIN: Yeah. We'll have to wait and see if Jothan types anything else here but so go ahead. Go on. He said, "Yes, that answers the question."

MATT LARSON: The only other question I'd like to put forward to the group was, did I miss anything? Are there certain graphs that you would like to see extended or dig into deeper details? And furthermore, if you thought this was a decent enough approach, then I will try and use this as an automation mechanism for generating these presentations for CORP, HOME, as well as the other strings that allow us to—maybe we have the first step of our case studies. So, I'll put that up to the group as a question. Yes, Steve Crocker, please go ahead.

STEVE CROCKER: Right. These are all extremely interesting, pretty, and informative at a certain point but the driving question it seems to me is getting

underneath these and seeing what's causing these? What are the explanations or explanatory theory underneath this, if you will?

So, I'm a little bit conflicted in that, it's very pretty to see all this and it's great work but it's not by itself sufficient for the big question which is what drives this work which is what kind of decisions should be made about whether or not these strings should or shouldn't be delegated.

And for that, one has to understand a great deal more than isn't answerable directly by these kinds of graphs. What causes these and if they were delegated, what kind of harm would happen and then how do you evaluate that harm?

So, that's the big picture from my point of view and I'd want to make sure that these exercises which take considerable amount of energy, I would imagine, are appropriate for—in proportion to the challenge and purpose of the whole exercise, whole inquiry that we're doing.

MATT LARSON: Thanks for that. Appreciate it, Steve. And yes, I agree. I think the million-dollar question is what is causing these queries, right? And that's where we're ultimately, hopefully getting to and by understanding what's causing these queries, then that'll help better inform us if there is actual harm or risk.

STEVE CROCKER: And a million dollars is cheap, actually.

MATT LARSON: Good point. Any other questions or comments?

WARREN KUMARI: So, yeah. I'm using the web UI so I can't raise my hand so I had just typed hand in the chat. But yeah, I mean, I unsurprisingly agree that what's actually important is knowing what is causing the queries and what the impact would be and that can also be scaled up to provide a more general answer.

But with that said, at the least this is a bunch of data and I think it's incredibly helpful and useful because we now have this data to look at. It doesn't answer the question but it gives a lot more flavor and insight than not having the data.

The original reason I put my hand up though was to just clarify for the Chrome probing queries, as you said, there were some changes but I believe that that was only in the Android version of Chrome currently. I believe that the larger set is going to be in the next 88 build of Chrome for the desktop as well.

STEVE CROCKER: I'd like to ask a somewhat off the wall question if there's time to do that. I don't know where we are in the agenda.

MATT LARSON: Please go ahead. This was the end of my agenda. I did not do anything else on .internal so I went deep down the rabbit hole on mail so I apologize. So, please go ahead, Steve.

STEVE CROCKER: I'm going to ask what would happen if—and I apologize if this is a well discussed thought that has come up before. What would happen if very purposefully, MAIL and CORP and HOME, for example, were delegated and the delegation resulted in no response whatsoever to any query to them.

WARREN KUMARI: What do you mean by no response?

STEVE CROCKER: Yeah, so that's kind of indistinct. Okay. So, I guess there's two versions of it. One would be that it's delegated in the usual way and the name servers that are pointed to then refuse to give any further response at all. So, that would be as if there was an entity in charge of mail that just simply was 100% unresponsive after that. And then a different form of that would be to jigger what the root servers do so that they give no response to the selected—that would be a change in the behavior—

WARREN KUMARI: I mean, that is largely what they do, right? They currently respond with NXDOMAIN and the DNS because there is—

STEVE CROCKER: And that's what I'm suggesting would be changed. That instead of getting an affirmative response that the domain doesn't exist, then we get a blackhole.

WARREN KUMARI: So, what would happen then is name servers—I mean, it depends on implementation but to some extent, name servers would probably become sad because some set of them would have a set of Qs which they are outstanding queries. Most of them would probably be okay but what would eventually happen is you would—

STEVE CROCKER: And I would recommend that if this were going to be done, that there'd be a notice sent to the world like a notice to airmen that these are being turned off and give a few months' notice to people that the behavior that they're seeing is going to change. And then see whether that this would quell all of these queries and you'd see a significant and permanent drop.

WARREN KUMARI: So, I mean, what would happen is DNSSEC would tell you that the domain exists but you're not able to resolve it so probably what you'd end up doing is basically something along the lines of roll over and die, right?

You would ask other name servers to please go off and answer this. And then eventually, depending on if you've got extended error implemented or not and if we have a separate extended error code for

this, you would just have a long delay and then I believe the final result that you'd end up with is at least from a recursive who's asking it, is so fail. Am I [inaudible] that?

STEVE CROCKER: Well, and the question is, would that have any positive impact on people saying, "Gee, that's really unpleasant. Why is my service so poor? Why am I having so much trouble?" And then when they ask somebody—

WARREN KUMARI: I think you'll just end up with a delay before you have the, why is my service so poor?

STEVE CROCKER: Right? And then, eventually, one would hope that the idea would get across that they should not be asking that question. It's a poke in the eye of a certain type.

MATT LARSON: Jeff Schmidt, I see your hand's up. Please go ahead.

JEFF SCHMIDT: Yeah. This, this is Jeff. Thanks. And, we did these experiments with obviously not top-level domains but if you believe that, for example, corp.com is a proxy for or a view into what happens at .corp, we

experimented with all different sorts of responses, both in a lab and a [Mart] Internet scale.

We reached out to folks that were talking to us that had issues etc., etc. And I would say mostly, the issue is entirely dependent on what underlying protocols are in play. So, for example, in the Microsoft world where a domain client is attempting to reach a domain controller, if the query fails either with MX or with serve fail which would be what happens if you just blackhole them eventually, they just move on through their search list anyway.

And all of these things are largely imperceivable to the end-user for behind-the-scenes protocols. Now for other protocols like SMTP, obviously HTTP, you would most likely get some sort of a timeout that the user would just give up.

WARREN KUMARI: Yeah. But, I mean, in general, the request of service is going to catch the fact that it tried this five times. We tried 13 different name servers and now has a serve fail and it will catch that for the negative TTL and so—I mean, you'd be adding a little bit of annoyance and pain to yourself but I don't think it solves anything.

STEVE CROCKER: Yeah. So, you don't think that it would have any positive educational impact?

WARREN KUMARI: A suggestion that was [inaudible] a long time back [inaudible] disposed off was the, you could point all of these at a webpage that explains why badness has just happened to you.

STEVE CROCKER: All right.

WARREN KUMARI: The infinitely long honeypot discussion of, you provide people a webpage and a mail bounce and an SSH banner and a Telnet banner and then everything else you can think of saying, your DNS is wildly broken, you should fix it because bad stuff could happen to you. And some set of people will break in an indeterminate manner and there'll probably be a huge number of lawsuits and bad privacy things will happen but you will at least have told people.

STEVE CROCKER: Well, lawsuits would be good in the sense that it would open up another communication channel.

WARREN KUMARI: That's true.

STEVE CROCKER: End of my curmudgeonly comments for today, at least on this call.

MATT LARSON: I would like to ask the group one more question, though. Let me rephrase this. For next week's call, would it be useful for me to redo this exercise for another string or a couple of different strings? That way we could do a compare apples to apples comparison and see if we find big differing results from CORP, HOME and MAIL or pick a different string entirely that we have on our list.

STEVE CROCKER: Sort of. Suppose you get the same kind of curves versus getting different kinds of curves, in both cases, the question I was asking before, it's sort of, what's the explanatory theory underneath? If you get different curves, then you have a signal perhaps that whatever the explanation is for one of them, it won't necessarily port over to the other.

And that's only a one-way test because you could get the same curves for different reasons but if you get different kinds of curves, then you have a good indicator that there are different things going on underneath.

MATT LARSON: Absolutely correct. And so I can speak to having looked at different curves and done outreach that those curves matter in terms of being able to do [remediation]. So, I'm suggesting that maybe we continue to look at a few more of these and look at those and just get a better sense of that. There isn't always one underlying cause and take a look at various different thoughts on this.

JEFF SCHMIDT: Hey. Sorry to butt in. Matt asked me to join today and be able to chat and I appreciate that. If the group thinks it's helpful, I would be happy to dig up and do a little bit of transmit mode on the work that we did for our second report.

In particular, we had about 300 pages of appendices that included what we called at the time a horizontal study and a vertical string by string study. I think many people have heard me reference before we did things like regular expression fitting to the strings because we were trying to get at what Steve said here, right? We've got to figure out the underlying why.

And so I'd be happy to do a little bit of transmit mode if people think that's helpful. If everybody's already looked at our study and it's not helpful, then that's fine too but I was wanting to volunteer it. We also had the unfortunate obligation that "decide" that CORP, HOME and MAIL were bad and I'd be happy to walk through what decision factors we used way back when to come to that decision, love us or hate us.

MATT LARSON: Thank you for that offer, Jeff. Just in the spirit of time, I know it's 3:00. I will circle back with you and the group but I think we will probably do a combination of both of those going forward over the next several weeks. I apologize. I have a hard stop here for another meeting so if there's any other business, please make it brief.

STEVE CROCKER: Thank you. Good meeting. Got to run.

MATT LARSON: Thanks. Thank you everyone and appreciate it. Let's talk next week.

JEFF SCHMIDT: Thank you.

MATT LARSON: Bye-bye.

[END OF TRANSCRIPTION]