

Ozan Sahin: Good morning, good afternoon, and good evening. This is the RSSAC Caucus RSS Metrics Work Party Teleconference held on the 5th of September 2019 at 17:00 UTC. On the call today we have Duane Wessels, Ryan Stephenson, Abdulkarim Oloyede, Warren Kumari, Jaap Akkerhuis, Jeff Osborn, Karl Reuss, Ken Renard, Paul Hoffman, Ray Bellis, Tom Miglin, and Jack Biesiadecki. From Support Staff, we have Steve Sheng, Andrew McConachie, and myself, Ozan Sahin.

I would like to remind you all to please state your names before speaking for transcription purposes and I thank you. Over to you, Duane.

Duane Wessels: Alright. Thank you Ozan. Welcome everybody. Russ is not able to make it today, so you'll be stuck with just me. And hopefully everyone saw the message that I sent out late last night. I have sort of four things that we plan to talk about today. Those would be the vantage points, the correctness metrics, staleness/freshness metrics, and then, assuming we still have some time, I'd like to sort of start some of the discussion about thresholds.

Paul Hoffman: Duane, this is Paul. I didn't see such a message. I don't know if everyone else did.

Duane Wessels:

Okay. I hopefully sent it out to the RSSAC Caucus List. So, did anyone else? Nobody saw it? Okay. Well... [inaudible]. Then I probably misaddressed it. My apologies. So, let me go into a little bit more detail about what I wanted to talk about.

So, for vantage points, in our document this is in Section 3. Previously there was just some placeholder text saying that the Work Party needs to come up with some Recommendations here, and I went ahead and put in some starting text that we can hopefully talk about or find problems with or get behind or whatever. Yeah, it's up on the screen there now.

The second point was the correctness metrics. So, for a long time we've talked about doing correctness based on DNSSEC and more recently there was some discussion of proposals to also have what we're calling a matching-based correctness, so this would capture responses that don't have DNSSEC Signatures. So, there are some new texts in the document around that. And we can discuss the advantages and the disadvantages of these approaches and maybe think about whether or not we need both.

For the staleness and freshness, this is really one metric and I keep calling it both staleness and freshness because I don't think we've decided what to call it yet. But anyway, there are currently four different methods proposed for doing this and for myself after thinking about these for a while, I'm kind of leaning towards the last one, the fourth one. And so, I thought we'd talk a little bit about that and maybe see if we can come to agreement that this is a reasonable approach.

And then at some point we need to sort of... You know, this is pretty terse so we need a lot more detail around how this would actually work.

And if you've been seeing Paul Hoffman's emails to the list, this I believe is the approach that he has implemented. Correct, Paul, in your Proof of Concept?

Paul Hoffman: Yes.

Duane Wessels: So, he has some data around this that we can look at as well.

Paul Hoffman: Yes.

Duane Wessels: And then for thresholds, really, I just wanted to set the stage for our future call where I would like to really get into the thresholds and also, we have a discussion point about the extent to which it makes sense to have thresholds for RSS Metrics. So, that's the plan. Assuming there's no significant issues with that, I guess we'll dive into the first part. Okay.

Yeah, so the vantage points. As I said, there were previously just some placeholders here, and based on my sense of what the Work Party has been discussing and previous calls, I went ahead and put some starting text here. So, for the number of vantage points, it says that the Work Party or RSSAC would recommend that there be at least 20 vantage

points, formally or also known as probes. And I wrote here that this number was chosen to balance the competing goals of having a lot of vantage points for good coverage but not too many vantage points that they become unmanageable, especially since this thing doesn't exist yet. So maybe as we get more operational experience with this, we would do decide that it's better and reasonable to have more vantage points.

So, the number on the table is 20. Anybody like to propose different numbers or support this or have anything to say about that number? Open it up.

Warren Kumari: This is Warren. I'm not sure if we're using raised hands.

Duane Wessels: Oh, I see your hand. Yeah, go ahead.

Warren Kumari: No worries. I still believe that 20 is way more than we actually need. Especially because the latency in reporting of that sort of thing where lots of visibility is needed, I think is going to be banded. So, I think this is way more than we need. If something looks okay from one location, it should look okay from the other and the additional fetch and management to manage that many seems annoying. I don't really care enough to fight that much more.

Duane Wessels: So, what would you suggest? Ten is a good number?

Warren Kumari: I would think that ten is more than enough and it's much easier to spin up and get organized. It also, if one has 12 and two of them die, you've still got ten. Whereas if you actually need 20 running, you probably should have few spares for when those fall over or are under maintenance. But if ICANN has additional funds to spin them up, it seems like as good a use of those funds as much other things, so whatever.

Duane Wessels: Okay, thanks. Paul?

Paul Hoffman: So, Abdulkarim asked any rationale for 20? And I, until almost just yesterday, would've agreed with Warren of ten is fine. But we haven't started discussing how the thresholds are going to be apportioned for the various tests. And one thing that I came up with yesterday is, "Oh, well if 95 percent of the probes are doing whatever, at which point you need 20 if we're willing to."

So, I think for now, 20 is a reasonable number. I could also believe that after we do the threshold conversation, we might drop that to ten or 12, but for now saying 20 gives us the opportunity to say, "In the

thresholds, blah to blah, 95 percent of the collectors saw blah." Something like that. So that's really, to me, that's a rationale.

Duane Wessels: Okay. That's a good point. I can add that to the document. I mean, 20 works out to be a nice number in terms of percentages but... Yeah, okay. And, so Abdulkarim, does that sort of answer your question? In my opinion, it's a little bit hard to come up with really strong rationales. Like I said, this is sort of a little bit pulled out of a hat but designed to balance these two goals of having coverage and manageability.

Abdulkarim Oloyede: Yeah, thank you. It does.

Duane Wessels: Okay. So, Warren, in the chat I see you said about 20, so I mean that's fine with me, I guess. Approximately 20 would be good. Alright, let's move on to the next section then.

This is about location of vantage points. I sort of took this from a previous phone call that we had where somebody mentioned the five ICANN geographic regions. They're listed here. Yesterday I did just a very quick research on the extent to which these five regions, how they divide up in terms of number of internet users and population of people and things like that. And based on what I saw, it felt to me like it would be appropriate to just say that the vantage points should be evenly distributed among these five regions.

It's not perfect, but it's probably close enough. Otherwise if we wanted to get more clever and say weigh things more towards more populous, or regions with more users, then we would have to agree on what data do we use for that and how does that work.

So, any comments about this very broad distribution of location of vantage points? Oh, and then I guess the other thing, you can see down after the list, it says the vantage points should be located within distinct major metropolitan areas within each region. So, don't put two vantage points in the same city, for example. Ray?

Ray Bellis: Yes. Kind of caveat that shouldn't just be reliable connectivity but high-grade connectivity. That's not going to be at the end of the wet piece of string, even if it's a very reliable piece of string. It should be 100 mega gigabit, for example, style IX Grade connections.

Duane Wessels: Yeah, that sounds like you're reading in Section 3.3 about the connectivity.

Ray Bellis: Yes. Sorry, I thought you had just moved on. I apologize.

Duane Wessels: Not quite but these things are related to some extent. But I guess I wanted to get agreement on the geographic diversity first. But, yeah,

definitely I agree with your point about quality of connectivity. Any comments for now about the location of vantage points? Tom?

Tom Miglin: Yes. Going back to the sentence about the distinct metropolitan areas, I understand you just talked about what you meant by that. But I guess when I read that, it tells me that you're limiting vantage points to only be located within major metropolitan areas. I think you were just trying to get not two of them in the same region or right next to each other.

Duane Wessels: Yeah, I guess my assumption was that they would be located in major metropolitan areas out of necessity anyway because that's where you would have data centers and concentrations of bandwidth and things like that.

Tom Miglin: Okay. So, your intention is to use that sentence for both those points then?

Duane Wessels: I guess so. I mean, really this is all new and up for debate. But if you think that it makes sense to, or if you think that's it okay to, have vantage points in not major metropolitan areas then we should consider that. But I kind of assumed that they would naturally happen anyway, I guess.

Tom Miglin: Okay.

Duane Wessels: Do you think it would be better if those two points were made separately and not in a single sentence here? Is that what you're getting at?

Tom Miglin: I think that not having two in the same metropolitan area, I think you could probably spell it out a little clearer.

Duane Wessels: Okay.

Tom Miglin: In a separate sentence.

Duane Wessels: Okay, alright. Warren?

Warren Kumari: Sorry, I was looking for the unmute. So, I mean, I don't really care. This seems fine to me. I still think that the latency banding is going to be large enough that if we stuck them all in far away places or near places

or major metro areas or not major metro areas, it's not going to skew the results enough. So, any of these seem okay to me.

Duane Wessels: Okay.

Warren Kumari: I still think we could even have all of them in one ICANN geographic region, but I think that a lot of this is now we're doing this for political reasons so...

Duane Wessels: Yeah, I hear what you're saying about the, I mean I think what you're saying, is that the thresholds are going to be high enough that they can be met almost no matter where you locate the probes in terms of latency, right?

Warren Kumari: Yeah. And also, that the fact that I've got a probe in Asia, if I'm trying to reach a root server located in Fiji, there is no guarantee at all that my traffic is going to be staying within a geographic area. So, I think it's a fantasy that we think that patterns are going to follow this in any sort of believable way. I actually chose Fiji because when I was in there, a huge percentage of my traffic hit both North America and Europe and didn't actually follow across to anywhere in the same, what would count as the same, region. So, but you know. If we leave [inaudible] out, people will be sad.

Duane Wessels: While I agree that it's likely that our latency thresholds are going to be high, at least for now, they may not always be so high, and some other party may decide to have different thresholds in which case the location might matter a little bit more. I don't know. Paul go ahead.

Paul Hoffman: This is Paul. So, we're having a problem here where we keep focusing on what our thresholds will be. But as you, Duane, as you brought up, you always want to discuss even today Root Server System metrics. So, if we are going to be publishing any metrics about the Root Server System as a whole, I do think we need, it's not just political to say, "Oh, we're going to spread the collectors all over the globe."

I think that if we say, "Here are some metrics of the Root Server System as a whole, and all the collectors are in a subset of the world.", that will just be wrong.

Duane Wessels: Right.

Warren Kumari: If I can quickly respond to that. You're not saying you're going to spread them all over the world. You're saying you're going to distribute them evenly amongst the five ICANN geographic regions. If you said you're going to spread them all over the world, I'd nod happily and move on. But divided evenly amongst the five ICANN geographic regions implies

that they have some sort of priority for specialness or that the regions were chosen for technical distribution, not political distribution.

Duane Wessels: So, Warren, is there a way to say spread around the world that everyone can understand and agree to?

Warren Kumari: If you dropped... Yeah, I guess if you just said, "It will be distributed evenly amongst", and then just listed those and just dropped the five ICANN geographic regions would make it seem more that what you're doing is trying to reach a technical decision, not a decision where ICANN's drawing on the map is what's important.

Duane Wessels: I see. So, your objection is the fact that it names these as ICANN regions, not so much that the regions themselves.

Warren Kumari: Yeah. And also, a little bit also the 'evenly' part. I mostly disagree with distribute amongst blah, blah, blah. That would be fine. 'Evenly' is less annoying or less concerning to me.

Duane Wessels: Okay.

Warren Kumari: And sorry for being so riled up about that particular point and I didn't mean to sound as grumpy as I did.

Duane Wessels: No problem. Karl?

Karl Reuss: Yeah, I was going to suggest possibly just removing 'evenly'. I think it reads a little better and 'evenly' is kind of up for debate, what you're being even about, numbers or...?

Duane Wessels: Well, I guess my concern about dropping 'evenly' would be that if somebody goes off and implements this, they can put one in each of four regions and then the other 16 in North America or something like that. So, my goal was to have something in there that talks about that they shouldn't be so unbalanced.

Warren Kumari: What's your metric for balance? Shouldn't your metric for balance be number of internet users? I mean, if you do want to balance. Or number of viewpoints?

Duane Wessels: I would love to be able to say that it's distributed in a way that corresponds to number of internet users. But I think that then we have to agree on how do we know where the users are and who's data do we

use for that and what if that data goes away, or whatever. So, if that would be preferable then I'm happy to put that in if the Work Party agrees.

Warren Kumari: I had kind of thought that we'd just assume that whoever's building it... Maybe that's a bad assumption. Yeah, I don't really care. I'll go crawl back under my rock.

Duane Wessels: Paul, is your hand still up?

Paul Hoffman: Yeah, so because in my work we are told to look at the number of internet users distributed, I have seen wildly different numbers including two wildly different sets of numbers from one organization, about counting users. And I propose that that's not our work here. If we're going to try to do anything based on, we're trying to distribute these, it would be by where the resolvers are, not the number of users, but the resolvers.

And when I've looked at those, they are pretty much just smeared all over the map because even a teeny little ISP sets up a resolver. So, I think it's best to not try to evenly distribute by users or resolvers. At which point, I agree with Duane saying evenly across whatever, whether they are the ICANN regions, or just those five regions. Although everyone looking at that list will recognize that those are the ICANN

regions because the Caribbean Islands somehow got stuck with Latin America, not with North America.

I think it is important to say ‘evenly’ for the very reason that you gave, Duane, which is that a lazy implementor would say, “I need one in each of these. Now where do I want to put the rest?” And I don’t think they would put them all in North America. I think they would put them all in North America and Europe and I believe that would make our metrics less believable.

Duane Wessels:

Okay. Thanks. Alright, I’m going to move on in the interest of time. I think we’re a little maybe behind schedule. I see there’s some stuff in the chat and I see it but let’s move on anyway.

So, Ray was talking about this section, this title, Connectivity and other requirements, and here it says, “The vantage points shall be hosted inside data centers with reliable power and connectivity.” And Ray’s comment is to say something about good connectivity or minimum bandwidth requirements. So, I can get behind that. That’s good.

And then the next part says that the vantage points may be deployed on bare metal or virtual machines. And this is sort of to address I think a comment from Warren about how sometimes if you’re using a virtual machine, you don’t have full control over the operating system so it says when VMs are used they should provide a dedicated address and the dedicated operating system environment.

Warren, your hand is up again or up from before? I'm not sure. Oh, Karl?

Karl Reuss: I was going to suggest that maybe we want to say, "Diverse connectivity providers". I'd hate to see them all put behind one ISP because of a contract award.

Duane Wessels: Yeah, that's a good point. Is the phrase 'diverse connectivity provider' sufficient or do we need to be more specific than that?

Warren Kumari: So, you might want to be slightly more specific because I think that finding 20 different providers might be tricky.

Duane Wessels: What about diverse connectivity providers within each region?

Warren Kumari: Sounds reasonable but you then run the risk of you end up with NTT and Level 3, like there just being two. Which might be fine. Just pointing out. I don't care.

Duane Wessels: Okay.

Warren Kumari: And that also wasn't meant to sound, "I don't care. I'm taking my ball and going home."

Duane Wessels: Okay, so thanks everyone for the input on this. Let's move on. It's already about halfway through the call. So, correctness metrics. I think folks are probably already pretty familiar with the old one or the existing one which uses DNSSEC. We did add a little bit of text to this describing in more detail how you would do validation and the things that you need. So, you need a trust anchor, for example, and you need the DNSKEY RRset. Can you scroll down a little bit, Ozan?

Yeah, so here in red is some of the new text about that. One thing also to point out here is that you'll see that in a lot of these metrics now, there's like a little paragraph that says reporting in it, and it talks about how for all these metrics, the idea is to report only whether or not the metric meets or does not meet the established threshold. Previously there was words here about, maybe not in this one but in other metrics, there were words about twenty fifth fiftieth, seventy fifth percentile and so on, and so those have been deleted.

So, let's scroll down to the next one. I'm sorry. That's too far. I meant, oh yeah, I forgot they're separated by... They're not adjacent. So, Section 5.5 is Root Server Correctness by Matching. I believe Paul is the author of most of this text. And the general idea is that you can do queries, and you also have a copy of the root zone and in fact if you

need here's some historical copies, here it says root zone copies up to 14 days, I believe, is that right Paul?

Paul Hoffman: So, no. You didn't see that I had actually mucked with this this morning. There's now two options. One is the one you were talking about which is that the collector keeps maybe 2 weeks' worth or whatever. But option 2 which is just off the screen, so Ozan if you could do this, is actually simpler which is to say the collectors just collect and they send the answers back to the place they were going to be sending the metrics reports anyways and that place is the one that checks the correctness.

And so quite frankly looking at the stuff that I sent yesterday about how quickly each root server gets the new zone, we can say you have to be in the last three. Not the last three days, the last three... That the checking will be done against the last three. That we can assume that if a root server is more than a day and a half behind, other bad stuff is happening.

Duane Wessels: So, my one issue with that and I think we've talked about this, is that you can have three or four root zones in a day, right?

Paul Hoffman: Yeah. And I believe from looking, I mean it's a very limited dataset, but I believe that every RSO that we're looking at probably has things up to date within 15 minutes. So even if we have two or three in a day, and

basically most of them have them within five minutes, these tests should be still sufficient. Or we could say five instead of three.

But the main difference between one and two is one checks the matching on the collector and two sends it off to somebody else, whoever's collecting the latency numbers anyways and they're the ones who do the checking.

Duane Wessels: Okay. Do you have a favorite?

Paul Hoffman: I do now, actually, which is why I wrote option 2 in just an hour ago. So, on my Proof of Concepts System, I'm doing option 1 and I discovered that in fact even though one would hope that a webserver would only say, "Yeah, I've got new data for you", when the data actually changed, it's actually saying every hour, "Here's another copy of the root zone." So, I think the idea of actually checking an essential place might be better.

It also means then, depending on how we do the other things, that then the collectors really are just collecting data. They're not doing any analysis. However, if we do option two here, I would hope we would do a similar option to equivalent for if we're still going to do the DNSSEC validation, which is to say, "The collector got this answer. Hey, you, central collector, you're the one who's going to validate all of them."

- Duane Wessels: Okay. Thank you. Warren, your hand is up?
- Warren Kumari: I point out that having this all go to a central collector means you need a couple of central collectors and you're going to have a fair bit of entertainment during data synchronization, etcetera. I think what would be reasonable is to get the last end zone files, you know, three or whatever, over the last six hours, whichever is larger. That way we deal with the oops, somebody mis-cut a zone file, here's a new one. Whoops, we did it again. Here's a new one. There's also, yes, if you get a copy of the zone file and then check and get a new one, you should be able to fairly easily tell that because to SOA is different when you download it from the server.
- Paul Hoffman: Yeah, Warren, to be clear. Yes, that's all fine. I was trying to have, I don't want our 20 systems hitting on the zone file distributor more than necessary. I mean, it is only two megabytes now anyway so it's not a big deal for a regular webserver. But option two I just put up as a sort of a politeness for the people who are distributing the zone file. And I agree with you that option one is more stable than option two because you're doing the calculation right there and so it doesn't really matter as much if the central guy has fallen over. So, either way should still get you exactly the same answer since we're doing exact matching.
- Warren Kumari: Yeah. I mean the obvious easy solution to a bunch of this is when a collector gets a zone file, or when a collector gets an SOA that is higher than its currently known SOA, it uses that as a trigger to go fetch the

zone file from the webserver. That way you only need to hit it once somebody has already told you there's a newer version of the zone file. You don't need to pull.

Paul Hoffman: Yeah. So, that's another way of doing that. So, yeah. But Duane, I think the more important part, unless you want to decide between those two now, is how to do the exact matching which is that next paragraph after the option two paragraph.

Duane Wessels: Okay. So, one thing I note here is that you are proposing only exact matching for NS queries.

Paul Hoffman: That's because you were doing all the others with the validation. If we end up not doing validation as the last paragraph says, we can do it for everything in the zone.

Duane Wessels: Okay. So, but as proposed here, it's NS queries, in which case as you note, the intersection should always be empty because it should be a delegation. And if the authority section contains the entire NSR set and that the additional section may or may not contain all of the glue records. But for every record that is present, it should match. Right?

Paul Hoffman: Yeah.

Duane Wessels: Sounds reasonable to me.

Paul Hoffman: And then if the Work Party wants to not do validation and do it all through direct matching, then number one needs to be modified. Or one and two need to modified for NS and non-NS. But other than that, I think the rest would be the same.

Duane Wessels: So, yeah. I think just to be clear for everyone here, I think what Paul is talking about is that if we have this exact matching approach, then you could just as equally use that for the dataset correctness. They're sort of a superset of that. And you could check signatures and everything like that so it may not make sense to do trust anchor base validation on the vantage points. Right, Paul?

Paul Hoffman: Yeah, that's the way I'm leaning but if we go that way, this list will have to be expanded to also check for proof of non-existence. So, that is if we do a DS Query for a ccTLD that does not have them, we also need to check that the answer section has nothing match against the root zone that says, "These records do not exist here."

Duane Wessels: Yeah, and it's probably even more complicated for non-existent names, for NXDOMAINs because then you're exact matching has to know a lot about how DNSSEC works and figure out which NSEC Record should be there and things like that.

Paul Hoffman: Right, whereas for this, if the query is for .example, all you need to do is to make sure that if the root zone had no records for .example, the answer section was empty.

Duane Wessels: Yeah. Well, I guess personally my feeling is that it would be good to try both of these and get experience with them and see how they play out. I think that they may give us slightly different results, slightly different measurements. And the people that have to implement them may learn some things that we're not thinking of right now that make one or the other harder. So, I guess I'm leaning towards keeping both the DNSSEC based one and the exact matching one and given that you're favoring option two here, I would suggest to keep that, and I guess strike option one at this point.

Paul Hoffman: I'm not sure. Actually, after what Warren said, I think I would tend towards option one again. Just on a... Once this thing is running, it would be better if the central server was just collecting metrics so that if it fails, they could be resent, might be better. Also, quite frankly, I don't know how to tool the doing the DNSSEC validation, which you seem to

be doing also on the vantage points themselves. So, I would say let's just leave option two in, but I'll continue to do option one because I already have the code for doing it.

Duane Wessels: Okay.

Paul Hoffman: And really the only thing is I noticed, oh my god, it is actually pulling this down every hour and it's pulling it down for the correct reason. The webserver says that it changed the ETag, which was for no good reason. But I mean in the real world once this gets deployed, I assume that the server that's going to have the root zones will be run by the RZM and it will only allow the collectors, that it'll use some password protection, so that only the collectors can get it and therefore it's really not going to matter if something's pulling down too often. We'll have better control.

Duane Wessels: Okay. Alright, thank you very much for that discussion. Any last comments about correctness before we move on? Okay. So, let's go to the staleness metric. Here there are four methods proposed, or four options. The first one is described in some detail. This is designed so that the vantage points can do a lot of the work. And the idea is that they send a set of SOA Queries and they wait for ten minutes and set another set and then they can use those two sets to find out if any of the servers don't have the most recent SOA. Scroll down to the next one.

The next one is, let's see. So, this one is initiators keeps the state for the most recent SOA Serial Number, every minute it checks all the RSOs. This one is very similar. Maybe not as complex because it doesn't have to send two sets of queries but it's not as well specified, I think.

The third method, from Fred, again, it's maybe not as well specified. I have a little bit of a hard time imagining how this would actually work until we can get more detail.

But then the fourth method, which is the one that I'm sort of in favor of, has the nice property that it reuses the queries from the other metrics, the other measurements. So, it reuses the SOA queries that are happening already on a five minute basis. But it does have the requirement that the central processing system has to do the metric calculation because it will have access to all the data from all the vantage points.

And I think as Paul has discovered, there are some interesting challenges when the SOA changes happen close to the date to the 24 hour boundaries, so it's not a matter of simply looking at all of the data you have for that day. The system will have to look at data that spans those boundaries or it will have to say if the change happened within a day and I didn't get it, they need to discard the measurement or count it as stale or something like that. Paul, is there anything more to say about this since you've actually implemented it?

Paul Hoffman:

No, but if people saw the message I sent out last night, the second one, not the first one because the first one, as I said in the second one, had

plenty of bad data and thank you Duane for noticing that. The bigger takeaways is that essentially, we can assume that a full change for the whole root server system is going to take more than five minutes. That is for the four in the message I sent out, none of them happened in one five minute interval and some of them took more than ten.

So, I think we can assume the idea of one minute checking is not going to be valuable because our thresholds are going to have be much larger than that. So, I think going with the five minute checks is good. And then the other thing is that we, when we're doing threshold decisions, are going to have to decide do we want to combine multiple collectors into a single threshold. And the reason for that is that, quite frankly, with many of the root server operators, their systems are frontend processors in front of a large number of boxes.

And at least in the very early data I've seen, those boxes are not all updated within the same second or two, that we saw differences going to the same place. So, I think that this method, even though it's collecting less often than we had originally said, is going to be perfectly sufficient for that unless other people think that we really want to be having tighter thresholds.

Duane Wessels:

Yeah, I think five minutes is adequate. Warren, your hand was up and then it's down.

Warren Kumari: Yeah, I was just going to agree and say that five minutes is more than sufficient. You don't want to end up using a micrometer to measure a two-by-four.

Duane Wessels: Right. And our section two has a metric around this as well. And the most recent version of that document says something similar like even though the number is reported with a resolution of seconds, it shouldn't be interpreted as needing to be that accurate. That on the order of minutes is fine. But, anyway.

So, one idea I had, this is a little bit of a tangent, but from looking at Paul's data, one idea I had was I think it would be very useful if the Work Party had access to some of Paul's raw data, maybe for example in a GitHub repository and then if we asked individuals to take that raw data and take the text in this document and calculate what the metrics should be. I think it'd be interesting to get a number of different folks looking at it and trying to code it up and to see where some of the difficulties lie to better advise the people that have to implement this for real.

So, that still may happen. It hasn't happened yet, but it would be good to do that not just for this metric, which I think is one of the more complicated ones, but even for the other metrics as well.

Paul Hoffman: So, this is Paul. I'm happy to publish it. I don't think that there's any one at ICANN could claim that any of this is PII. I could either do it as a

GitHub repo or in some central place. But I've heard informally that other people also either are doing this or have been doing some others, so I think having multiple datasets, especially if they disagree, would be useful so that we could figure out what kind of data collection should be done.

But I'm happy to put up mine. Just as a warning, I've just rejigged the system, so I only have about two days' worth of data now, because I changed the database format for various reasons. But if, Duane, you're thinking of maybe starting this in about a week or so, I think that that would be enough. And if other people are doing it, then we could all just tell people, "Here's where our data is."

Duane Wessels:

Yeah. I think that's good. And I would also advise that when we do this, not that it matters a whole lot, but maybe we should change the operator designations, scramble them up a little bit and not call them ABCD but call them something else, NOPQR, whatever. So, anyway.

Alright, so getting back to freshness/staleness, any other points on this before we wrap up with the last topic? Is anyone, I guess let me ask directly, is anyone opposed to just proceeding with method four and abandoning these other three methods at this point? Okay, I don't think so.

So that will be the proposal going forward then. Alright. Let's talk a little bit about thresholds. The goal that I would like to get to is that the RSSAC has a workshop coming up in about four weeks, three or four weeks, and at that workshop we want to spend a lot of the time talking

about thresholds. We want to have a lot of this stuff settled so that the RSSAC Members there can have a discussion about what are appropriate thresholds for these metrics.

So, first of all, I would like this Work Party to take a first stab at what some of the metrics should be. So, I'd like to do that on our next call in two weeks. Everyone please send some time thinking about what is an appropriate threshold for latency, what is an appropriate threshold for availability and for freshness and correctness, and we'll have that discussion there. And then we can use that as a starting point for the discussions with the broader RSSAC Group at the Workshop. Warren, you have a comment, go ahead.

Warren Kumari:

Quick question. So, for these thresholds and things, is it figuring out what the threshold is for the ideal most perfect root server system, or thresholds for everything seems to be working mostly kind of okay at the moment so something that sort of keeps the status quo? Okay? So, we can find when there's something that's actively breaking, what should the metrics that we're thinking about for the next few weeks be aimed at?

Duane Wessels:

Well, I would encourage you to think about more along the lines of your first one, but I wouldn't probably use the word perfect. I would use the word good. There's a phrase that we used a lot in the RSSAC discussions which is what does good look like. But I would discourage you from designing thresholds that... I would discourage you from thinking of the

current system when thinking of thresholds. We shouldn't necessarily be developing thresholds that just the current system could meet.

Karl Reuss:

So, the foundation of this Work Party was centered around the fact that 37 had a pass/fail type measurement issue that needed to be dealt with. And the point behind coming up with the metrics in the first place was to design really where that fail point is, which to me leads not so much deciding what good is, but if we're going to try and meet that goal of helping 37 eventually get realized, the metrics that we really want to study are, at what point do we need to be concerned and that's very different than what does good look like.

Duane Wessels:

Okay, a fair point.

Karl Reuss:

That doesn't mean that we can't sort of do both but in my mind, getting ahead of what's going to be stood up for that model is really, the requirements for that, unless and correct me if I'm wrong anybody, but the requirements for that are much more centered on what does bad look like.

Duane Wessels:

I guess. I struggle with that we've never, I don't think we've talked about it that way within RSSAC really. We've always talked about what does good look like, and maybe that's just because we were being too simple in our discussions.

Karl Reuss: I think we have talked about it and I think if you think about when the designation removal function is triggered was based on an audit fail for you are not performing up to needed metrics, minimum needed metrics, the minimum bar, right? And you've been warned once, now you're being checked again and if you fail this, then the DNR function will kick off.

Duane Wessels: Yeah, and maybe what we're struggling here with, too, is that our thresholds, at this point at least, are binary. They're either pass or fail. We did talk previously about something a little more, I guess, complicated. We talked about light colors, right, like green, red, yellow, or something like that. But, at this point, we're still focused on pass/fail. So, I guess taking that into account, that's why I said it would be better to have that in mind rather than, certainly not what is perfect, what is ideal. Paul, your hand is up?

Paul Hoffman: Yeah, so Duane, do you want us focusing more on the RSO thresholds, or RSS thresholds? Because Warren was just asking about RSS thresholds and I personally hope we don't touch those. RSO thresholds, I think, are what we've discussed more so far. Even though we haven't discussed thresholds, what we've aimed at more so far in the Work Party. But RSS thresholds are a completely different beast.

Duane Wessels: Yes, I'm inclined to agree that it's a different beast. I think that the RSO thresholds are easier to contemplate. RSS thresholds, at least the way they are defined now, is just sort of aggregating all of the RSOs together and sort of doing the same math and then you get a value for the RSS.

Paul Hoffman: Wait, wait, wait. That's the RSS metrics.

Duane Wessels: Yes.

Paul Hoffman: I don't think you can multiply the RS or divide the RSO thresholds by 13.

Duane Wessels: No, no, no, that's not what I'm suggesting.

Paul Hoffman: Okay.

Duane Wessels: I'm not suggesting that at all. No. What I'm saying is that for something like availability, if you wanted to calculate RSS availability, you just take all of the RSO measurements as a group and do the same aggregation to them, right? And so, you just get a percentage, a number, for them all aggregated together.

Paul Hoffman: But those aren't thresholds, those are metrics, and I thought you wanted us...

Duane Wessels: That's right, those are the metrics.

Paul Hoffman: I thought you wanted us thinking about thresholds. So, I'm sorry to be pokey here.

Duane Wessels: No, what I'm trying to get at is that given that an RSS metric is just an aggregation of all the RSO measurements, assuming we do have RSS thresholds, should they be any different than the RSO thresholds? So, for example, if you had set RSO latency at 300 milliseconds as your threshold for an individual RSO, is there any reason to think that the aggregated RSS threshold should be higher or lower than that? Or should it be the same?

Paul Hoffman: So, in the remaining two minutes, I think that this is a list discussion or a face-to-face discussion, but Ken Renard asked a really good question in the chat which is do you want us to have two sets of thresholds, minimum and desired, or do you just want us to have one?

Duane Wessels: I get to decide this?

Paul Hoffman: You're the one who asked us to think about it.

Duane Wessels: I know.

Paul Hoffman: Maybe we also have to have that as one of our questions, which I think is reasonable. But given that the first thing that Warren blurted out mixed those two up, and I'm not saying that's bad of Warren. I think we do need to be way explicit about what are you asking us to think about?

Duane Wessels: Yeah. So, it's a good question and so far, my sense is that we have only been discussing having single thresholds, but if the Work Party would like to say, "Here's two types of thresholds to consider." I think that would be very valuable input as well. So, I think that's a good suggestion, Ken. Let's feel free to do that and then we can take that to RSSAC and see how they like that idea.

Alright, so we're out of time. The discussion about thresholds is going to be very interesting and drawn out, I'm sure. Hopefully we can make a little bit of progress ahead of the workshop, so let's try to do that on our next call. And for anyone here on the call or in the Work Party that's not an RSSAC Member, that will of course be an extended Work Party Session you'll be welcome to join remotely. You're welcome to join in person, I believe, if you can get yourself there. Unlike previous

meetings, there's no travel support but if you're in the area, you're welcome to attend in person. And we'll continue that discussion there.

Was there anything else to discuss before we wrap up, Steve or Ozan? I guess the next call is in exactly two weeks.

Ozan Sahin: Correct, Duane.

Duane Wessels: Yeah. Meanwhile, everyone please continue to feel free to leave comments in the Google document. The input is always welcomed and with that I think we can wrap up the call, unless there's any last comments. Alright, thanks everyone for your time. Talk to you later.

Warren Kumari: Thank you.

[END OF TRANSCRIPTION]
