
UNKNOWN SPEAKER: So, this afternoon, as Duane indicated just before lunch, we are going to start off with the discussion about RSS Metric. The first one that we want to try to tackle is availability. We had one of our remote work party members, Paul Hoffman, has put his hand up to begin the session this afternoon, and he said he had his hand up before we started, and that was an accident. So, Paul, why don't we let you go ahead and start.

PAUL HOFFMANN: So, I wanted to hop in before we started on RSS anything, to have an observation of where we got in the last day or so on root server operators. And people may disagree with the way that I'm slicing and dicing this, but this just sort of came to me while you folks were eating lunch and I was taking a shower this morning.

We have two types of metrics now for an RSO, those that can be externally verified and measured, and those cannot. So, those that can, we have the list that we started with, latency, availability, correctness, and staleness. And what we came to yesterday and today, I think, that is sort of in addition to that, is that some of those of interrelated. So we might want to look at those interrelations more. We also have metrics that cannot, bandwidth packets and queries, BPQ, we started with, and then yesterday and today we started talking about another metric that might be reported as a metric, that is not externally verifiable, which is your instance count.

That is a metric, that is a number, it can't be externally verified, and it may or may not be part of what the working group wants to do. But

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.

then today, especially in the last session, as we were going around like what are people going to optimize for, two other metrics that cannot be externally measured came up, which was instance location geography and instance location topology. There be dragons there and all that, but they came up often enough where I want the work party to at least think that these are things that we need to deal with even if only to reject them.

And then the other thing that hit me, that gout wound up, and I wish I had thought of this a few hours ago, was we actually have a new set of metrics that we care about, which is the metrics of the probes themselves. Daniel brought up some interesting ones of the non-DNS latency, but as soon as you start thinking of that, but also get non-DNS availability and correctness and staleness. So that's like a new metric that we care about that we need to describe. And then there's, of course, the nonmeasurable ones of those which is also geography and topology of the probes.

But as we got into the discussion of where do we put probes and is it good to have them in Miami versus in Quito, or whatever, it became clear that we have metrics of those, as well. So, anyways, metrics are the things that we can probe and collect, great. Measurements are the things we get after we do some mathematical complication on the metrics, but it's pretty clear that the work party is going further than the statement of work, and I think that that's completely necessary, but let's do it with some good definitions. Thank you. Let's go into RSS.

UNKNOWN SPEAKER: Okay, Paul. Thank you. We not only have been recording all of our work party sessions here, in this meeting, but also a number of us are taking notes and getting things written down, and having this nicely enumerated summary you just did is very helpful. Yes, Brad?

BRAD VERD: Really quickly, Paul, I've said this again numerous times, I'm happy to say it numerous more times, statement work is meant to be a guideline and it is up to the work party if they want to change it, add it, modify it, do whatever they want to it, in order for them to feel that they've given a good, thorough answer.

UNKNOWN SPEAKER: Good, thank you. Okay, so, with that, let's jump right into our first RSS metric. We built these based right out of what the SOW says, in terms of the order that they will be discussed. They can be rearranged and such, but we're starting with availability. Right now, I don't know if folks have actually read, it's on the screen in the Zoom Room and it's available here on the screen in this room. We have had quite a range of inputs from various work party participants about what does the RSS Availability metric, or how should you actually described it?

The range I think covers everything from if your probe however many you may have, if your probe can get an answer from any one of any instance of any RSO, the RSS is online. Closer to the other end of the spectrum is that your probe has to be able to get an answer from every RSO for the RSS to be considered online. So there is a wide spectrum

here of inputs we've had so far. I think I saw Brad or Jeff first, I don't know. Brad.

BRAD VERD:

Yeah, I've heard some of these discussions on kind of defining what the RSS is. I think in hindsight maybe this could have been worded differently, but I think the intent, I'll look at Wes when I say this, I think the intent, as the other shepherd, the intent is the service is online. So if you get an answer, the service is online. We need to think of this as like a black box, and the reason that maybe we're having such a hard time doing this, is because we've spent 30 years focused on the individual root servers and not the service and then in just the last year or two, we've tried to change that focus to look at the service as a whole. And so that's what this was trying to do.

UNKNOWN SPEAKER:

Thanks, Jeff?

JEFF OSBORN:

Plus one, because Brad said it better than I could have. But, in addition, recognizing that if there is something there that's answering the system is up, begins to address the ICANN Board problem of what if there's an outage, I'll be ruined, where we kept trying to say you're not recognizing how incredibly robust the system is. And so anything other than the measure of if you get a response back from anywhere, the system is up, it's creating an artificial frailty that isn't there in the system. This thing is so robust that it doesn't matter that you got it from nearby, it doesn't

matter that you got from where you wanted to get it from, you got it. I think that's a really important point in proving.

UNKNOWN SPEAKER: Wes, I think you're next.

WES HARDAKER: Yeah, then Brad and Fred, watching the hands raised, you had your head turned. So, with my shepherd's hat on, thanks Brad, I was determined not to talk this hour, but I will. I think if you look at the statement work, the goals of this, you know, we started this work specifically pointing to the PMMF which is specifically designed an RSO and its performance level. But the reality is I think that we have to do both, and we have to look at how do we measure this, the service as a whole, and how do we measure through RSO, and unfortunately that doubles the workload.

UNKNOWN SPEAKER: Well, let me just comment to that before we go on, and that is right now, at least the way Duane and I structured the work and the work party is there dealing with RSO, and the same set of things, and a hunt going with RSS, and our objective here this afternoon, the RSS discussion, we can change that if necessary, but I think having them separate is an attempt to make it easier to chew through that problem. Okay, Fred, you're next.

FRED BAKER:

Well, I'm going to repeat myself from an email that I sent to the caucus, plus this morning. I think at least part of the issue in this particular metric is the way it's phrased. We can measure an RSO in a binary sense. I can send something to an address and I'll get an answer, or I don't, old style. With the RSS, that's significantly more complex. But I wonder if it's sort of asking the wrong question.

And what I suggested this morning was that we send a request of some kind, whatever it is, to each of 26 addresses and then we report on what we got back. We got 26 correct responses, or we got none of them, or this guy I got one from, from that guy I didn't, and be able to say I can talk to five RSOs, I can talk to three RSOs and three more with an unreasonable latency, you know, that kind of information.

The RSS isn't a block, that I can talk about it in a binary sense, and I think an axle that we get wrapped around is trying to describe it in a binary sense. And I think maybe the issue is the description. We want to talk about our ability to interact with 26 different addresses, send a request and get a response. What's my experienced with the RSS as a group of 26 addresses, 13 RSOs.

BRAD VERD:

So, Fred, an interesting way to look at it, I think maybe I would tweak it a bit, and that is I think there is a metric on the service, meaning a client doesn't care if he gets an answer from all 26 or 1 of the 26. It just wants and answer, and it can carry on. I think that's important for the messaging to the Board, to the community, that this resilient service is available and you can get an answer. I think there is another aspect that

you described very well of the user experience, which you could use in collecting metrics to find out, this goes to the underserved topic of one of our work parties.

At one point you talked about having a client on the phone that you could hit a button and say what is my experience with the root service. That experience would be the experience with all 26, so that you could kind of see what works and what doesn't work. But it concerns me a bit, because I'm concerned with the messaging to the Board and the community about this idea of focusing on the single letters or identities, rather than just the service, which is just an answer.

FRED BAKER:

Well, understood, and I don't disagree with you on that, but there are other aspects. For example, upstream of me there may be a network that doesn't carry IPV6 and as a result there are 13 different addresses that are never going to respond to me. Not my problem, not the RSO's problem, but somewhere in the network there is a problem, and if we need to diagnose it, we're going to need to know that.

And there are other aspects there. So I think if we design the metric too narrowly, I know, the one question I've got is the Board, can I convince them to peel themselves off the ceiling? In the long run, I don't think we wind up with as useful information as if we write the question, what I'll call correctly, and then the metric that results from that to help form questions that we have down the line.

UNKNOWN SPEAKER: Let me ask Fred at this point, if we use the approach that it's 26 addresses and somewhat you combine them in some manner to talk about the RSS, can you explain how that is different from the RSO metric of the same nature? The RSO availability metric?

FRED BAKER: Well, and what I'm doing is using the RSO metric 26 times. And then characterizing the RSS currently, composed of 21 companies, 13 identities, 26 addresses. And my experience with it is this. The metric that I'm then describing, available is a word, but I guess user experience might be a related word.

BRAD VERD: I think what you're describing, to me, the available is do they get an answer, one answer. That's the available metric that you have. I think what you're describing the health of the system.

FRED BAKER: Okay, and maybe health is the right word, I don't know.

BRAD VERD: As an ops person, I want to know exactly what you're describing, because I want to know the health of my system. And when there's a network problem, I want to be able to go diagnose it, work with that carrier, and fix it.

FRED BAKER: Right, okay, but now I also find myself thinking about political questions. China, Russia, certain countries in Asia, have talked about, gee, I'd like to just shut off everything in my border, and it's not just China, there's a list of them, and frankly, China makes a wonderful example between the government, the firewall, and Yeti, wow, that's a wonderful example, but Russia is building the same thing. And who is to say that other countries are not? So if I were to discover that I only had access to A, because the firewall shut it all off, I'd actually like to know that, that's something I need to understand. And the fact that I can get an answer from A is interesting, but not I think ultimately all that informative. Wonderful as A is.

BRAD VERD: I don't disagree, but that's a separate discussion than available of the service. I 1000% using [inaudible] agree.

FRED BAKER: Okay, well, and Wes is a wonderful mathematician. I don't know.

JEFF OSBORN: Let me ask a hypothetical situation. You've got 100 probes and for now let's ignore V4, V6, let's say measuring 13 servers, some number of times an interval. So let's say one of those probes can only reach half the servers, and all the others can reach all of them. So, who do we use that to characterize the availability of the root server system?

FRED BAKER: From the perspective of that probe, and I think this is according to an observer in a location, topological or whatever. From the perspective of that probe, the service is not all that available, where from the perspective of other probes, it is. So now can I come back to the Board and say the service is available? Well, sort of. But as an engineer, I'm trying to manage the service, I think I've got to identify the location that probe is in and go diagnose it.

DUANE WESSELS: But I see our task here is to define a way to characterize the system with a simple metric, not, we're not going to say Probe A was this, and Probe B was that. We want to say the system, the service was this, it was 98% available or it was available or unavailable. So I don't think we want to get to the point where we're reporting metrics on individual probes, we have to aggregate all those probes together somehow, right? So that's one of the questions here, how do you aggregate all the probe measurements over a time period to come up with the characterization of availability for the system?

KAVEH RANJBAR: So, I think, I disagree with you, Fred. When you're talking about the system, we need to think of the system as a whole, or service, sorry, as a whole. A good example would be like let's say GSM a team of ours, I'm sure I'm connected to at least three or four towers right now. I don't care if two of them go down or disconnected because I still have my service, and phone might have issues, so I cannot get connected to T-Mobile Network, but still T-Mobile is up and running, correct?

So we want to see if T-Mobile Network US wide is working or not. And of course at any given time there are towers which are failing, there are phones which are failing, but that doesn't matter. You have to say okay, the system is healthy, up and running, the network, T-Mobile Network in the US. And of course there are issues, but in general, it's working, or not.

And that's why I agree with Duane, that yes, we have to figure out how to combine this data and say okay, from our point of view, the system is working. I don't care if one state now says oh, you cannot turn on your mobile phones because of whatever, that would be your China example. The system is available and working, and we have to be able to measure that and understand if the network or system, or service, sorry, is up and running.

DUANE WESSELS:

I was just going to say, to extend Kaveh's analogy, what would be interesting, though, is imagine after a hurricane or natural disaster to know how many of T-Mobile's towers are down. One could argue the whole system is still up, but certainly from some people's perspective, T-Mobile is down, and you can get an idea of where the overall health of T-Mobile's network by knowing, okay, 18% of their towers are not up and running.

So, could I suggest, I'm just going to put a straw man out there, that we measure V4 and V6 separately and again, from the probes' perspective, if you can get to one of the 13 V4 addresses, the service is up. Likewise for the 13 V6 addresses, it's either up on V4, up on V6, and then the

total number of probes, it's either up or down, that's how we characterize the availability of the overall system, but we have to be very careful that we're talking about it from the perspective of the probes, so that we don't get this idea of the entire system is somehow 90% down or 90% up, or something, that's pretty dangerous. But I'll put that out there as a straw man way of thinking about this.

BRAD VERD:

I'll look to Matt and Duane, both Matt's, because Matt, you were there when we had many of these discussions with VeriSign. We have a lot of monitors, tools that monitor the system, think of them as probes monitoring the system, and then we have what we call our real-time monitors, and we don't alert off of our real-time monitors, our real-time monitors are kind of real-time, you could watch them. They showed, there was like this delineation between alerting system and health system. And the real-time was really like real-time health, what's going on, which I think is really what you're after. You want to know the health of the system, I want to know it too, but the health of the system, if one piece of it isn't necessarily working, it doesn't represent an outage of the service, because we are redundant and resilient as we described.

UNKNOWN SPEAKER:

And to exactly that point, I'm wondering whether we shouldn't be describing this instead of as we're 97% healthy, as we're available from 97% of locations measured.

DUANE WESSELS: That's what I meant about having to characterize this very carefully.

BRAD VERD: I still think what you describe needs to somewhere, if not in the metrics party, another effort somewhere. It feels like it could fall in here and be maybe the monitor and the metrics could be defined here, and then the data collection and use of that data to understand the health of the system, could be figured out later.

FRED BAKER: Which comes back to my previous comments about the distinction between the metric and the use case.

UNKNOWN SPEAKER: Well, I think, because Brad is of course one of our shepherds, and as I read the charter, although health of the system is not in any way specifically mentioned, conceptually the idea is very, very important and I think easily fits within the work party. So, I think we at least need to add it to our parking lot and see what we can come up with, with respect to this, and it might be a recommendation that another work party be formed to look specifically at it, but yeah. And do you agree, Fred, with Brad's characterization that it's really a 'health of the system' type of thing that you're describing?

FRED BAKER: I'm willing to see those words.

UNKNOWN SPEAKER: Okay. So, we've also had the suggestion from Matt and Jeff restated it slightly differently, about doing the measurements divided by our major protocols, V6 and V4, and if the probe doing the measurements receives an answer from the respective protocols, that it will be considered a success and the system is available if they get any answer to a query.

JEFF OSBORN: At least one has to respond.

DUANE WESSELS: One would imagine it would have to be pretty catastrophic of a situation for the service from that perspective, from any perspective to be down.

JEFF OSBORN: Which technically sounds silly, but I think politically really sends a message. It's robust enough that the system is working in that catastrophic situation.

PAUL HOFFMANN: I think we should be really explicit on what constitutes an answer. Are [inaudible] or refused answers that qualify?

DUANE WESSELS: I think we can address the other work party, but my initial reaction is no, they're not. You would have some sanity check that you're getting an answer that you expect.

WES HARDAKER: Wouldn't that be available, but not correct?

UNKNOWN SPEAKER: We definitely are delving into the 'how do the metrics fit together' space here, with this discussion and it's a very important discussion, do we want to go there right now, Duane, or do we want to leave it until we get further through the metrics for the RSS?

DUANE WESSELS: I don't know, I guess the question is if we that's something we want to tackle here today in this group or for the work party later. To me, that level of detail seems like more work party work, not the best use of this time, but I can be convinced otherwise.

UNKNOWN SPEAKER: I don't think we need to discuss it right now.

UNKNOWN SPEAKER: Okay, let me suggest we go on. If we do have time at the end of this session, maybe we can come back to it and discuss it a little more. One of the items that has come up in the work party discussion particularly relative to this is does it make, that a reasonable approach to keep the

RSS, the RSS and RSO equivalent metrics fully separate or do we want to, for instance for availability, look at combining or evaluating the RSO availability information and mathematically determining the service availability, as opposed to having a separate and explicit metric inquiries and responses for the probes. Brad?

BRAD VERD:

Just for clarification, you're asking if we should ask, if a probe should ask the same question for each different service, is that what you're asking? One for the RSOs, same question for the service, correct? Or should we derive them from one query? Is that correct?

UNKNOWN SPEAKER:

The RSS availability question could be answered by either of those means, and the question I'm reading at this point is does this group have views on which would be the better way to approach this?

BRAD VERD:

My answer would be that there is no point in throwing away resources, the cycles, and doing the query again to get an answer that you already have.

UNKNOWN SPEAKER:

Paul, go ahead, Hoffmann.

PAUL HOFFMAN:

I'll go further than what Brad just said. If the metric, if the measurement that we are getting here is based on individual RSO metrics, then we are holding those individual RSO metrics. It would actually look like we weren't using those if we probed again. So I think that on any of these RSS questions that we have, if they are based on either in this case it's not mathematical, it's logical, did any of them, you know, raise their hand, or a mathematical one of latency, we should absolutely point to the individual RSO metrics that we collected and say we made this determination today based on those metrics.

UNKNOWN SPEAKER:

So, one of the themes I'm hearing here that we heard earlier today is that in fact making reuse or multiple use of a given query or set of queries that were collected for one particular metric, into determining another metric is the general preference, rather than totally separating the metric measurements and as a result having more queries, and so forth. I think that's what we've heard each time we've had a similar discussion. So, I think that's very good guidance, and probably what I think we will have to figure out is what are the right words to get down to describe that, but having the method or the desire to combine and multiple use your data is what we need to figure out how to get written down here.

DUANE WESSELS:

I think there's another point that's worth making, just so everyone has all the options on the table. Can you scroll down a little bit, Steve? There's some text in this document that describes a way you might do

these measurements differently, not using queries directly to individual operators, but instead you have like a local recursive resolver installed, you could just ask it your question and then let it do its job of deciding which root server to talk to.

So, for example, that's kind of how the ripe atlas measurements that I showed this morning, that's how it works. So I'd welcome input on, well let me back up, I think this approach has both advantages and disadvantages to doing the probes individually. One is that it can be a little bit simpler, because the recursive name server does a lot of the hard stuff, including the server selection and the timeouts and the retransmissions, but it also makes things a little bit complicated, because then you have to say well which actual software am I going to use, how is it configured, how is it maintained, and that sort of thing. But I just wanted to put this out there as an option, and if people have thoughts or opinions or preferences about this versus the other, now would be the time to say them. Brad?

BRAD VERD:

I think both are good, and if we can do both, we should do both. And I think it would be nice, I mean, if I was writing the paper I'd put both options down and say the benefits of both. Having both gives you an option to audit each other, to see if you're sane, and it's a question of the LOE, right? The level of effort to do one or the other. If we're already collecting the RSO metrics creating that logic to say did we get an answer, that seems a little LOE. If this is a little LOE or just a little bit more, it seems like a double check would be great. Because then you're

truly representing the client experience of getting an answer from the service by going through a resolver. That's good.

DUANE WESSELS: That's good feedback, thank you. Anybody else?

UNKNOWN SPEAKER: Regardless of which of these methods may be used, the measurements interval is listed as something we're looking for input on, but I would suggest as a starting point there is no need to use any different interval than what the RSO equivalent metric is. Does anyone see anything different than that? Okay, that's good. We've already decided we're not going to try to decide what 'good' looks like today.

DUANE WESSELS: Maybe tomorrow.

UNKNOWN SPEAKER: Maybe tomorrow, yesterday. Because what 'good' looks like involves really all aspects of the metrics including the thresholds, once you get all the numeric values that you can expect. How far does one go before you decide it's not good enough. Personally I don't think that we're close enough now to knowing what the numbers are to be able to say what's good enough and what's not good enough. Do you disagree, Brad?

BRAD VERD: When is the discussion, then?

UNKNOWN SPEAKER: I think the discussion is after we get the paper, or the work more fleshed out in the work party, both the incorporation of the ideas that have been discussed here today in this meeting, yesterday, today, as well as the next step of getting the numbers wrapped into and getting agreement on how we want to treat the distance from the center point or the distance from perfect. Is it going to be some percent, will it be different for each one?

I think it will tend to be different for each one. So I see that as much closer to the end of the work party effort, which our charter says August, which I think we'll certainly try to achieve that, I don't know that it's achievable, but I would say when we hit, just to pick a number, 3/4 of the way through the definition of the work and I'd say now we're about a third to half at this point. Go ahead, Wes.

WES HARDAKER: Just as a reminder, my proposal yesterday was that we get through the individual metrics, and then we go back and discuss what 'good' is based on what we finally got through defining. In part because I think that when we get to the numbers, we will possibly go back and need to reflect some of our definitions based on it. But it's a question, it's a chicken and egg problem, so rather than boiling the ocean all at once, one bucket at a time is sort of what I was shooting for.

JEFF OSBORN: I don't know, Brad looks frustrated, please...

BRAD VERD: I'm frustrated because your proposal yesterday turned into, "we're going to talk about thresholds at a later date," from yesterday until today. So that's why I'm frustrated.

WES HARDAKER: The way I tried to word it was after we got done with this.

BRAD VERD: This is not a result of what you said, I'm just saying that your proposal yesterday, heard by others, interpreted, is that we're going to get three quarters of the way through the metrics work party before we talk about thresholds, so we're going to build it and they will come. So, we're going to build it first and then we'll talk about thresholds. This is frustrating to me.

JEFF OSBORN: I'm just trying to understand why three-quarters of this work has to exist before you can start talking about what's good. And the thing that was thrown out, I think it was Brad, that I thought was really interesting, is the performance we're currently providing to the world at great effort, over a great number of years, good? And I think Brad was saying something aspirational about how much better we could be, but you know, I'll stick a stick in the ground and say the current system as it's operating right now is good. We haven't had a global outage in forever,

the current system as it works is good. So whatever the measurements come out to, the measurements come out.

But you know, let's start from somewhere and then go from there. If it's not good, why isn't it good? Is it not good because 8 terabytes sustained 4-hour attack would leave us in wreckage, or is it not good because has been stated in a couple ICANN generated documents, the growth is outrunning the RSS's ability to grow to suit it? You know, there's some really dismal stuff that gets written off handedly by people who aren't in this room when describing this group, and not many people talk about it. So straw-manning somewhere, is this good, is this inadequate, is this too big? Do we need to scale back?

UNKNOWN SPEAKER:

Well, I guess let me say a little more about the threshold being later on in the work. Two points that I think certainly, Duane, the support folks have talked about, and that is that when we get to the point of saying how good is good enough, when we're basing it on the numbers, the numbers will be determined by how you define a measurement, what is the metric, what are the parameters around that metric, and at least it's my perception that the definition of those things will be itself quite important in being able to then say how good is good enough, how tall do you have to be, is it 32 or is it 958, I just pick numbers out of the air there, because we don't have numbers at this point.

Now I'd like to be able to figure out how to do it sooner, and if what maybe, it was really driven by a later definition of how tall is tall enough, or good enough, came more from what we've been looking at

for the work party leadership than Wes' suggestion yesterday. But if there is a way to figure it out before we get to the numbers, I'm perfectly willing to do that. Fred?

FRED BAKER:

Well, this is a comment on numbers and thresholds. But it seems like we should at least characterize what we have before we decide what numbers somebody else needs to match. I'm not sure we'd necessarily match them ourselves if we haven't asked the question. So it seems like for thresholds and such we need some data.

UNKNOWN SPEAKER:

Okay, so one of the things also that I wanted to just sort of bring up is we, the RSAC and the RSAC Caucus have the opportunity to really treat this work, this effort as the training wheels for what eventually will get used. We can create a lot of these things ourselves. We can publish them right away, but if we choose to, we can go off and do at least some prototype implementation of the sensors, do some checks on the metrics, and we might want to do that prior to officially publishing a metrics document itself, so we can make sure that the things we produce have at least a degree of sanity to them, they aren't totally out in left field.

UNKNOWN SPEAKER:

So, that's completely novel, we become informed, and then inform other people? You've got a question.

HIRO HOTTA: Thank you, I just want to seek clarification and to also make a suggestion. I'm thinking I want to tend to agree with Wes in terms of the measurements, the threshold would probably have to be defined later, because the question I've been asking myself which from the temperature filling the room is solve these metrics and not actually [inaudible]. So I'm thinking until when we can answer the question, is [inaudible] able to measure all of these metrics before the threshold can now come in? That's my thinking.

UNKNOWN SPEAKER: It sounds like another advocate for let's get data before we actually publish something.

BRAD VERD: Again, that choice of words is really frustrating to me. I am not talking about publishing something before we talk about thresholds, okay? I'm not, that's what you just said, though. But, I am fine, I am fine talking about thresholds at a later date. I feel that it is another elephant in the room, and I stated in the admin meetings, I've stated actually in previous meetings that we had in Barcelona and wherever we were last, Kobe, where I said I think identifying what the metrics are and what the measurements are and the intervals are, is the easy part. I think this group is going to thrash when we talk thresholds, and we haven't gotten there yet. So I'm just going to point that out.

WES HARDAKER: So, I've been trying not to agree with you, but now I have to. And the reason why is that I didn't want to trump the Chair's plan for walking through this conversation. But the reality is when I said yesterday we should get through the metrics first, that included even before trying to figure out how to do it at the global service level. So I'm actually all for upward stepping right now, but there is an agenda and order in plan. We've met the threshold that I tried to say and failed yesterday.

UNKNOWN SPEAKER: So, with respect to this metric, I think we have covered everything except what 'good' looks like. Does anyone have further input on the RSS metric that they want to bring up before we move on? Jeff?

JEFF OSBORN: Did we determine anything or did we just have a discussion?

UNKNOWN SPEAKER: No, we determined that from a system perspective, answer for V4 and answer for V6 would be marked as available, the system is available with respect to that. And the query time will essentially align with the RSO query frequency, and so forth.

JEFF OSBORN: TCP and UDP?

UNKNOWN SPEAKER: Well, we didn't actually cover that. We were grouping it with the RSO, whatever we decided for the RSO would translate into here for the RSS.

DUANE WESSELS: Right, so what we have on the individual RSO measurement, is we have all four, all combinations. So I would propagate here as well. I think if you're going to split it by V4 and V6, you should also split by TCP and UDP.

UNKNOWN SPEAKER: Yeah, you're right, Duane. It was something that my head had done the carry forward from the RSO and just had not said it like I should have.

DUANE WESSELS: I don't know, do we want to break early, we've got 8 minutes, or should we just plunge ahead? Okay, we'll take a break, about 15 minutes, and come back around 2:35, or so.

Okay, thank you. So, welcome back everyone to the second afternoon session on metrics. What we'd like to talk about next is a correctness metric for the root service. You'll probably notice a lot of similarity between the last discussion where for example the first proposal here is to use the measurements from individual RSOs to then calculate a corrective measurement for the root service itself. A lot of same questions and caveats apply here. There is also as you can see here, there sort of a proposal for doing it based on the individual measurements and also for doing it based on using a local recursive.

And per Brad's suggestion, I think we're happy to keep both of these in and use them as a check on each other.

The text here is I guess unfortunately very specific to DNSSEC, that's what we had talked about previously and I know Paul Hoffmann made the point on behalf of someone else that we should also have a correctness check for data that can't be signed, for unsigned data, so a check just on comparison and I'd be happy to do that for this case, as well, we can add that text later. Let's see, not sure where to go with the discussion, exactly. So, in this picture there is no long list of PDB items, because I think a lot of those would be the same as some of the other metrics. So it would make sense to use the same measurement interval for the RSS as we use for the RSOs.

Probably the only, maybe the first thing we should talk about is for the last metric we had a proposal from Matt Larson on a way you can get to a number where you take all the results and sort of assemble them and analyze them, and then from those you do a single result that says some percent of probes received correct responses from the server system. Does anyone see glaring issues or points they want to bring up with respect to this metric at this time?

BRAD VERD:

Certainly no glaring issues. Isn't there a way to check the correctness without what was described earlier, the exactness of the "crawling the zone," basically, with that digest or that hash that is being considered in the [inaudible] IETF type of thing? Maybe this isn't the right place, I just...

DUANE WESSELS:

Yeah, so what Brad is referring to, I think, is actually in an address that I'm coauthoring with Wes and Warren, which is adding a hash over the contents of the zone, so that would allow you to say that the zone that I've received is unmodified and whole, and correct. I can see that being used by whatever thing, whether it's a probe or a collection point, one of those things will have to have a current zone file that it can do comparisons against.

So, that would certainly be useful for those. It wouldn't be useful for like an individual DNS query response. It doesn't provide the right level of coverage for that, but I do think that would be useful and maybe we could even put that in our draft as a use case for that feature. Yeah, unfortunately, I think our only two choices are DNSSEC validation or strict comparison of response data. So, if everyone sort of okay with mimicking this metric based on what we discussed for availability where you say, you know, let me, I should be a little more specific.

So, here's an example case. You've got a probe doing measurements to end servers, maybe it's 13, 26, whatever. Let's say most of them give correct responses, but one of them gives an incorrect response, for some definition of incorrect, whether it's DNSSEC or not. So, what would you say about that probe at that time? Did that probe get correct data from the root server system, or did it get incorrect data from the root server system?

I think you could argue that it got correct data, because the way we know, again, this implementation dependent, but a lot of

implementations will retry in the event of a failure. They will retry their query until they get a correct answer. They will treat is in DNSSEC validation as a sort of temporary failure and retry. Matt, I think I interrupted you right when you were ready to...

MATT LARSON: I was just going to ask if you could repeat your proposal, I didn't quite get it.

DUANE WESSELS: Well, I'm not so much on a proposal right now, but a question. So, let's say we've got a single probe and at time T it's going to send out queries to all the root servers and it's going to look at their responses and make a judgment as to whether those responses are correct. But then based on the selection of responses, it needs to make a statement about whether the service was correct.

MATT LARSON: So, you're assuming just as you're explaining this, that you would go one record at a time through the canonical root zone and query all?

DUANE WESSELS: No, not exactly. I'm saying that, let's just keep it simple, let's say it's going to do an SOA query to all 13 root servers; 12 of them come back correct and one of them comes back incorrect. So what did that probe get from the service?

MATT LARSON: Well, I would think any incorrect data is a pretty big deal. This doesn't seem like something that should be aggregated up, but something that should be individually reported.

DUANE WESSELS: Okay, so in that case, one incorrect response means the probe got an incorrect from the service. Then we could say something about the percent of probes that got incorrect responses.

LARS-JOHAN LIMAN: I would argue that it got 85% or whatever the number is, correct responses, and that's bad. We should react to that, I fully agree with Matt there, but we cannot, when you have multiple entities involved, you cannot report on or off. You have to report a percentage. We have to work with that higher up in the system.

DUANE WESSELS: Well, didn't we just do exactly that in the previous metric, where we said if you've got a probe querying 13 or 26 servers, and if it gets a response from at least one of them, then the root server system was available to that probe at that time. That was Matt's proposal.

MATT LARSON: My gut reaction is that bad data is in another category of worseness, beyond, that's the best choice of words. Bad data is worse than data, so

it seems to me that it should be treated, I don't know what that way is, but that's just my gut reaction reacting to what I'm hearing.

UNKNOWN SPEAKER: So let me do one other slight modification to Duane's scenario there, and if the response that it got from 12 of the RSOs was the proper, and it's an SOA base, so proper SOA, and the 1 that it got on the 13th RSO was the one back. Is that, how do we treat that? I that bad, is that incorrect data?

DUANE WESSELS: I'd suggest that's our staleness value. What I'm reacting to is like a root server instance literally giving out bad data, something that's not in the canonical zone or canonical -1 or -2, or however far back we want to go. I guess the wrong data or fails DNSSEC validation, like something that was never in any root zone.

UNKNOWN SPEAKER: Is not or should not be in the content of the zone, which has an inference to it that the probes will have to pull the full set of root zone data from each of the servers that they're sending queries to.

MATT LARSON: As we talked earlier, you could do this multiple ways, but I'm just talking at a conceptual level. I think you need to note somehow, somewhere you need to make the comparison between what's coming back from the instance the contents of the current root zone, and maybe N -1 or 2,

so that you know a little bit of past history, so that you don't false alarm on that for somebody who is stale, or maybe if somebody is stale, maybe you figure out first that they're not stale, and if they're stale, that's separate issue, but they're not stale, then you could imagine, as we talked earlier, doing that either by, the probe does know the current zone or you ship it back to somebody who does know.

UNKNOWN SPEAKER:

Part of what I am trying to run through in my mind in this scenario is, are we in this metric inferring that somehow when a probe is doing the correctness metric, whether it's against a singular ISO system that it, in so manner has to get all of the route zone data, from that server to itself. Somebody else may analysis it, it may be shipped off elsewhere, but to actually measure this metric, you have to get all the route zone data back to a probe, somehow. To judge correctness, if we are going to have exact correctness.

DUANE WESSELS:

I don't think you have to get the route zone data to a probe, I think you have to get the route zone -- you have to match them up, they don't have to be on a probe.

UNKNOWN SPEAKER:

But if the probe is what's sending the query and getting no response, even if it ships bad information somewhere, it's still, I think, going to have to have all the data come back to it, instead of going elsewhere,

but if you send a query and direct the response sample that is not a standard DNS query.

RUSS MUNDY:

I think this is a matter of are you going to sample specific records or you're actually going to validate the correctness of the entire zone, every record.

DUANE WESSELS:

I would not propose to validate the correctness of the entire zone. I would do some sort of sampling. If it were up to me, I would do a number of queries per interval, three, four, five, something like that. Probably, I would do a SOI query, and some other randomly chosen names, maybe even responses that should be, like delegations, and maybe something like that, that's how I would approach it. So Russ, it sounds like you and I should talk later.

UNKNOWN SPEAKER:

We also I think have and I don't remember who all raised it, I know Paul did, but other people did too, who as far as the correctness and for the full zone data and if that is the goal for the work party and it is going to be based on a sampling set of queries versus getting the entirety of the zone back to a probe, from a given -- from the server it's querying, I'd like to hear from the other people, especially Paul and others on the phone that might want to speak up about this.

UNKNOWN SPEAKER: Paul and then Daniel. Paul, go ahead.

PAUL HOFFMANN: Okay, so I've had my hand up for a while, not about this, but, Russ I never said that I wanted it for the full zone, I think that's silly. I think sampling is just fine. But related to what Matt and a few others have said about incorrectness for the RSS, is important. I think that when, if there is ever a metric, not a measurement, but a metric that says less than a hundred percent of the route zone was correct during a particular interval, I think it is incumbent on us to say what was incorrect. Was it an NS record? Was it an A record? Was it something else? That is that the community will want to know, not just a percentage of correctness, but was it aimed at something?

And going back to the troop, I unfortunately agree with it about States supported actors, countries, regions, whatever, trying to inject that responses, it will be useful to see if a bunch of probes in a region, where getting a wrong answer at a certain time, of a certain thing. So I believe that this metric, that in reporting it, we will have to go to a level of detail, if we want it to be useful to anybody, that we will have to go to a level of detail, not in the other metrics.

UNKNOWN SPEAKER: Thank you, Paul. Daniel.

DANIEL MIGAULT: Okay, so I might then be the one proposing something that is stupid, but the problem I see when we are assembling what we are testing is that,

is if someone is serving a corrupted zone, if not probably take a random name into that zone and changing that value. I think most probably it's going to take a very specific name, because it has a target. I mean, that is targets orientated so, I think if we are not doing the full zone, we won't be able to say the full zone is being served correctly. So what I might agree is that if we are sampling, we should physically say exactly what we are doing, which name we are taking, and maybe we don't need to choose that name randomly, but, maybe the only answer will be enough then.

UNKNOWN SPEAKER: Thank you.

DANIEL MIGAULT: It's physically having a CRC vs a signature, I think.

DUANE WESSELS: Thank you, Daniel. I want to reiterate something that I just heard Paul say. He was talking about state actors and that kind of thing. I think the concern here is that you may have something in between a probe and an authoritative route server that is a source of bad responses, and I guess my question to the group is, do we want to consider those potential bad actors as something that we need to measure? Are they part of the route server system to be measured? Or not, should we attempt to somehow identify when those sorts of things are happening and discard those cases or those measurements?

RUSS MUNDY: So they definitely exist in the ATLAS data. Definitely, we should be discarding them. We don't always know we may get a correct response, correct data from a middle box or whatever you want to call it, that could be making the latency better, making latency worse. Maybe not the job of this group to sort of police those things, if the data can be used elsewhere, fine but maybe out of scope for our purpose.

DUANE WESSELS: Sorry, I think I heard you say two different things. First I heard you say they should be discarded, but then you said out of scope.

RUSS MUNDY: Discarded for our matrix and determining latency or determining correctness, if they can be identified. But as far as keeping them around to identify bad actors, do we really want to cross that boundary and get into that business?

DUANE WESSELS: I think this is an important topic and I want to make sure everyone's understanding and we're in agreement, because I think it affects the metric quite a bit. You know, like Ken said, when I look at the RIPE ATLAS state, I see this as well. Unfortunately in the right ballad state it is kind of easy to identify the cases where there's something other than an actual route server sending responses. They are intercepting and you can identify that.

So the work party could be task with, one of the things is probes and the system has to do its best to identify data, but that doesn't come from a

legitimate route server. Or we can say, well that's the reality we live in, and those are things we want to know about. Those are things we want to measure when those are happening. Maybe the answer is both, I don't know, but I think it is an important point.

LARS-JOHAN LIMAN:

Half joking, half serious, suggesting using six zero to secure an alternative connection to the server. It just popped up in my mind, I haven't really thought this through, but if someone is willing to help me to take it to either end saying it won't work or yes, it is doable, then it might be a way forward, I don't know.

UNKNOWN SPEAKER:

This is a tricky problem, but it would be nice to check the catastrophic situation, which is hopefully very, very unlikely, of the actual data itself, on a router and set them change by that actor or, just wrong for whatever reason, right. But with the network path in between the monitor and the router incidence, cause you never know. I go back to what I said earlier, it seems to me the worst possible thing that we could all be doing is serving bad data.

DUANE WESSELS:

Okay, so I think that's a good point. What I hear you saying, it would be good if the metric or the work party could do some work. Think of ways to identify, data, bad data on the server versus bad data in the past that they might be getting in.

UNKNOWN SPEAKER: Paul Hoffmann, then Daniel. Paul.

PAUL HOFFMANN: So this is pretty much why I said that we need to report what the data is, what that data is, is so that either we or someone else can say that didn't come from that route server, or in this case, that route server instance, because the probes is only talking to one instance, that came from an intermediary or that is likely to have come from an intermediary. If there is a way to determine that early, great and we can say N percent of records, which we know are all from that intermediary, were incorrect.

But this really goes back to what Brad pounded on this morning or, I'm sorry, right after lunch, about -- that we were talking about the route server system or service or whatever. I believe we're talking about the service. The service, we have not got a direct connection with the customers and the service, and given that we are going to see some of these things. In the same way, there might be a middle box that adds a butt load of latency. That's still is part of what a customer is going to see.

UNKNOWN SPEAKER: Thank you, Paul. Daniel.

DANIEL MIGAULT: Thanks, unfortunately I couldn't hear some of this, what Matt just mentioned but, do we have an agreement that, what I wanted to say is that maybe an instance can serve corrupted data, if we agree it's out of

scope, then I am fine with that, if we reach such assumptions. But on the other hand, it's true and I think was something that I tried to explain also in some of the comments, is that when everything is working fine, you can state that the routes of the system is working well, but when you have an error or whatever, it doesn't necessarily mean it comes from the end points.

So everything can be involved inside so it can be a middle box. In the case of corrupted data, it is hard to detail if it comes from the middle box or the end point so it really needs some further investigation, and as Paul just mentioned, at that level it's important to mention what exactly is being measured, and specifically. Thank you.

DUANE WESSELS:

Any last comments about the correctness metric? I think not, in the interest in time we will maybe move on to one of the other ones, so we can get through them all. Okay, guess not.

Right, the topic here is RSS Metrics of latency. So, again, a couple of proposals here. One is to calculate a latency or somehow characterize a latency for the RSS based on the individual RSO measurements. What it says here is you could assemble individual measurements and then calculate the median or maybe other percentiles. We've talked about that quite a bit as well. Qaurtiles, other percentiles, what have you.

Alternatively, or in addition, you could do the same thing by sending a query to a local recursive resolver which then does its usual thing, chooses which server it wants to talk to at that time, over which transport and you can measure the latency from that recursive resolver.

I would suggest all of the latency measurements from individual probes be then assimilated and reported as [inaudible] or other percentile which sort of leaves us in the situation that we're looking at medians of medians; which, hopefully that's okay. Hopefully that's still a valid thing to do, but I think that's here we're at. It's not that [inaudible] question but again, for this metric I would probably mimic what we do for the individual ones which is report these separately for V4V6 and UDP TCP and that's the gist of it. Any questions?

UNKNOWN SPEAKER:

Let me ask, Brad, I think that your comment you made earlier about the two types of measurements could be done using the local recurrence resolver or using a probe type of device. If Brads here, we should try to do the metric so they could be done either way.

BRAD VERD:

Yeah, I would do the [inaudible].

DUANE WESSELS:

I think that would be very interesting because you could see if the resolver base measurement is better or worse than the medium of the individuals, taken together, it would be very interesting to know.

UNKNOWN SPEAKER:

And folks are generally agreeable with the same before V6 [inaudible] division so same set of end points on everything.

FRED BAKER: What is the argument for separating between V4 and V6? Is that to detect places that turn V4 off?

DUANE WESSELS: That wasn't my argument but I think the rationale, my rationale for having them separate is that it saves you from having to choose which one. If you're just going to do one of them, you know, you have choice which one.

FRED BAKER: Well, which comes back to my thing, you've got a list of addresses and you go after them. The thing is that we have networks that are starting to turn V4 off and we certainly have a lot of networks around that only run V4. It just seems like an odd choice.

DUANE WESSELS: An odd choice to turn off before or an odd choice to measure these separately?

FRED BAKER: An odd choice to worry about them as a guiding principal. Ten years ago it might have made sense. I don't think it makes sense now.

DUANE WESSELS: So, let's revert for a minute back to the individual RSO case, because that is sort of where a lot of these measurements derive from. So, if I'm a probe and I need to measure latency to a root server, how do I choose my transport?

FRED BAKER: If you're going through the T-Mobile Network, you only have V6.

DUANE WESSELS: Okay. But -- so, are you saying I -- but I have to choose a transport over which to make my query. Maybe I try V6 and it doesn't work, then I try V4 and it doesn't work, and I try the other. But, I have to choose one to make my query. So, how do I make that choice?

FRED BAKER: You also have to choose -- and you've got 26 addresses to choose from.

DUANE WESSELS: I'm measuring just one. I mean --

FRED BAKER: But you're picking one.

DUANE WESSELS: I'm a probe, and my goal is to measure the latency to [inaudible].

FRED BAKER: To something. Yeah.

DUANE WESSELS: To [inaudible]. Okay. So, I got to make a choice.

FRED BAKER: But you have to make a choice among 26 addresses, not among two.

DUANE WESSELS: No, I'm saying I'm a probe, and I've been told, "Measure the latency to a root."

LARS-JOHAN LIMAN: Why aren't you then told, "You're supposed to measure a root of a TTP?"

DUANE WESSELS: That's fine.

LARS-JOHAN LIMAN: It's just a matter of why are you make the decision. Either in the instruction to the probe, or by some random process in the probe.

DUANE WESSELS: Okay. So, I'll try again. Somebody has to decide, the transport --

LARS-JOHAN LIMAN: A decision has to be made.

DUANE WESSELS: Yeah, right. So, I think the work that needs guidance is how those decisions gets made.

UNKNOWN SPEAKER: So, let me just add one slightly different factor to V4, V6. In almost any set of graphing of results of measurement data you'll see where they separate V4 and V6. You'll see differences between V4 and V6. And, at least in my mind knowing that there are differences and being able to observe what changes over time in terms of those differences, historically, the kinds of things we were seeing were almost identical between V4 and V6. I'd say don't bother to separate that way, just sort of do the coin flip, "Which one am I gonna use this time. If I'm going to send a query to [inaudible] root or whatever." But since there has been over time, differences shown up consistently. I think that's valuable information to have.

FRED BAKER: And to me it's an argument for measuring both. And especially knowing that there are networks that are turning V4 off. Seems like that's going to be something to track.

DUANE WESSELS: I think we are proposing -- do propose to measure both. And over UDP and TCP as well.

FRED BAKER: Okay, I'll resist the urge to quote what you've said two minutes ago about having to pick one.

DUANE WESSELS: I think that we must be misunderstanding each other because what I'm proposing is that -- Earlier I called them sub-metrics. So, under latency you would have four sub-metrics. One for each permutation of transports. And I guess I misunderstood because I thought earlier I heard you say something like, "Why are you measuring during both TCP and UDP?" So, I misunderstood, I'm sorry.

UNKNOWN SPEAKER: So I think we're under violent agreement misunderstood that we were I think.

DUANE WESSELS: Any more dead horses to beat over latency? There's probably a lot of repetition here, so this is RSS Metrics [inaudible]. As before, the gist of it is that you can do SOA serial queries to individual rot server operators and then collate the results and come up with some metric on staleness for the service, the root service system. I'm going to have to read this to refresh my memory on what -- Let's see. So, "At TBD Intervals collect the measurements from individual RSOs from our probes, aggregate the results to calculate the percentage of probes who observed SOA serial-based staleness."

So this is maybe a little bit simpler than sort of where we have been going with some of the previous ones where again going back to 'correctness'. For example, I believe we agreed that if a probe receives at least one incorrect response then you say that that probe has incorrect data at that time. And in the availability one it was sort of similar, that if the probe got at least one response from a route server then it had availability. Should we do something similar in this case so that, for example, if a probe observed staleness from at least one root server operator, do we say that that probe was experiencing staleness at that time?

I'm not sure that staleness is as serious as correctness, so maybe that's too strong of a -- maybe that is kind of how stale it is, yeah.

UNKNOWN SPEAKER:

We had also a good bit of input on the RSO staleness metric about how to reinitiate the query in a given timeframe tied to the SOA. And that just being out of date by just the previous SOA is something that could be accommodated in the metric itself, I think. So, probably a similar approach here. One thing that- it may have been discussed, but I am not remembering it at the moment, is: 'what about a SOA response that's badly out of whack?' That's back say three days, five days. Something like that. [CROSSTALK] Or ahead, yeah. Have I forgotten an earlier discussion, or did we not cover that?

DUANE WESSELS:

No, we didn't really talk about that as far as I remember either. I think being ahead would be initially the case. We should definitely describe

what to do in that case. Being behind by, well -- In my mind there's some range of being stale, which is sort of okay. It's not, you know, that detrimental. Until you get to, I think, seven days is sort of the point at which [inaudible] zones start to expire from secondary [inaudible] instead. That's a pretty serious cut-off right there.

UNKNOWN SPEAKER:

Putting on my route-zone operator, maintainer hat. There have been emergency route-zone changes lately. And, you know, when that happens there's -- Even if you're one SOA or serial behind, you know, if there was an emergency route-zone change and the recursive server goes to the root server that -- or an instance set has an old copy that [inaudible] purposes of supporting an emergency change. I would argue that [inaudible] is up to date, short of some threshold that we agreed upon. I don't know, just thinking out loud here.

DUANE WESSELS:

Yeah, in the earlier discussion I -- Were you here for the earlier discussion about [inaudible] or were you out of the room then? Yeah, so we had a very good discussion where we sort of settled on using the SOA refresh perimeters as sort of our cut-off for saying you're stale or not stale. So that we can probably get that here. But again, I think one thing we need a little bit of direction on here is how do we take the individual measures and use them to make a measurement or statement about the system itself in terms of staleness.

UNKNOWN SPEAKER: Well, I think an aspect of that also is if it's just -- so if it's one off, a new one has come out and you just -- one behind that there is parameter for handling that but the further away question which Matt was just giving us with the case study as to why -- do we want to keep it simple? Which I think is my personal preference -- and just say if it isn't in that sort of turning over phase of things, that it just is marked as bad. That's a fail. And the data is available and can be looked at with other things, but it just gets treated as a fail. When we combine them, that there was some percentage of fail.

DUANE WESSELS: So, you discard the fail?

UNKNOWN SPEAKER: Yeah.

DUANE WESSELS: Yeah, I think that's something we should do but let me pose a hypothetical with made up numbers. So, if there's 10 probes and 10 servers and 99 of those results came back fresh and one came back fail. One probe from one server was stale. So, would you say that the root server system was one percent stale? Or was it 10 percent stale? It was stale for 10 percent of the probes? What sort of statements would you come out -- the system in that situation?

MATT LARSON: It depends on, again, how many Anycast instances root server supports and maybe only a small handful are quite out of date. So, it depends on where you end up going, which attachment -- worst case scenario would be 1/13th is stale, right? You could jjust flag the whole server as stale. Go ahead.

LARS-JOHAN LIMAN: On the other side, if roots of operator has very few instances, more probes are going to end up at each server, so you will get a larger percentage. So, the number of failures will actually tell you something about how large fraction of the internet was hit by this, or possibly hit by this. So, I don't want to aggregate is up to -- just because one probe fails for some server, that doesn't mean the entire fleet in all instances is bad and that should be reflected in the report.

BARD VERD: I agree with Matt that serving the wrong answer is really bad; Larson said it's like the worst thing we could do. However, in this discussion about the service and staleness. Is this an effective metric at all? I'm not sure it is.

MATT LARSON: Either you're stale or you're not, like if you're two hours stale or 10 hours stale --

BRAD VERD: Trying to know -- but I mean trying to have -- trying to say we're 1/13th stale or this or that is just -- for a root server operator the metric makes sense. I'm curious, and what I'm asking, is does it make sense for the service? I don't think it does.

DUANE WESSELS: I agree.

UNKNOWN SPEAKER: Paul Hoffman?

PAUL HOFFMAN: So -- again, my hand's been up for a while but I -- Thank you, Brad, you just actually got what I was going to say like five minutes ago which is I am not convinced that this is a useful metric for this service. That is that assuming that we do the RSO measurement -- the individual RSO measurements -- the way we said before which is if it looks like that a single RSO is -- from that one perspective -- is behind try it again in a short period of time and use that as a yes/no.

I don't think we can translate those yes/no's into a service because it isn't something that somebody's going to say, "Oh my. This service is having a staleness problem, therefore I want to do X." They can certainly say that about correctness. They can certainly say that about latency, not that I'm going to agree with what they say about them. But not on staleness.

DUANE WESSELS: Okay. No more hands up at this point so, I'm hearing suggestions that this metric sort of be dropped. Is there anyone who wants to advocate for it instead? Or should we just take it as droppable? Okay, less is more. The symmetry will have been broken but we can go with that.

LARS-JOHAN LIMAN: We could do with ripping out more.

DUANE WESSELS: Okay, and we still have 45 minutes in session. So, we can talk about -- do we have any other parking lot items to talk about or should we get our hands dirty and start talking about thresholds maybe?

FRED BAKER: Would you like to have data before you start setting numbers?

DUANE WESSELS: Yes and no. I mean -- sure, yeah.

LARS-JOHAN LIMAN: Liman here. I think we either need some existing numbers to start from, or we need an algorithmical path towards figuring out what the optimal numbers be. I think we have neither right now.

WES HARDAKER: -- interrupt Fred was off mic when he asked, "Should we have numbers for deciding what the metrics should be.

UNKNOWN SPEAKER: Well, in terms of figuring out some kind of threshold, the hope that I had earlier -- I have not been able myself to think of any way to design an algorithmic sort of approach to these. But there's a lot of people around here that are a lot smarter than me and thinking about this sort of stuff a lot more. So, I think it would be a useful think to at least walk through them individually and say, "Is there at least a beginning of an algorithmic approach to how we might develop thresholds for these various metrics?"

WES HARDAKER: Let's start with correctness. I'm pretty sure that we've always been 100 percent correct. Should we stay there?

LARS-JOHAN LIMAN: But that's taking it from existing numbers which is fair in that case.

DUANE WESSELS: I don't know. To be honest, I'm sort of struggling about where to go with this because I'm sensing hesitation that some people don't want to talk about, data would be nice to have. I feel like we should take advantage of our time here together to have a discussion. So, maybe we could just start having the discussion and not necessarily have to come up with actual values and thresholds, just discussing of how we would go about it.

LARS-JOHAN LIMAN:

Liman here. I think we should look for -- before setting actual thresholds, I think we should have an understanding of what the system looks like today. Now, that said, we can still do work here because I think we can discuss around as a -- once we have the numbers -- first, which numbers should we try to find out? Second, once we have those numbers how should we use them? And, given those numbers what thresholds should we try to set?

I think we should start from the stick in the ground that's just put down which is the current system and we should look at the numbers for the current system and use them as some kind of starting point and we should probably try to figure out which numbers we want to know. We probably have defined pretty much what we want to know but we don't have the measurement gear in place to actually get the numbers. But given the current numbers, how much fluctuation, how much change from those should we allow before we raise the red flag? That's probably a discussion worth having.

UNKNOWN SPEAKER:

Thanks, Lars. So, I don't know if this is helpful but during the discussions for the transition, for the IANA Stewardship transition, the [inaudible] creative was called the design team A that specifically worked on the numbers and the measurements I'd like to see [inaudible] for the performance of the naming function, and so they asked us to supply data off our current performance. I think we supplied data for about six months. It wasn't a large enough sample but that's the data we were able to collect in time for that team.

And so they used that 63 measurements that we report against currently, the report I sent earlier today. They weren't necessarily with the measurements so [inaudible] in them, so they gave us a bigger SLA to live up to through the transition -- but it was helpful because it was based on the current system. So, I agree, among the 13 roots it was right, and we have lots of data, so I think it should be based on the current system, my personal opinion. And I think it should be aggressive and, you know, what I would like to see coming out of the system in terms of in terms of how do you set the threshold, again personal opinion.

But one thing that would be really helpful with the current system for us in reporting on resolved performance -- on naming performance -- the design team A and now in working with the CSE would give us good thresholds to work with. So, for some of the measurements we have to meet that 95 percent of the time we have to meet that threshold. For some it's 100 percent, but the correctness one, there's no negotiation on that one.

So, just looking at the measurement and having thresholds, some are down to 90 percent even. That's really helpful and I think in that last two, three years it's been a much more predictable and open process because we know what we have to report on and then the community knows what to expect from us. So, I don't know if that's helpful, but I'd encourage you to jump into the numbers if you can but also base it on the current system as well.

LARS-JOHAN LIMAN: Liman here. Thank you, that's useful. There is in my view a major difference though between the PI's of a system and the root server system. One is that you have 12 players involved here. The other one is that we only control so much of the system because a lot depends on the network infrastructure that we control. We cannot possibly adapt to it, we cannot control it. The PDI is in a better position to control and actually stronger influence the numbers that you need to change. So, it's just an observation that we should keep in mind when we work with this. Otherwise I think this is very good advice. Thank you.

JEFF OSBORN: Although I'm the one who put the stick in the ground first and said the system is good, I think we need to recognize when we present that it's to a group that has publicly said it's not good or at least it's a great risk. So, having said we can just stick a stake in the ground, we also are going to need to defend it. So, I don't know if that means we can just throw our heads back or we actually have data, but we need to do something.

BRAD VERD: In the spirit of baby steps and small victories there are four metrics, right? Essentially? And I've seen numerous people agree -- I've seen everybody nod their head -- I've heard numerous people say it already today which is that the worst scenario is to give out a wrong answer. It sounds like the one metric on correctness is 100 percent. The one threshold. So, that's 25 percent done. Small victory.

UNKNOWN SPEAKER: Go ahead, Russ.

RUSS MUNDY: So, I can aim for 50 percent, so we can get an F plus. I just actually measured the entire root server system from everywhere, really quickly -- I actually did it for the last day and so the result was n. So, my question is what multiplier do we want to put in front of n to get a latency threshold?

JEFF OSBORN: Is this going to be an aspirational number or a never exceed this number? Because that makes a big difference.

RUSS MUNDY: Thank you for pointing out what I've been trying to say about individual versus global that the hard thing. Because do we want to hold everybody the same to the must versus it would be nice if you were about this.

LARS-JOHAN LIMAN: Liman here. The n, was it derived according to the algorithm that we talked about or how was it --

JEFF OSBORN: Absolutely. I preconceived that --

UNKNOWN SPEAKER:

In terms of -- another way to answer your question, Jeff, is the first set of numbers that we come up with -- and that's with thresholds and everything -- they can be totally kept within this group if we choose until we feel that we've got enough of the right kind of information to even -- I mean our set caucus list is public list, so anybody that wants to go and read the list can see what's going on and that's okay. But, I guess in my view, we can publish things when it becomes an official RSSAC thing only when we feel there's sufficiently good enough for us in terms of reflecting either today's reality or what we want to see in the future.

PAUL HOFFMAN:

So, just as an observer here, I think almost every other speaker who's been talking for the last 10 minutes has switched between, "Are these thresholds for the system or thresholds for an RSO?" I don't think that they did that consciously but we ain't going to get anywhere unless we pick one and I think it's much easier to pick thresholds for an individual RSO than it is for the system, especially -- other than correctness, I'm not going to take away our 25 percent success on correctness or 33 percent for the system but my feeling is that we have no idea how to mathematically collate latencies and availability for the system, whereas I think that even though we may have a fair amount of disagreement, we do have ideas that might turn into a consensus without a heck of a lot of beating on latencies and availability for an individual RSO. Once we have those then thresholds might or might not make sense.

One of the things I've heard different people say is, "It's okay if RSO --" let's not use letters from A to M -- "If RSO t has crappy latencies

because they are purposely serving a very different audience and they don't want to spend money on throwing up five well connected root server instances that the probes are going to find" So really pick between whether we're trying to do thresholds for the system or for RSO's, and then I think we will have better focus.

FRED BAKER:

Well, I have a suggestion. [Inaudible] root operators, Dan, has ping ponged running through all our instances. We could probably get that measure which would give you an idea of what the latency is worldwide as opposed to from Washington DC. And that would be a similar test to what Russ has been running for the last day. We could probably get him to throw that in and that would at least give us something that was wider, and I imagine Verizon can do the same.

JEFF OSBORN:

For what it's worth, ThousandEyes offers a free probe or test for all the root servers. They use 20 probes, I think, for V4 and 20 for V6 and that's been running for a couple of years now. So, anybody can log in and create a free account and take a look at it if you want. Or I can show you as well.

FRED BAKER:

Okay, well if we can get a number from it. I'm just, very uncomfortable picking a number out of the air. I would really like to see some data.

DUANE WESSELS: Well, I think that's -- Wes is going at it; let's say we had a number. What next? Go ahead, Jeff.

MATT LARSON: The numbers range from nine -- sorry, the numbers range from 9 milliseconds to 123 milliseconds; average latency for the root servers, from ThousandEyes from 20 probes for what it's worth.

UNKNOWN SPEAKER: Nine to what?

MATT LARSON: 123 milliseconds, on average.

JEFF OSBORN: You'll pardon me for trying to use an analogy for this but I'm more comfortable with budgets than this.

FRED BAKER: Is it baseball?

JEFF OSBORN: It isn't, it is money. In the budgeting world, there are two ways to look at a thing. And there is a budget limit, which if you ever hit it, you're fired and banned and horrible things happen. Nothing personal, but government's budgets tend to work that way, where it's a thing to never exceed. Whereas I found that commercially a better budget is the

one that -- it's the thing you expect to hit and all deltas are interesting. "Why did you spend less, why did you spend more?" That's interesting; we thought it was going to be along this slide.

So, what removes a punitive aspect where you want to do more things? When you came up with n , my immediate thought well, n is the perfect number to start with and then all the deltas are interesting. Or, if there is a punitive limit, then you simply say, $3n$. I think the $3n$ is kind of dumb but if we are going to be in a position where you're fired as an RSO if you fail to maintain your place in the threshold, then put it up high so we never bump our heads. Does that make sense, because it's clear to me but this [AUDIO BREAK]

KAVEH RANJBAR:

As Duane mentioned, I showed the graph. There are a lot of numbers. I don't think we have enough numbers either from ThousandEyes or from ATLAS, millions of measurements already towards all of the root servers all kind of stuff. We even have on others, we even have some visualization already, where we can also go back in time for the system or per letter, per identify. But I'm not sure if we have to do that work right now to be honest, because I think, maybe we need to do a bit more on our methodology and have something before coming up with those numbers.

So, looking at the scope of the work part [inaudible] like, it's good to document that we want to do that, but we don't do it right now, that can be a separate effort maybe for the next workshop or something but we come to the workshop with some decisions, some idea, that that's

how to get the numbers, where we want to get them and how we want to set them. Something like that. I find it a bit rushed if we are trying to do it --

MATT LARSON: No, I agree, I think, I mean, I think what's happening is what we all kind of thought was going to happen right. I just can say this is just an open discussion in throwing ideas out. I don't think we need to have the goal today of actually reaching agreement on any particular numbers. I just wanted to get people's approaches and you know, sort of where you're thinking. Brad.

UNKNOWN SPEAKER: I think Brad was next.

BRAD VERD: Yeah, rather than having the discussion on what the threshold might be or the numbers would have, what if we had a discussion on philosophy approach type of thing. And what I mean by that is availability. So I think we could all [inaudible] think that availability of service should be 100%. But, I'm not throwing that threshold out; I'm just saying I think philosophically, we should believe the service should be available.

WES HARDAKER: That matches past measurements too.

BRAD VERD: I think what would be an interesting discussion is what do people think the availability of a single RSO should be, philosophically? I mean, I can tell you from me personally, I'm not speaking for other people at Verisign, but I think we should believe, it should be 100%. We believe that an analogy of RSO is kind of unacceptable. That's our philosophy.

WES HARDAKER: That included under a targeted attack?

BRAD VERD: Well no, I can't guarantee that, but the goal is --

WES HARDAKER: No, but you just said 100% right, that is --

BRAD VERD: That's the goal, yeah.

JEFF OSBORN: I think that's -- I'm looking at Fred and it's fair to say, I think we certainly operate under that same intention. That's our goal.

RUSS MUNDY: So, let me -- let's pull up a level. We started down this work train to get ahead of the PMMF -- PMFF. I still don't know. Let's think about what they would do. So they are not going to do constant review. They are not going to say, you know, "These 30 seconds, unless 30 seconds

something changed, I'm going to take action." I think in 37 we said once every three months or once every month, I don't remember. But at some point, they have to look back across the aggregate past and months of data. Well say months -- I'll use n, n months of data and they have to make a determination, so is crossing that threshold once on a day bad.

If you look at any alerting system, you know, built on the planet today, there is not green and red, there is always a green, yellow and red. What do we want? Do we want a mane a must? What happens? How many days do you cross the threshold in our day aggregation metrics before a must action takes place as opposed to a warning letter. So, eventually all this stuff is going to be used by that body. And I think if we pull up a notch and say, "What does that body need to make a decision, and at what level does the decision trigger can we get back into how to define these numbers?"

DUANE WESSELS:

Yeah. I've had similar thoughts. I think there is a couple of ways -- I think there is some disadvantage to having a single threshold right for these metrics. One, things you do instead is what you just said; you could have a red, green, yellow or some scale of how serious things are. You could also have sort of different threshold, you could say something like, "Well, availability must be 90% for 95% of probes and there must be 50% for 100% of probes," or whatever. So you know you could have a couple of different options there and all of those need to be met. So I think there is a lot of different things we can do to make it better than just a single threshold, a single bullion.

MATT LARSON: Like what?

RUSS MUNDY: Like what would you make better? I mean, what would you give a proposal -- I'm trying to extract the exact answer.

DUANE WESSELS: I don't have a specific -- I mean, I'll just repeat myself. You can have different levels of thresholds. You may have one threshold at a 90% level, so 90% of probes have to meet a certain threshold, 50% of probes meet a different threshold, something like that. Or you can say if availability is 99% and above, it's green, if it's 80 to 99% it's yellow. You know, so, some kind of grading scale.

WES HARDAKER: Anybody objected to Duane's numbers?

DUANE WESSELS: Trying to nail me down?

LARS-JOHAN LIMAN: So, for the ICANN's TLD program, what the contract publish, is what they call the emergency threshold. And then before the emergency -- so this is the ceiling, and then before ICANN will send notification when that 10%, 25%, 70%, 75%, so as you work that threshold. And then if it

was 10%, you send an email saying, you know, "What's going on?" I think we should not just think of the threshold as the absolute number and then there is a lot of steps before that thing happens.

WES HARDAKER: What happens at the emergency level? What actions are --

LARS-JOHAN LIMAN: You will trigger the overall process. They have an emergency transition backup process.

WES HARDAKER: And that's a hard crossover in what amount of time; a single instance, in a single second or you know, a day aggregation? I looked at the stuff two years ago and I apologize for not remembering.

LARS-JOHAN LIMAN: I can get, you know, if it is useful for this group, we can get those.

WES HARDAKER: I think it's highly useful because that's a model that ICANN already knows and understands. It would certainly make sense that, you know, they wouldn't want us not to invent something 1,000% new.

LARS-JOHAN LIMAN: If it's good, but I agree. We should definitely have a look at it. It would be very valuable. I will assume that most of it is actually good.

WES HARDAKER: Yeah. Well, you know, it's funny because you're bringing back a memory that when we were talking about the 500 millisecond number that was thrown into the same program, I had a question earlier that was asked which was, "How many people are, you know, in the ICANN community think that number is good versus how many think it's bad and how many don't care?" I'd love to see the percentage breakdown between those three sets of people.

UNKNOWN SPEAKER: And maybe there's a fourth set, "How many have no idea what it means it all?"

DUANE WESSELS: Are we done? No. How much time until the next break. 20 minutes. Okay. Brad has an idea. Go ahead, Brad.

BRAD VERD: I don't have an idea. I just -- I keep hearing that the PMMF will only do monitoring sporadically. As I go back and I read 37, you know, the PMMF is doing ongoing monitoring,.

UNKNOWN SPEAKER: Never [Inaudible] decision -- you're right.

BRAD VERD: It says, in the case where measurements indicate that an RSO does not satisfy the defined thresholds, the PMMF will report -- will produce a report on the RSS performance. If a documented pattern or poor performance arises the SAPF may engage the DRF for possible action. It's an ongoing monitoring and then the thresholds continue to fall under and remediation doesn't occur, things have escalated.

WES HARDAKER: No, no, thank you for that. But there's not a -- we don't have a timescale for how long until remediation occurs, because it's been there for a while, but...

BRAD VERD: My only comment was about the monitoring that is happening all the time.

FRED BAKER: Well, my recollection is -- and from what you just said, the monitoring is happening all the time, but the numbers get interpreted by the SAPF not PMMF.

BRAD VERD: The number, the thresholds are set by the SAPF, or here right now, we could set them and then the SAPF would validate them, change them, add to them, whatever. Yes. And then it's executed by the PMMF.

FRED BAKER:

It is.

UNKNOWN SPEAKER:

Yeah, and this is one of the reasons why I've said that as we develop these things, and we reach a certain point of having a number, we can try it out before we actually publish it.

BRAD VERD:

I only bring this up because I keep hearing that monitoring was not all the time. Monitoring is all the time. All the time, we're collecting data. That's all I want to like, kind of --

WES HARDAKER:

I had two things. I'm sorry for miscommunicating the fact that monitoring doesn't happen all the time. It has to or otherwise the system doesn't work and; b, I am sorry, for misremembering that there's not a cron interval that's different than the monitoring intervals.

UNKNOWN SPEAKER:

I think the work party has benefited hugely from these sessions. An area that we did discuss some this morning that I'm not that comfortable with the guidance that we've gotten has to do with the probes and is there -- do we have -- should we try to establish some type of algorithmic relationship between, not only the quantity of the probes but potentially the geographic locations or topological locations of the probes with respect to what we know about the RSOs because,

you know, obviously, the probes make a big difference in what the end numbers come out being.

Does anyone have any further thoughts on on that in terms of how hard that work party needs to work at figuring out where -- how many probes there are and where they should be? I know we have a task to do that, but I'd love to get more input. Anything from the room? I don't see any hands in there. Go ahead, Brad.

BRAD VERD:

I feel that the work party should not spend a lot of time on that. Really don't. There are tons of platforms out there that you could use the anchors for RIPE. You could use Catchpoint, you could use ThousandEyes, you could use all these different platforms that have already done this and invented the wheel. We don't need to reinvent the wheel. We should not spend a lot of time on it, in my opinion.

LARS-JOHAN LIMAN:

I tend to agree to it to a large degree. That said, I think it might be interesting to try to perform some of these tests on different platforms and compare the results and see if there are large differences between the platforms because that will -- that's good input for the discussion continuing from there.

My hope is that there isn't a lot of differences and that would support your statement. And if there is a difference, we need to figure out why that's the case and that will help us to better understand the monitoring system that we are trying to build.

UNKNOWN SPEAKER: Okay, thanks. Kevin? Oh Paul, okay.

PAUL HOFFMAN: Assuming you're -- this is Paul. Matt just stepped out, Matt Larson, that Matt, just stepped out of the room but folks had asked earlier this morning about the testing done by GDD for the servers, and we got an answer back, which is a JSON file, which has two interesting things or what could be interesting to this group.

One is there's 38 entries of where the probes are and doing a quick look at, each entry has a city name, an IPv4 address and an IPv6 address. In doing a quick look, I recognize every city name, like I've been to many of them. I haven't been to Düsseldorf or Johannesburg. But basically, if you wanted to think in terms of what is ICANN currently doing for checking the gTLDs, it's not a large number. I don't consider 38 to be a terribly large number, and it's all from major cities, not like things, you know, off the beaten path.

BARD VERD: Paul, assuming there's a V6 and V4 probe, it's 19 sites, one could assume.

PAUL HOFFMAN: No, it is 38 sites, each of which has a V4 and a V6 address.

UNKNOWN SPEAKER: Great, thank you very much, Paul. Any other thoughts or comments on this? Okay, unless Duane has anything here, we can declare the work party effort for this meeting finished?

DUANE WESSELS: Well, I think -- what I'm starting with now is we have two slots to fill. I mean, our agenda for tomorrow is to keep talking about metrics. Maybe people can think about make suggestions if there's things you still want to talk about with relation to metrics, or are truly done with it, and we'll declare victory, I guess. But we do have those two morning slots allocated for this. If there's something else we should talk about, we certainly can.

One of the things for us on our list was as a point, and I forget -- sorry, I forgot who brought this up, but it says, "At what point should the RSS service coverage party be reactivated given this work on the metrics." Maybe that's something we can talk about tomorrow, if we want.

MATT LARSON: I think we want to maybe circle back around to real brief time tomorrow to talk about the independence and the updates that were made.

BARD VERD: Do the work party have enough output from this group to carry on?

DUANE WESSELS: I'm feeling a little overwhelmed, so I'd say yes.

BARD VERD: I mean, if we're going to end early tomorrow, or like run out of sessions, I don't want the work party to be like, "Well, geez, we really needed, you know."

DUANE WESSELS: So I think we have a lot of notes and information back and revise our document that we're looking at. Just I guess, just keep everyone in the loop, tentatively, what we're talking about is that Russ and I will meet with the ICANN staff probably next week to go over our notes and refresh our memories how this went. Then maybe a week after that, we'll have another work party meeting to again, to brief the work party on the results of this day's work.

At this point in time for both the benefit of our, our co-chairs and the shepherds to the work party, we are going to continue to try to have telecons every two weeks, we're talking about possibly making them 90 minutes instead of 60 minutes. Continuing to try to keep the priority and the movement as high end and as fast as we possibly can.

And as John says, he's feeling overwhelmed, I'm feeling overwhelmed with how much data -- we've gotten a very large amount of input, very useful input from this, and we got to adjust it and get it into the document. And we'll be putting it out again. Has this actually gone to the RSSAC Caucus list, this document? I don't know that.

MATT LARSON: Yeah, I think a PDF --

UNKNOWN SPEAKER: I know it went to the RSSAC list in preparation for this meeting, but I'm not sure that it went to the Caucus list. I think just to make sure we get all of our Caucus members covered. Unless anybody has an objection, I'd like to ask the staff to send this one as it is right now out to the Caucus list, pointing out that it'll be updated and a newer one provided. But since the Caucus as a whole hasn't seen it and it was a major item for today. If that's okay, we'll do that.

DUANE WESSELS: And you're talking about this Google Doc right here that's being displayed?

UNKNOWN SPEAKER: Yes.

DUANE WESSELS: Okay.

BRAD VERD: Well, I think [Inaudible] maybe it's too early. I'm not sure, but it sounds like given the rough timeline I'm hearing and what wants to get done before we talk about threshold sounds like maybe a fall workshop on thresholds.

FRED BAKER: Okay, so we'll think about that. Co-chair?

BRAD VERD: Yeah. It's something to talk about tomorrow as we work through. That's all, I just wanted it on the notes.

FRED BAKER: Right. Okay. I'm looking at the agenda for today and the agenda for tomorrow. Russ, correct me if I'm wrong, but I think you just wrapped up what happened today?

RUSS MUNDY: Okay, unless folks want to think of something else they want to cover tomorrow. I think that the co-chairs think we're done.

FRED BAKER: Okay. As the agenda reads, we have eight minutes until we have a break with snacks and then we have a wrap up, which we've accomplished. Then we have dinner this evening, courtesy of Verisign, and I'll let you talk about that in a minute. But Ken, coming back to your thing, what that means is that tomorrow morning, we have a slot for the hours metric three latencies staleness which we might not use.

We have [Inaudible] where we have some other things including reactivating, including reactivating the RSS service recovery part. It seems like we can probably use either of those for you. Would that work for you?

KEN RENARD: Yeah, I'm sure I'll be much less than either one of those slots, but that was my estimate the other day. It was not very good, so. Okay. Yeah, some portion of that, not all of it certainly.

UNKNOWN SPEAKER: Now then at 11:30, we have our monthly RSSAC meeting. I don't know about the rest of you, but I will probably not have read the independent document by then.

RUSS MUNDY: I was planning on talking through the deltas. You can look at the delta in the document. I'm hoping that's not a huge amount of changes. I don't think the read ahead could be during the working breakfast tomorrow. No homework.

FRED BAKER: Oh, cool. And what I'm thinking about there is one of the things we've got scheduled during that meeting is everybody vote. We might very well decide to put that off to an electronic balance the next week or whatever. But yeah, let's use the 30 slots then for your independence discussion. I don't know what you want to do with the [inaudible]. I mean, we can talk about that.

BRAD VERD: Well, I mean, you got the question about the reactivate [Inaudible].

DUANE WESSELS: Yeah, I think that's probably really the only thing we haven't talked about yet with relation to the Metrics Work Party. We'll talk about it anytime, but that's the only thing remaining from our parking lot.

FRED BAKER: Yeah, yeah. We can talk about that tomorrow and then maybe take a long lunch or something.

BRAD VERD: Well, then I think we should talk about in the parking lot of kind of the timeline, since we're going to need to have all these numbers before we talk about thresholds, we can start kind of laying that out.

Yeah, I think there's some -- there's probably work that can be done in Marrakesh and that probably will fall, if we have a workshop or if we do, we want it to be wherever we're, is it Montreal or something? But things like that, we can start talking towards what the timeline is. That's all.

FRED BAKER: Okay. Then at 11:30, we have the monthly, RSSAC meeting and that it will actually be the main meeting. We will not have a meeting in May. After which, we all assume a way -- well, most of us certainly, but we have to talk, but I will be home. Right, I think that's pretty much the plan for tomorrow. With that, do I just let them go?

UNKNOWN SPEAKER: [Inaudible].

FRED BAKER: I'm seeing the running shoes coming out.

UNKNOWN SPEAKER: Sounds like [Inaudible] too. Ozan, Carlos, anything?

UNKNOWN SPEAKER: Brad, do you have any details for dinner?

BRAD VERD: Dinner is here at Morton's, right down the way; if you don't know where it is, it's on Google Maps, you walk right there. Just walk around you'll find it. It's right next to Starbucks; if you know where Starbucks is, it's right next to Starbucks. We have a private room in there. So, food, beer, wine.

UNKNOWN SPEAKER: Can you all shoot me a [Inaudible] please?

BRAD VERD: Yeah, it's seven.

UNKNOWN SPEAKER: I do want to say this Morton's is where the hearing board has most of its meetings, and I've been there, about 100 times, and they actually have credible vegetarian choices. They don't just have a token choice for those of you who don't like meat. I have no fear, you'll be fed.

[END OF TRANSCRIPTION]