
OZAN SAHIN: Good morning, good afternoon, and good evening. This is the RSS Metrics Work Party call held on the 11th of July 2019 at 15:00 UTC. On the call today, we have Duane Wessels, [inaudible], Abdulkarim Oloyede, Daniel Migault, Dessalegn Yehiala, Fred Baker, Jeff Oborn, Karl Reuss, Kazunori Fujiwara, Naveed Rais, Ray Bellis, Russ Mundy, Shinta Sato, and Tom Minglin.

From ICANN staff, we have myself, Ozan Sahin. I would like to remind you all to please state your names before speaking for transcription purposes and mute your microphones when not speaking. Thank you, and over to you, Duane.

FRED BAKER: Paul Hoffman dropped off for a moment. He's coming back. He's back.

OZAN SAHIN: Yeah. Thanks, Fred.

DUANE WESSELS: Okay. Thank you, Ozan. So, the agenda that I have for today is, first, I think we will go over what happened at the last work party meeting which took place in Marrakech. I sent an email out to the caucus list but we'll briefly go through that. I know Paul Hoffman has some data to share regarding some experiments that he's doing. I also have some real data that I can talk about a little bit. Then I thought we can spend the rest of the time going through comments in the Google Doc and see if

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.

we can whittle those down a little bit. Does anyone else have something that they particularly would like to talk about today other than those things? I should probably open the chat. Okay.

So, I'll go into what happened in Marrakech, I guess. So, we had two meeting slots on the schedule in Marrakech and then we also had a little bit of extra time which we took advantage of. But in the extra time, we didn't really ... Since it wasn't a scheduled part of the caucus meeting, we didn't do a lot of work there.

In that extra time, we talked. I presented some slides and graphs on standard deviation versus percentiles and why I think that we should use percentiles going forward in this document. Everyone in the room seemed to agree with that conclusion. And I attached those PDF files to the email that got sent out, so hopefully you all had a chance to see that.

The other thing we talked about was we spent a little bit of time in the BPQ section of our document and what's there now is just a little bit more detail about how you can take real [packet captures], tally everything up, and then determine the relationship between raw [packet] counts and queries over EDP and TCP. Then, the idea is to sort of apply those in reverse using RSSAC 002 data as input and that would allow you to estimate bandwidth and [packets]. So, I think that's pretty straightforward.

The thing that we got hung up a little bit about in Marrakech was the genesis of the BPQ was it was about service capacity more than the current load on the system. Also, the recollection that a lot of people

had was that when we were talking about BPQ, it sort of applied to the whole system and not to individual operators.

So, even though both of those things are true, my feeling is that if we are going to say anything at all [inaudible] we're going to be limited to talking about current load and using data from individual operators. Anyone who maybe wasn't in the Marrakech meeting want to comment on any of those two things? Doesn't sound like it.

PAUL HOFFMAN: Actually, Duane, I raised my hand.

DUANE WESSELS: Oh. I'm sorry, Paul. I didn't see the hand raised.

PAUL HOFFMAN: So, two things. We all got interrupted with some weird audio just as you were summarizing the BPQ stuff. But the other is—

DUANE WESSELS: I think that's [inaudible] phone does that thing. Okay.

PAUL HOFFMAN: Okay. But the other was that I think what you were saying to summarize was that because we couldn't – there isn't a good way to measure the capacities which is what RSSAC 037 wanted, we could use the RSSAC 002 numbers to get current levels. Is that what you were saying?

DUANE WESSELS: Right. That's basically it. I think that's where we're going to be – I don't want to say stuck. But I think the best we can do given what we have at this time.

PAUL HOFFMAN: So, then my question is does RSSAC or the caucus need to go back and revise 037? Because 037 was really clearly about capacity, not just in saying we want capacity but in the justification for wanting that. How will we handle the fact that the BPQs that the work party might end up helping them specify doesn't match what's in 037?

DUANE WESSELS: Well, I don't think that ... For myself, and I think for other RSSAC members, I don't think we need to go back and modify 037 on this point. The BPQ that is spoken about in 037 was intended as a way for the ICANN Organization to estimate costs of running the service. They can still do that if they can come up with their own ways of estimating BPQ as capacity.

The other thing that was sort of talked about in the room in Marrakech and even before that was that one thing, one approach you could take is to use what's described in our document to calculate the capacity of a system and then extrapolate that out to – I'm sorry, not on the capacity, the load of the system – and then extrapolate that out to capacity using [inaudible] is not defined in our document, admittedly. Does that make sense? Russ, your hand is up, I see.

RUSS MUNDY: Yeah. Thanks. I think that since there's a lot of people that will be saying things about 037, it will need to be looked at and examined with respect to all those inputs that come from the community review process and so forth. It seems to me that the right path for the work party to take, rather than trying to do a revision or suggest a specific revision to 037, to note where we see that there may be a different output from the work party than what the work party thinks was anticipated from 037. Raise a flag and say this needs to be looked at in the future in some manner, since it didn't align with original expectations as we read 037.

DUANE WESSELS: Yeah. Taking that a step further, our RSSAC chairs have even offered that if this is – I don't want to say too hard. But if this is something that we can't do, then the work party can change its chartering and say we don't have anything to say about BPQ. We can choose to not talk about it at all if we want.

RUSS MUNDY: Right. And if I also recall, they alternatively said we can say it's too hard for us or we can remain silent.

DUANE WESSELS: Yeah. Any other comments or questions about that part before I move on? Okay. So, back to the Marrakech meeting. One thing that we spent a lot of time on was talking about things like location of probes and almost number of probes, too. We sort of put forth a proposal – and we

talked about this last time on [our own call]. We could consider that there are two types of probes, near probes and far probes, and that certain metrics are much better done from near probes because it eliminates certain uncertainties in the network that are beyond the operator's control. So, we did spend a lot of time talking about this and going back and forth, but there wasn't really any consensus on where probes should be, if there really should be two types and how many there should be.

So, in my mind, this is still something that the work party needs to tackle. We need probably some more specific proposals for, when we say near, what do we actually mean by near, how near, and so on? I think both Paul and I later on will show some data that starts to have this discussion, starts to address things like number of probes and maybe even location of probes.

I apologize. I know that we have a suggestion to use the word vantage point instead of probes but I haven't made that transition yet, so I keep saying probes when maybe I should say vantage point. Ray, go ahead.

RAY BELLIS:

Hi, Duane. Ray Bellis here, ISC. In the discussion of vantage points or probes, was there any discussion as to how well connected those should be? One of our concerns was that those probes shouldn't generally be ones that are, for example, at the end of, say, a consumer broadband line with unknown and unreliable latency.

DUANE WESSELS: So, yes and no, I would say. Yes in the sense that I think one of the things that we sort of almost agreed on was that for things that are – for metrics that are tied to SLAs, you do want to use the nearer probes. You want to use well-connected probes, so you eliminate some of those uncertainties that you talk about. But, at the same time, there were also a number of people that would like to see measurements that sort of do their best to convey the end user experience. So, in that case, you kind of do want these probes from farther away. But again I wouldn't say we have consensus on that.

RAY BELLIS: Yes, [inaudible] which sort of end user you're talking about because most end users are not running their own resolvers [inaudible] root system. Most end users are talking to an [off net] recursive resolver [inaudible] talking to the root.

DUANE WESSELS: Yes. That's a very good point. Yes.

RAY BELLIS: That may change, of course, but ...

DUANE WESSELS: The other point to make on this, I guess, is that for latency, it's especially tricky because I think the nearer that you are to the thing that's being measured, obviously the latency is lower and better and probably all the operators become essentially equivalent the closer that

you get. The latency is sort of – the limit goes to zero. So, to the extent that you want to use latency as a way to – in an SLA – you may want to be a little bit further away.

So, I don't know. Like I said, there's still work for the work party to do. I don't think we're ready to tackle the whole thing in this call. I think we have other things to talk about. So, hopefully, we can have the discussion on the mailing list, some proposals, and maybe we'll have some ideas after looking at some real data here in a little bit. Jeff, go ahead.

JEFF OSBORN:

This is kind of obliquely raising an interesting point. For those who are in the caucus and not in RSSAC, I've been kind of tasked with putting a financial model together that involves the SLAs and the RSOs and ICANN. I'm sort of sitting on the side going the metrics process is going to end up with a set of results that make it a little easier to talk about money changing hands because service-level agreements are based on metrics and measurements.

But the very interesting point is coming up now of if you're in an SLA situation, you always want to be able to meet your SLA. If you're in a research environment measuring things, you're looking for different levels of accuracy. I just wonder whether we don't need to state as part of this group this is for research reliability measurement points only or to recognize that we're doing the draft work of what is going to become contractual. Is that logical? It made sense in my head. I want to make sure it made sense in English.

DUANE WESSELS: Yeah, Jeff. I think you're right. We're sort of starting to realize this. There is a little bit of text that talks about this. But I think it's a question about whether this work party should be trying to do both or should it limit itself to one or the other. Maybe there should be a separate work party for the other one. I got a bunch of hands up. Warren, were you next?

WARREN KUMARI: I believe so, yes. I started asking these questions in the chat instead because I wasn't at the Marrakech meeting. I was at another conflicting meeting. So, I wasn't sure how we've gotten here and if this is all settled. Apparently, not quite as much as I had assumed.

One of the things that I think is worth keeping in mind is depending on how the measurement system is set up and the number of probes, etc., gives a very different set of outcomes to things like what Jeff was talking about. If we're using something, we're building our own set of monitoring infrastructure and we have a limited number of probes on the order of ten, if I am trying to meet an SLA, I will make sure that I put my [serving] nodes right, right, right near the monitoring service or right near the probes.

If, instead, we're using something with a much larger set of probes, [ATLAS] or something [inaudible] something similar where there's a significant number of distributed probes, I am not really going to be able to optimize my service thing for making my latency look good or for meeting my SLAs.

So, I think that a lot of these discussions that we're having about what exactly gets measured, and also things like Jeff just brought up about SLAS, is quite largely influenced by the number of probes that are going to be deployed or are going to be used. And apologies again if this was covered during the meeting. As I say, sorry I wasn't there.

DUANE WESSELS:

So, we did talk about it, Warren, but it is not settled as you say. Russ?

RUSS MUNDY:

Yeah. It seems to me that we can make our best progress here by thinking of this rather than in terms of SLA or contractual agreements because if what we produce out of this work party is viewed as something that's more in alignment with the service-level expectations, the documents that have been published by RSSAC and IETF and so forth. If, eventually, as a result of this someone in whether it's called a PDF or something different decides to make use of these and writes contracts with them where it becomes an SLA, then that's a completely separate and independent and subsequent action.

So, it seems that the best way to progress is to have us think about it in terms of an SLE and existing published documentation and how we're going to measure against things of that nature. Thanks.

DUANE WESSELS:

Alright. Thanks, Russ. Warren, again.

WARREN KUMARI: I either violently agree with Russ or I violently disagree. I'm not quite sure exactly what his position was. But if this is going to be used as an SLE, then yeah, okay. If it's going to be used as an SLA, I have a really strong incentive to figure out where the probes are and I will put a [serving] node in the same data center on the same circuit as close as humanly possible to the set of probes, because otherwise, I get dinged for it. I couldn't figure out if Russ was saying that we should plow ahead and figure out what the metrics are, and then if it turns into an SLA later, eh. Or if he was saying what it ends up being has a strong impact on what we should be measuring and how.

RUSS MUNDY: What I was trying to say – maybe not very well at all – was that we should approach this from an SLE perspective and think about it in with respect to the existing already established publications, not think of it as something that will be an SLA. If someone else chooses to make that decision in the future, fine for them. But that's not what we're focused on. [inaudible] should be looking at SLE kinds of things.

DUANE WESSELS: Alright. Thanks, guys. Again, I still think we need to take this to the list or a future meeting and hash this out some more. We're going to move on to some other topics.

Back to the Marrakech meeting, we have John Kristoff as an ICANN research fellow and he has written some code and written some data analysis on the RIPE Atlas measurements. We specifically asked him to

look at the RIPE Atlas measurements and see to what extent these metrics can be applied to those measurements.

So, John wasn't at the meeting but I presented his results. Those are also attached to that email. We gave John a little bit of feedback on maybe how to improve the presentation of the data and whatnot.

As I said in the email, when John did this, he looked at the RIPE Atlas, the anchors only, which I think there's only 500 of those. He produced these latency graphs. If you haven't looked at this, I would certainly encourage you to because it's very interesting how different the different root servers show up in this data. There's a very wide range of measured latencies.

So, we've also asked John, like I said, to improve those graphs and then also to look at some of the other metrics and see if they can be applied to the [data] as well.

UNIDENTIFIED MALE:

Note that RIPE is heavily biased towards where they're located. Europe is very well covered and America is sort of second. So, when you're looking at things like latency and stuff like that, you're very biased toward where stuff is coming from.

DUANE WESSELS:

Yeah. I think that's a very good point. One thing we could ask John to do is to maybe try to hand pick or somehow pick vantage points that try to balance that out a little bit.

UNIDENTIFIED MALE: You could apply some sort of waiting metrics I think to the measurements. That's true.

DUANE WESSELS: Yeah.

FRED BAKER: Well, I was looking at that a bit and would up deciding that, at least for the stuff that I was doing, I would have to report what each probe saw individually. I didn't see an obvious way to aggregate probes. Maybe somebody can clue me in on that.

DUANE WESSELS: Yeah. Fred, I guess the kind of thing that I was suggesting was that you could, on a very broad level, you could say, "Well, I'm going to pick ten probes from Europe, ten from the US, and ten from each continent," or whatever and do something like that. We'll ask John if he can do that for us.

Okay. One last thing we talked about in Marrakech was – this was sort of born out of some sort of individual side discussions between myself and Russ. In the document, there were a number of alternative measurement techniques that used, or proposed using, a locally operated recursive name server on the vantage point. We kind of ended up talking ourselves out of this for a number of reasons. One reason being that we felt like you kind of end up measuring the recursive

software's behavior maybe rather than the operating system. So, if one implementation uses a certain [time-add] value, for example, that might show up very strongly in your measurements.

Of course, there's also this issue that if you use this you have to pick which software you're going to use and you have to sort of hope that software doesn't drastically change its algorithms over time and so on.

Anyway, after discussing this, we all sort of agreed that these alternative approaches can be stricken from the document and if you happen to look at the current version in Google Docs, you'll see that those sections are now struck out and if everyone on the call here is on board with this, then we'll take a step of action in removing them and that will simplify our document somewhat. Any comments about that or I guess anything else that we talked about at Marrakech before we move on? Paul, go ahead.

PAUL HOFFMAN:

One thing that was discussed at Marrakech a little bit – unfortunately, Brad is not on the call right now because Brad was in and out of the room. But at one point towards the end, when there was the discussion of the Atlas probes he said, “I just want to measure from ten well-connected data centers.” He said it a couple of times. A couple of people in the room seemed to agree with him. That's a very different alternative. The only reason I'm bringing it up now is when I talk a little bit later about the quick and dirty stuff I'm doing, it was based on that view.

It's fine if the work party doesn't go that way but there were a couple of people – or, I mean one person vocally in favor and a couple of people going, “Yeah, yeah, that sounds right,” as another way. And this would be for what we would call the far probes only.

DUANE WESSELS:

Yeah. Thanks, Paul. That's right. That was Brad's suggestion for how you could [inaudible] probes. If you could rank data centers or data center locations, you could pick the top ten or top whatever and put probes there or vantage points there. So, that's a good lead-in to the data you wanted to share. Does anyone have last-minute questions before we let Paul show what he wants to show? Are you going to be able to share, Paul, or do you want me to do something for you? Hey, it works!

PAUL HOFFMAN:

And I'm unmuted. Even better. Although I can't see what I'm sharing. Okay, life is that way. Anyhow ...

Based on what Brad said, I realize that one of the services that we use in OCTO is DigitalOcean and DigitalOcean has 12 different locations where you can set up your VMs. It turns out even within a particular city – like, there are three in New York City – there are actually all ... And I verified this. They're all in separate physical location data centers with different connections.

So, very quick and dirty, like yesterday, I put up a probe service. What you are seeing here is the result of just the numbers – less than 24 hours numbers – from one of the probes. They're actually running at all

12 now. But this is just the latency plus timeouts from NYC-1, which is one of them, using the various numbers.

I'm not at all suggesting this is the way we should go or anything like that. It's mostly just to say that if the work party goes towards the more limited view of – or the view of we don't have to have a bunch or if we can't pick where the [Atlas 1s] are, it is definitely possible to do this kind of thing ... Like I said, I spent less than a day on this. So, once the work party chooses a way to do these, we should do much better. That's really all I wanted to say is, look, there are some pretty numbers.

DUANE WESSELS:

Alright. Thanks, Paul. I see in the chat Warren is sort of reiterating his point about gaming the measurements. We did talk about this in Marrakech also. Some people would say, okay, if you game the system by putting root servers at these locations, then that's great. This is maybe where root servers should be. You probably improve the overall service.

I think the point at which it maybe becomes a problem is if the system is gamed in a way that only the measurements are impacted and the actual service the users get is sort of unchanged. I think, Wes, you mentioned I could put a raspberry pie right next to it and configure it in a way that it only receives queries from [probes] and things like that. I agree that that's a problem.

So, I guess my question to the group is—

UNIDENTIFIED MALE:

Actually, Duane, before you go in, let me sort of flip that over, though. I think one thing that those discussions all – and again, I’m not advocating towards the let’s just do ten where people know where they are. But these are all Anycast clouds and part of the discussion came from Jeff Osborn in Marrakech saying what if there was a root server operator who primarily wanted, because of the way they view the way they want to serve the world, if they primarily wanted to put most of their probes in obscure, hard-to-reach places? I’m sorry, most of their instances in obscure, hard-to-reach places so that they are serving underserved places better.

One of the things – again, I certainly don’t want to put words in Brad’s mouth because he knows how to find me. But I think one of the things Brad’s responses was, yes, but this is Anycast, so that as long as you have some instances that are more easily reachable by large data centers or whatever, that’s where the probes are going to hit.

Again, these numbers are showing that, in fact, from New York there are different things to different places. But if the model that the work party wants to go for in measurements is to say we want to be able to measure the most hittable systems, which is exactly what Anycast does, that’s different than saying you need to put a probe at every data center, because in fact, putting a probe at two or three places near two or three of – I’m sorry, putting an instance at two or three data centers near, not even in it, will probably get through the threshold that we want.

I know this is making the discussion harder but Anycast makes the whole discussion different about you don’t need to put instances

everywhere. You only need to put instances close to things that will generally get something. Does that make sense?

DUANE WESSELS: Yeah.

UNIDENTIFIED MALE: Okay, so I guess I don't need to show pretty numbers anymore, so I will stop sharing there.

DUANE WESSELS: Okay. I have two thoughts about this discussion about gaming system, I guess. One is that certainly RIPE Atlas has been doing these measurements for years and if you wanted to go and find – even I think on the DNSMON site, you can get the latency measurement. So, the latency to root server operators has been measured for years and, as far as you know, no one has felt the need to game the system yet, right? So, it seems to me we're talking about maybe these measurements having more visibility that would cause people to want to game the system. Warren?

WARREN KUMARI: I think I disagree with you that nobody has tried to game the system. I think anyone who's using the RIPE Atlas measurements is gaming the system whether they mean to or not. It's one of the standard sets of metrics that people look at when they are trying to figure out, "Is my letter doing okay?" I know that I've spent a significant amount of time

looking at DNSMON and being like, “Wow, that doesn’t look good. I should fix it.” And then a new node gets deployed and it looks better. I don’t think people are intentionally going, “How do I game the measurement from RIPE Atlas nodes.” But it is a data set which people use and so they automatically try and make it better. So, maybe gaming is not exactly the right word. But anytime you provide a set of metrics which exists and which people can use to determine how well their system is going, they will take advantage of them. [inaudible] nodes that I think it gives a much better representative view than ten big data centers.

Related to that, I should also say the top ten data centers in the world are very, very, very different to the top ten set of eyeball networks in the world, and I think that measuring from the top ten data centers gives you a completely synthetic set of measurements which don’t actually reflect anything at all other than the top ten set of data centers reaching root servers. So, I think that some of this is kind of related to the fundamental are we trying to actually measure if the root server system is getting better for the users or are we measuring stuff so that we have metrics? I think that that’s still somewhat of a fundamental disconnect on are we collecting stuff so we’re collecting stuff and can point at or are we collecting stuff to understand what this actually looks like for the world?

WES HARDAKER:

So, way back when we were originally starting this document and starting to think about it, the original thinking or the original numbers of people were posting for latency issues were like 500 milliseconds,

something that anybody could hit easily. With the goal of not being able to say, “Are you better or worse,” but is your service sufficient that you are reachable? I think we really wandered away where we’ve come down to – it came down 200 milliseconds at one point and then 50 milliseconds and it’s like the target for latency has become more important and there’s an awful lot of people that say latency to the root is not actually that critical in the first place.

So, I’m sort of beginning to wonder, should we re-question the goal of what is it that this metric is trying to do for us goal-wise? Are we trying to prove that the system is really good or are we trying to prove that the system is functioning? And there’s two different end goals to this.

Then, Warren, you’re 100% right about where you measure stuff. Eyeballs does not equal the rest of the DNS. In fact, if you look at the top number of TLD requests or the top number of requests that come into the roots compared to the Alexa top 100, they don’t align. There’s a ton of stuff that comes into the roots that isn’t anywhere on the Alexa list. Like, dot-ARPA, for example, is a huge quantity of traffic that hits the roots and that [inaudible] to dot-ARPA [is a] webpage.

DUANE WESSELS:

Thank you, Wes. Jeff, I know you’ve had your hand up for a little bit. Go ahead.

WES HARDAKER:

Somebody has a lot of background noise, by the way.

DUANE WESSELS:

Yeah. Somebody is doing dishes, it sounds like that.

JEFF OSBORN:

Clacking plates and children are there. I'm empathetic but it's hard to hear. I'm wondering at what point we don't say: what are we trying to accomplish here? I'm coming from a [fun] direction, Duane, because I'm sort of sitting on this going, what can I use that the metrics work group will come up with to then go and make an SLA? So, if that is not the purpose of this group, then somebody needs to help me either ask Brad and Fred or somebody and say a work group that does address the work of the SLAs is necessary. And if this isn't that, if rather this is just a sort of self-improvement exercise, then we go a slightly different direction.

The problem with gaming comes up because it used to be the root server operators had one motivation which was to do the right thing and they could figure out how. And very diversely they did. But we're talking about now having a paid SLA in which you get money if you do something. That's when gaming comes into the system and that's why going forward would be different if we're connecting this to the SLAs. So, I kind of feel like we need to get to the root of the matter.

The other thing that's odd is, if this was a commercial SLA, we'd simply go to the customer and say, "Where would you like the probes? What would you like them to measure?" So it's kind of odd that we're the ones coming up with this.

DUANE WESSELS: Thanks, Jeff. Paul, I know your hand us up, but I want to ask Jeff one quick follow-up question. What you just said, did I hear that you almost suggest that maybe we should focus only on the service-level parts of this and not so much on the research, what we call the research aspects of it, the bettering the system parts?

JEFF OSBORN: Well, I'm just wondering if we don't need to clarify to ourselves what we're trying to accomplish, because I think at least implicitly I think ... I was getting the direction, "Go follow the metrics group, and from that you should be able to fall out something you can make into an SLA."

DUANE WESSELS: Okay.

JEFF OSBORN: So, if we are not going to go that way, that's fine. I just need to go off and see if I can assemble something that does go that way.

DUANE WESSELS: Okay, thanks. Paul, go ahead.

PAUL HOFFMAN: So, that's really interesting, Jeff, because I had not heard that we were already – somebody was already at the point of being tasked with trying to figure out the SLA. But really I want to give a big plus-one to what Wes said. I, too, came in ... Again, since I'm not on RSSAC – I've only

been seeing what's been happening here in the caucus, in the work party. I, too, came in with that different goal from the beginning of we're not trying to fail people, we're trying to measure whether the system is good and whether the individual parts of the system were good.

So, I think he's right – very much right – that we need to figure out the goals again because they really sound like they've been shifting. But the other thing is that one goal that we didn't have at the beginning, which I've now heard a couple of people put in on, is what is the user experience? I do not believe – I have actually done research in the last couple of weeks and I do not believe we have any idea on what the user experience is of the root server system, given that the large majority of users apparently are behind resolvers that cache. And the fact that they cache means that all their positive responses for things from the root last for approximately two days. Even the ones that cache with a hard maximum of less than two days, such as one day or God forbid even one hour. We don't know what it means for the user experience to then be the one person from a large ISP who is waiting the however many milliseconds for the answer on that query.

And anyone who [inaudible] server data knows that the vast majority of queries are for things for NX domains. I could not find any research at all that explained which applications were doing all of those queries and were they waiting for the answers, and if so how did that come to the end user?

So, again, going back to the goals, if we're going to put user experience in, I think we're going to be years away from understanding the user

experience of the root server system, much less of individual root server operators. Thanks.

DUANE WESSELS: Alright. Thank you, Paul. Ray?

RAY BELLIS: Yes, hi. I just wanted to point out [inaudible] the discussion we had at the RSSAC caucus meeting at the last IETF relating to the service [contract] work party because that's somewhat stalled. I think [inaudible] say so at the time was, ultimately, it was overlapping significantly with this group. I've got that service coverage [inaudible] in front of me know. It has things like what indicators, performance factors, and coverage to find adequate and/or inadequate service by the RSS? There's a lot of overlap, though. So, there is [inaudible] service party that will [inaudible] need anything or the output from this group if [inaudible] to continue.

DUANE WESSELS: Yep. Alright. Thanks, Ray. Warren?

WARREN KUMARI: Yeah. I think I violently agree with Jeff that the purpose of this, goal of this, seems to have been understood differently by different people and, as far as I can see, also changed over time.

But I think that if we do end up with this largely being used for just SLE, SLA – is person doing okay? Then I think sitting as an example, latency at a number, and then the only thing that gets reported is how much of the time you actually meet or do not meet the latency number would be a reasonable thing. As soon as it ends up being this is the latency metric and the latency actually observed is ... Then I think you start ending up in people gaming this for artificial reasons.

So, if, for example, Wes's thing of the latency number is set at 500 and no matter where you are, you should easily be able to meet 500 and you only report pass/fail, then you can easily [assure you're] within the SLA. If you also show present/past and their average latency was 26, then there's a strong incentive for people to fiddle with things so that their latency is now 25, and then 24, and then 23 and I think that ends up with people working towards the metric instead of working towards the desired outcome.

DUANE WESSELS:

Alright. Thank you. I like that. I think Fred was before Ray, maybe.

FRED BAKER:

Hi, there. Thinking about user experience, if we're going to talk about the user experience of the roots, then it's going to be the experience of the users of the root which are typically not random Internet consumers. They are resolvers, as Paul pointed out. I'd be really surprised to hear us trying to talk about individual users as opposed to resolvers in this case.

DUANE WESSELS: Alright.

FRED BAKER: I hear Paul every so often saying, “I’m not a member of the RSSAC, so I don’t have an inside track on all of the information going on.” We tried to – we, you, and I, and the RSSAC – tried to give the work party direction on what was going on. If the work party has questions that it would like the RSSAC to answer, it might be good for the work party to formulate them and you can send them or I can send them to the RSSAC.

DUANE WESSELS: Thank you. Daniel? You’re muted, Daniel. I don’t know if you’re speaking. Ray, I’m not sure if your hand is a new hand or an old hand.

RAY BELLIS: Sorry, an old one.

DUANE WESSELS: Thanks. Warren, did you want to say something again?

WARREN KUMARI: Sorry, no. I thought I put my hand down.

DUANE WESSELS: Okay. Daniel, I see your hand is up but you're muted. I'm not sure what's going on.

JEFF OSBORN: In comments, he's saying he can't find the mic.

DANIEL MIGAULT: Okay, I found it. Got it. So, the problem I have is that if the metrics and the measurements are defined for a very specific platform, it looks very much that we are ending with something similar to [inaudible] providing its measurements with a specific software so that it can match its own measurements, the measurements he claims and so on.

Maybe the metrics should be defined independently of both a platform and SLA and what we end up is that the platform we're going to choose for those metrics is going to depend on which point of view we are really measuring and what the values are necessary to be met. It depends on the platform as well.

So, I think we should split up the metrics, the platform, and the SLA. I'm trying we should split the documents, but we can end up into some [inaudible] directions about what the metrics imply and means, given a few examples of platforms and maybe associate some various numbers to those, different platforms. So, that may be a way to move away this discussion, which platform we consider and so on and so on.

Also, recall that the reason we introduced far away nodes and closed nodes is mostly because we ended up in a very high dependency on the

network between the platforms, the probes, and the root server system.

So, we tried to get closer to remove the impact of the network, but fundamentally, some of these measurements can only be self-assessed if we want to remove this aspect of the network and really be focused on the root server system itself.

I think whatever measurements we came to, if it's not self-assessed for some of the metrics, might only be one way to sort of control and correlate self-assessed measurements by the RSOs. I think it's worth what it's worth, but it's still [inaudible] third party to check what is being claimed and audit, what is being provided by an RSO for example. So, I think it's still valuable.

DUANE WESSELS:

Thanks, Daniel. I wanted to call up a little bit about what you're talking about, separating the platform from measurements. It is definitely ... In my mind, what we produce should be platform agnostic as much as possible. We shouldn't specify any particular platform. But I do think we need to consider a few things, such as number of vantage points and something about location of vantage points because I think we all understand that some of these ... The measurements are going to depend on those characteristics. They're going to depend on how many locations you measure from and maybe nearness versus fairness.

We should make it as separate as possible, but we do have to still, I think, make some recommendations. For example, there should be a minimum of [end] locations and they should be distributed across

geographies and things like that. Does that still work with what you were thinking about?

DANIEL MIGAULT: Yeah. We may have different considerations for ... Maybe we came with some generic considerations that apply for any kind of platform, but I don't think it's bad that if we mention that if we have a widely spread number of [inaudible] this is what we're going to measure and it's going to be heavily influenced by the network latency and so on that is not dependent on the root server system, while if we take another platform where we have, let's suppose, a given number of data centers, this is what we're going to measure. The platform is going to ... Some of the characteristics of the platform can be identified, and given those characteristics, we can say this is how we will have to interpret what the metrics shows.

DUANE WESSELS: Okay, thank you.

DANIEL MIGAULT: But I think it's more or less ... I mean, I think it's not the opposite of what you were saying.

DUANE WESSELS: Yeah. Okay. We've only got a few minutes left. I want to mention that the next work party meeting has been scheduled to take place during IETF in two weeks. Let's see if we can find the exact date. It's a Monday.

Monday the 22nd. So, if anyone will be in Montreal for the IETF, I'll have a meeting room that we can go to and work from there. In between now and then, I think that the plan is that myself and Russ and perhaps Fred and Brad will meet with ICANN staff to go over this call and start making some of the changes that we talked about or some of the changes that we've already agreed on. For example, deleting some of the sections. We'll resolve some of the comments in the document.

One thing we didn't get to talk about today, but I'll just mention briefly, there was a suggestion from John Heideman to move some of the examples to a separate section and I like that idea, so I'll propose that we do that in the document as well. Then we'll take the ideas here and start some discussions on the mailing list. Russ, anything else to add?

RUSS MUNDY:

I think the only thing that I wanted to add is, hopefully, we'll be able to arrange for remote participation for the work party meeting during the IETF. I don't know that – I think Ozan's original email said we would have a Zoom room. Ozan, is that correct?

OZAN SAHIN:

That is correct, Russ. Thanks for raising it.

PAUL HOFFMAN:

Assuming that all things work, I should be getting the two microphones that we have used at other meetings in the mail from whoever had them at Marrakech in the next day or two. So, I'll be hand carrying them in to Montreal, so we'll be able to do that. For other people, since

Duane didn't say when the meeting was, it's at 8:00 in the morning local Montreal time and we don't actually have any meetings until 10:00 that morning so we can have whatever time folks want for that meeting.

DUANE WESSELS: Yeah. We can make it a little bit longer than usual. That's a good point.

RUSS MUNDY: Excellent. Thank you, Paul.

DUANE WESSELS: Alright. We're basically at the end of our time. Any last-minute comments before we close this Zoom call? If so, speak up. Alright. Thank you, everyone, for your participation. Look for more communication on the list, please.

RUSS MUNDY: Thanks, all. Bye now.

[END OF TRANSCRIPTION]