
STEVE SHENG: Welcome to the conference call. On the call we have Afifa, Amir, Daniel, Dessalegn, Fred, Ihtisham, Jaap, Jeff, Karl, Kazunori, Ken, Kevin Jones, Kevin Wright, Lars-Johan, Paul Hoffman, Robert, Shinta, and wess. From Staff we have Andrew, Ozan, and myself. Did I miss anyone? Anyone on the phone who has not joined the Adobe Connect?

PETER DEVRIES: This is Peter DeVries, I'm here, I'm not on Adobe Connect.

ABDULKARIM OLOYEDE: Hello, this is Abdulkarim Oloyede. I'm not on Adobe Connect.

STEVE SHENG: Okay. Welcome. So, over to you, Duane.

DUANE WESSELS: Alright, thanks Steve. Thanks everyone for joining the call, it's good to see all this participation. Today basically my plan, and I have to apologize for my co-chair, Russ, I know he's busy with other things today, so he's not able to join us. The plan was to discuss some parts of this document that I hope you all have in front of you. This document is essentially a combination of the slide deck that we talked through at our previous meeting, plus some additional notes and some sort of straw man proposals for some metrics.

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.

There are about six or seven sections that I think we should go through and then if we have time we can talk through the other things, but I'm guessing we're going to be a little tight on time, we've got basically an hour. We we'll try to limit it to about 10 minutes per section, if possible. I think let's just go through these in order, if that okay. Any comments or questions before we get started on that?

So, the first section, this General Comments / Discussion section, and some people have been typing comments in the Google Doc, which is great. This was mean to capture some of the discussion from the previous meeting. The only thing that I think maybe is worth talking about on the call today is this last paragraph, there was a question about whether we should be focusing on measuring root server operators or resolvers.

I think it's an important point, and the way I interpret this is, do we want to understand how the root servers are functioning and operating and returning things, or are we more interested in how a root server client, i.e. a resolver, how it interacts with this system and what sort of thing, if by return. I feel like for the most part we're sort of heading in a direction of measuring the RSOs and not so much measuring resolvers.

And also, by the way, when I say measuring resolvers, I don't mean to measure all the resolvers out there on the internet, I mean, for example, to measure behavior of that isolated resolver in a controlled condition, for example a probe or an agent, or something like that. So there has been some input to this already on the document, but if anyone else has comments on this, I would welcome those at this time. Paul, go ahead.

PAUL HOFFMAN: Hi, this is Paul Hoffman. So, I think that we are better off only looking at the root server side, simply because we have a general picture of how resolvers should act, such as priming with the root servers and asking various ones, but for those of you who have been following the DNS operations mailing list in this last week, there are many resolvers that use nonstandard configurations I think more than many of us thought there were, and certainly in some areas. So, I don't think that it is a good idea for us to try to model resolvers as a way of measuring the root system. Maybe years from now, when we've done more research, we can do that, but for now I would say we don't know and we will probably guess wrong. Thanks.

DUANE WESSELS: Okay, thank you very much, Paul. Alright, I think that captures the sentiment in general, as well. Unless anyone else has a comment on this topic, I'll move to the next section.

So, the next section talks a little bit about a Measurement Platform which, of course, does not exist today. The reason it's here is to set some ground rules for what may or may not be possible, how such measurements might work. Hi, who just joined?

RYAN STEPHENSON: Yeah, it's just Ryan. Hey, how's it going, guys?

DUANE WESSELS:

Hey Ryan, good. We're going through the doc, we're only on the second section called Measurement Platform right now. There's some text in this section also, similar to what we just talked about, which is whether or not resolvers should be used to make some of the measurements. And I put some text in there about some of the advantages and disadvantages. Does anyone have comments or questions about what's written in the Measurement Platform section at this point? A lot of this, of course, is to be determined, to be figured out later, which I think is fine at this point.

Okay, let's move on to some of the more specific metrics sections. These are again in order of the way that they were written in the Statement Of Works, so we'll just go through them in that order. The first one talks about what does it mean to say "that the root server system as a whole is online?" And these bullets here are some of the responses that we got from that little survey. There's a short discussion, and then there's a straw man proposal for an Onlineness Metric. And there are some open questions here about how this should be done, whether or not, for example, if you're measuring the servers, do you hit them on both v4 and v6, and so on.

The proposal that's here is to say these probe clients and queries, and then you can report on things such as the percentage of clients for whom at least some percentage of the root servers responded. And there is sort of a long discussion here about whether or not it makes sense to think about things like 5% of 13, or not, I think that's maybe a detail that can be worked out later. Any comments on what is written in this section at this point? Or suggestions for improvements? Paul?

FRED BAKER: I do have a question. When we talk about 95% of the servers, or whatever, is that from 13 addresses or 26?

DUANE WESSELS: I think that's open for debate. For example, if it's 26, then that necessarily assumes that all of the probe clients have both types of connectivity. Do you have a specific feeling one way or the other, Fred?

FRED BAKER: Well, personally, I'm coming at this in part from IPv6 operations chair in the ITF, I'm interested in all 26. And I think that tells me a variety of things including whether the probe itself has v6 capabilities or whether it's in a network that has v6 capabilities. It might be on a LAN that has v6 capabilities and completely surrounded by a region that doesn't. There's any number of possibilities there. It seems like at least one of the things that we would want to get out of this is whether, let's imagine that we're in some locality that has decided that they want to have a national root server. Could I only get to that one? And why would that be? Would that be, is there a firewall, is there simply a lack of connectivity, what does that mean? I would be interested in seeing 26 respondents.

DUANE WESSELS: Okay. Paul, go ahead, you had your hand up?

PAUL HOFFMAN:

I did. So, two things on this, I put both of these in the document, but I don't think they got answered. And this might actually go to the meta question of maybe RSSAC is going to change these after you all meet in a couple weeks. But measuring the root server system as a whole means 100% have to pass, unless we can say it's fine for the root server system to have only N percent pass, but then it's not the root server system as a whole, working. It an 'in general,' the root server system in general is online.

And then the second going to Fred's question, we actually discussed that you can see the thread started by Ray Bellis there, about IPv6. As much as I like IPv6 deployment, I believe that if a probe can reach a particular RSO over v4 but not v6, or over v6 but not v4, that's actually not relevant, that is, that means that there is some path, and one path is working and one path is not. That has nothing to do with the root server system as a whole being online, it has to do with v4 or v6 routing being broken. So it would be good if we didn't have these metrics stretch into internet health in general, if we kept them focused on the root server system, because I think we have enough problems as it is with defining that. Thanks.

DUANE WESSELS:

Alright, thank you, Paul. Robert, go ahead.

ROBERT STORY:

I made some comments in the document and I'm on the opposite side of the fence from Paul. My comment basically was if we're talking the root system as a whole, I think the reason we have the diversity of 13

operators is that if there are issues with one or more, the others pick up the load. So for me, if the majority of users can get an answer from somewhere, then the system as a whole is functioning; it may be functioning in a degraded state, but it's functioning.

And whether or not we're getting status of the system as a whole, where we could then evaluate the percentage of operators that are responding or not responding, or latency is too high, is a different metric than are users getting responses for the system as a whole. And so I think instead of the number of clients, the number of RSOs a client can get to is inverted. It should be the number of RSOs that get an answer from a client.

DUANE WESSELS: Do you mean get a question from a client?

ROBERT STORY: I'm sorry, the number of clients to get an answer from an RSO, at least one RSO. So, if 95% of the clients are getting an answer from one or more RSOs, then the system as a whole is operating.

DUANE WESSELS: Okay. So, that's kind of similar to this first bullet which says "The percentage of clients for whom at least 5% of root servers responded," so you're saying the percentage of clients for whom at least N or one root server responded?

ROBERT STORY: Right.

DUANE WESSELS: Okay. Daniel?

DANIEL MIGAULT: So, I think going back to Robert's comment, one of the reasons to evaluate the role RSS, we are taking the path to evaluate each RSO, but it doesn't mean that if one is failing the whole system is going to fail. So I think it's not that we measure something on RSO, one RSO failing doesn't mean the RSS is failing. I think that is the difference we have to be clear on.

The other thing, back to Paul's comment on if a resolver is not able to reach an RSO or the root server system via v4 or v6, it's not because a resolver cannot reach the root server system that necessarily the root server system is failing. If you don't have connectivity, the RSS can't do anything for that. We have to be careful that what we measure is always assuming there is no networking problems between the client and the RSS. When you have one failure, it doesn't mean the failure is on the RSS side all the time. When it works, it means the RSS is up. When it doesn't work, then additional tests or investigation should be happening.

DUANE WESSELS: Yeah, Daniel, the measurement platform section has something along those lines, but there is the idea that the probes will test internet connectivity and if they're not connected, then the measurements are

not taken or they're discarded, or something like that. So I think that's good. So, in the interest of time, I'm going to move onto the next section, we're already 20 minutes into our call here. And I know that this probably isn't enough time, we could spend a lot more time on this document. So, if you have comments that you don't get to say during the call, please type them into the Google document.

Okay, the next section says "What does it mean that the RSS as a whole is serving correct responses?" I think there is a very important point in the discussion which is does it really make sense to think of the system as a whole as being correct or incorrect? Because if we're probing individual operators and one is doing something that you would call incorrect, then how do you classify that? Do you say that it's mostly correct, is it is some percent correct, or does one incorrect make the whole thing incorrect?

So, there is no proposed metric for this topic or this item, mainly because I think that this is an important point that we need to solve. So, any comments or questions about this section? Daniel I don't know if your hand is a new hand or an old hand.

DANIEL MIGAULT: Sorry, I wasn't paying attention.

DUANE WESSELS: Fred? Comment?

FRED BAKER: Well, it seems like correctness has to be defined by RSSAC 1. It would be that we're literally distributing the data that we received from the IANA or from the root zone maintainer, and that would include the signature and the signature when it arrived was correct. If we don't define what it means for a response to be correct, we're going to have a problem.

DUANE WESSELS: So, if one operator is not doing that, then is the system as a whole correct?

FRED BAKER: Well, I think there are at least some of these metrics which are the sum of the RSOs. I'm getting responses from 13 RSOs, I'm getting responses from 2, and it's going to be difficult to tease those apart. There are other measurements that are in fact of the system as a whole, and if the only response that somebody asking a question can get has the wrong signature, then it seems like they're not the IANA data, they're getting something, but it's not the root data.

DUANE WESSELS: Right, but if you look at it from a recursive band server point of view, if it gets a bad signature, it's most likely going to retry somewhere else, and then it gets the right answer. So that resolver got the right answer that it wanted and ignored the bad data.

FRED BAKER: Well, it ignores the bad data, but it seems like the fact that it's getting bad data is also a relevant thing.

DUANE WESSELS: Yeah, but that can be handled in the part where we say our individual RSOs...

FRED BAKER: Well, yeah, if we want to take it that way, I'm willing to do that, but seems like that question is part of the question.

DUANE WESSELS: Yeah, we need to, as you said, we need to get the definition for this, and I don't think we have it at this point. Robert?

ROBERT STORY: So, for this question and the system as a whole, the comment that I made in the document is kind of the inverse of my previous one. And that's where for our users to have confidence in the system, they need to have confidence in the whole system and all operators. So if there is an operator that's not serving correct responses, as opposed to maybe being modified on path or something, then to me the system as a whole is not meeting the metric. For RSOs to be serving bad data is a pretty serious thing, in my opinion.

DUANE WESSELS: Okay, thank you. Any other input on this section? So, Robert, let me get back to you. I guess given what you just said, it sounds like a proposed metric would be that you do these probes to all the operators, you check their signatures and what not, if any one of them has a signature that doesn't validate, then you say the system as a whole at that point in time is not serving correct responses?

ROBERT STORY: Right.

DUANE WESSELS: Okay, thanks. Paul, you've got a comment on this section?

PAUL HOFFMAN: So, something that you just said brings up again, this is a higher level comment. So RSSAC asked us to come up with some of these measurements with respect to RSSAC037-038. So, let's take a specific example that you and Robert just went through. An operator for some reason is giving out a wrong answer for one thing. And if we go with the definition of "then the root server system as a whole is not serving correct answers," what the heck does that mean with respect to RSSAC037, if anything?

I guess my question is once we have these metrics, how are they going to fit back into RSSAC037, if there are times when the root server system as a whole is not fully available or is not giving correct answers, what will the effect of that be?

DUANE WESSELS: Yeah, I think that's a very good question. So I'm going to make sure that we have time to talk about that at the RSSAC workshop in a couple weeks. Jeff, go ahead.

JEFF OSBORN: I'm trying to figure out where to put this in. Paul, that's a really excellent point. I'm current struggling to flesh out a financial policy as part of 037 and you just picked a really good hard example. If we are somehow as RSOs beholden to ICANN to some level of service level expectation, one would presume that some of those expectations is that what is being served is correct.

So, currently it's certainly expected, but there is no legal document requiring it. So this is the perfect example, I think, in two weeks for us to say, what consequences should there be for serving incorrect data and how does that relate to the contractual obligations we're trying to put together? And if that strikes anybody as complicated, you'll understand why I'm not sleeping well this week.

DUANE WESSELS: Did you want to respond, Paul?

PAUL HOFFMAN: That makes sense, I really was asking a question, not a leading question. Because I'm sleeping well this week, but I can see how we as the caucus work party not being able to come up with good answers for these

metrics, could make the job of figuring out how this all fits into 037 much more difficult. And by the way, correctness is going to be the easy. I think the "is the root server system as a whole, operating," is going to be a really, really gnarly one.

Quite frankly, so is, we haven't gotten there yet, and I know you're trying to get through Duane, but so is an individual operator. If an individual operator is not available from any of the probes simple because somebody is doing a concerted DDOS attack on that operator, what's the consequence of that? So, I'm not proposing any answers, I'm saying it would be good for us, when we come up with metrics, to understand why we're coming up with these metrics. Thanks.

DUANE WESSELS:

And Paul, I also think that not everything that comes out of this work party needs to necessarily have a tie in back to RSSAC037. I think that's certainly the driver for this, but I don't think there has to be a one to one mapping. But again, we'll talk more about that in a couple weeks. Alright, Fred, you get the last word on this topic, so if you want to respond, go ahead, and then we'll move onto the next section.

FRED BAKER:

Well, yeah, I'm basically responding to Paul. I put something in the chat room. RSSAC037 talks about measurement data being used to evaluate RSOs which could result in somebody deciding that an RSO isn't being functional and shouldn't be part of the root zone. If there was a DDOS attack against a particular operator, it seems like that would be a whole other discussion. And so if the measurement data, we're getting bad

data from somebody or we're not getting data from somebody, that raises a question, it doesn't indicate a necessary outcome.

DUANE WESSELS:

Thanks, Fred. Alright, let's go on to the next section which is the last point about the system as a whole. It says, "What does it mean that the RSS as a whole is serving timely responses?" And here, there are actually three proposed metrics. One of them is a straightforward latency metric, or maybe not really straightforward, but a little bit complicated, but essentially measuring roundtrip times of queries and back and somehow reporting on that, either reporting on the number of clients whose minimum latency was less than some threshold, or a set of thresholds, if you. And then there are a couple of metrics around what I would consider staleness, which is essentially checking the SOA serial number to see that the system or the operators are serving up to date copies of the zone.

There's this Alternate Proposed Staleness Metric that goes into a little more detail, and it's probably better, because it accounts for the fact that you might be doing one of these measurements right at the time when a new one is coming out. So it has an approach to deal with that situation. Any comments, questions, suggestions on timely responses? Robert?

ROBERT STORY:

So, one comment that I made in the document was that a fixed value for the metrics, be it 200 msec, or whatever, it seems like it should be relative, that could take into account the connectivity or latency that a

probe is in. So if a probe is seeing over 200 msec latency to all sites, then that doesn't mean that there is anything wrong with any of the operators, it's just the latency at that time.

And also, if we're talking about the system as a whole, it should somehow take into account to figure out where the probe is and what RSOs it's talking to, and which ones have any cast in that region. Not all operators have any cast everywhere, so I don't think that being slow in a particular area might necessarily indicate that that's a bad thing, as long as there is adequate coverage from some of it.

DUANE WESSELS:

Yeah, so the proposed latency metrics tries to address that. For example it says a probe would query all identities, and it would report on the minimum latency and/or perhaps the median latency. I have an idea to think about regions, if you can write down a way that can be done, I would say please do that. It sounds kind of hard, because I think you're now getting into geography and things like that.

And then again, your other idea about the relative latency, if you have an idea for that, please write down something more specific. I will say that I did, and ICANN staff pulled out some of the metrics that are in the gTLD guidebook, and they do have a fixed metric of 500 msec for UDP, just as one data point there. Daniel?

DANIEL MIGAULT:

Yes, I agree with the comment from Robert, and I think it's something like that in the Excel sheets, so I will complete the Google doc. A

number seems to be quite meaningless. We should rather look at the ratio given some other type of connectivity. For example, the probe might do some connections to let's say Google or other cloud providers, to have an estimation of the average or acceptable connectivity in that region, and then compare that to the one it obtains from the RSS. I will set some lines on that.

DUANE WESSELS:

Okay, thank you. If there is no other comments on this section at this time then we can move one. So, the next one, we're getting into metrics based on infiltrate operators. So, the first one is the Statement of Work said, "assess the performance that each RSO provides." There is no proposed metric in this section, mainly because performance is an undefined term at this time. Does performance just mean latency or does it mean something different?

Any further thoughts on this section? Daniel, I don't know if your hand is from before or if it's for this section; if it is, please go ahead. In the Google doc, Paul, you had a comment that, "If we can't agree on single-RSO performance," that's kind of a big deal, and I'm inclined to agree. So, what do people think performance means in terms of individual root server operators? Paul, go ahead.

PAUL HOFFMAN:

I understand, Duane, what you said about the RSSAC002 are self reported, and therefore they can't be externally monitored. On the other hand, knowing that a root server operator is responding to a particular sort of order of magnitude of queries, and then they say that

they're not, could be relevant. I'm trying not to grasp here, and again, this really goes back to RSSAC themselves saying more about what this metric is used for, particularly with respect to RSSAC037. But in the past, when we've had discussions about root server operators, some people have said, well, that letter is barely doing X queries, based on the 002 data, compared to this one who is doing many more. That might be relevant, might not.

DUANE WESSELS:

Yeah, it's tricky because in a way servers don't really choose who queries them, it's a little bit outside of their control. Jeff, go ahead.

JEFF OSBORN:

Paul raises the question and then kind of veers away from it, which is an action I've seen hundreds of times now. But with the publication of 037, one of the things that is going to be sort of forced on us at some point is what we've been referring to as "you must be this tall to ride this ride." And the section we're on now sort of looks like how must you be to ride this ride. And there are two possible answers, how tall that line should be, one is low enough that all the existing players can pass it, or not.

To go back selfishly to one of the parts I'm working on of whether there is funding involved, much of that wraps around the idea of if anybody is currently inadequate, do they get some money to fix themselves up? Well, if we make this line so that everybody passes, that's moot. But if we don't make the line so that everybody passes, the new standard becomes the smallest or least something operator. And this is the part

where this is going to start getting awkward, I think, or personal, so I'm just going to leave that there.

DUANE WESSELS:

Okay, so Jeff, I want to respond to what you said, because I think you're right, this sort of gets to the heart of how tall do you have to be. But the thing we're struggling with is how do we measure tallness, or the equivalent? Is there a way that performance, or the tallness can be measured by a third part for a given operator? How would you do that?

JEFF OSBORN:

That is such an excellent question. And the obvious ways, we'd sort of roll in that people have done are things like how many instances and how geographically are they distributed, and then on from there. But any time somebody starts doing this, they start with number of instances and where they are. Anybody that thinks this is going to a simple proposal that gets a unanimous vote 5 minutes later hasn't been paying attention.

DUANE WESSELS:

Yeah, so certainly something like number of instances and geographic distribution is easy to figure out, I guess. Either you can trust what the operator says, or you could do research and things like that. But it's quite a bit different than some of the other things we've been discussing which is based around probe queries that happen at regular intervals and things like that.

JEFF OSBORN: Understood. It just seems like you have more control to get inside an operator and measure things differently than you do the system as a whole that has to be viewed in the wild. To continue with how many instances and where, you could do how much band width, and where, maximum traffic capacity, these are just corporate network type things you could simply apply to an RSO as if it was a commercial entity, a lot of those same tools would work.

DUANE WESSELS: Alright, I'll let someone else have a change. Wes, go ahead.

WES HARDAKER: Thank you. Jeff, your mic is really echoing, I don't know if you know that or not. So, a couple things. One, 037 actually already has some notion of what do we want to measure in terms of deciding performance levels, and that's the BVQ measurement, which we should come back to and flesh out better, because it's a very high level thing. That's the basis for which we decided, so it's not just band width and latency, which is one of the things that's important to measure, and I think latency is one aspect of measuring BVQ, but the reality is can an RSO handle all of those metrics accordingly.

Back to a comment about what Jeff said, Jeff, you and I have had many discussions, so I don't think you meant it this way, but you said where do we design for a particular point at which all these people below this line fall below the current line and they need to be beefed up in some way, something like that, and we may define the line as being below the minimum.

Well, that's not the right way to design a system. The right way to design a system is to design it for what do we want, and then that line will be put somewhere in that lineage of where we are. And it maybe that everybody is below it, but may be that everybody is above. But I know you, so I don't think that's the way you actually meant it. But we should be thinking that way.

DUANE WESSELS:

Thank you, Wes. Fred, you're next in the queue.

FRED BAKER:

Well, I wanted to respond to a comment that Paul made, and has made in one form or another several times. This needs to go back to the RSSAC, the RSSAC puts parameters around questions, then we talk about it. The reason we've got the caucus involved in this is we're actually interested in the caucus' opinion. So if any of us have something that seems to need definition or further fleshing out or something like that, it would be useful both to have the question and to have what you think the response or the specification should look like. We're interested in the caucus of opinions.

DUANE WESSELS:

Thank you, Fred. Paul?

PAUL HOFFMAN:

So, thank you, Fred. I appreciate that, but when this work party was started, we were given the Statement of Work, which actually didn't go

into detail about 037/038. So I think with spent the beginning time looking, and still today, looking at the actual metrics that were mentioned, not their use.

So maybe a good thing that would come out of the meeting you folks are having in a couple weeks is a better description of whether you want to hear from the work party about things such as what Jeff mentioned earlier, or a couple people mentioned earlier, not just Jeff, of like, if we come up with these metrics and someone falls below them, do we believe that organization should stop being a root server operator? That's a much deeper and more serious question than is 50% good enough. I appreciate you all asking, I really do, but I'm certainly hesitant to start jumping in with the "well, I think the root server system should act like this," without being sure that that's what we're being asked. So, some clarity on that would be great. Thanks.

DUANE WESSELS:

Thank you, Paul. Jeff, I don't know if you're waiting in the queue, if you are, go ahead.

JEFF OSBORN:

Yeah, thank you. Sorry, I didn't take my hand down. But while I'm at it, Wes, you are correct, thank you. Wes was correct.

DUANE WESSELS:

Okay, thank you Jeff. Fred?

FRED BAKER: Well, again, responding to Paul, the Statement of Work doesn't specifically reference 037, but 037 is very much part of the context. So I think it's useful to look at and think about the statements that are made in 037. And yes, I think at some point we need to talk about thresholds or something like what does this measurement tell us about the root server system or the individual RSO, and what defines 'good enough?' that's not where we're starting out. The first thing I think we're looking at is what are the proposed measurements, but essentially we have to come up with that kind of analysis.

DUANE WESSELS: Alright, thank you. I'm going to close the hand queue, if I can, and Jeff, if you could mute, that would be good. We've got a couple more sections to get through and only about 10 minutes. So, Wes, quickly, go ahead.

WES HARDAKER: So, a couple things, one, I think that the goal of this working group is to define metrics so that a decision about what do to based on the RSSAC037 architecture could be done. So, I don't think it's in this working group's remit to actually decide what to do. The goal is to create measurements so that can be determined, and that the ICANN process of fleshing out the 037 model and eventually standing it up will take care of what do we do at a point when something passes or fails.

One other quick point, Paul, the meeting in a couple weeks with respect to further studying this work, I think there are five caucus members that are coming, so what I want to know from Fred and maybe ICANN Staff

is, is there remote dial-in participation, because if it's open I think to anybody that's able to participate, although we've had to cap the number of physical participants, but maybe somebody else could correct me.

DUANE WESSELS: Alright, we'll try to find out.

STEVE SHENG: Wes, this is Steve. We will find out and get an answer. Is there a desire for people to participate remotely?

WES HARDAKER: That's my other point, yes, I don't know.

DUANE WESSELS: Alright, I'm going to move on. Let's try to make it through the last two sections on individual RSOs. The next one talks about "the availability that each RSO provides." And the discussions are that this relates to timeouts and things like that. There was a straw man proposal for an Availability Metric which is probes send queries at fixed intervals with some sort of timeout on the order of seconds, for all permutations of v4/v6, udp/tcp, and then to combine all those results and report the percentage of probes that received a response within the timeout value. Any quick comments on the availability of the individual RSO?

Alright, let's move on to the next one. This last individual metric, in the Statement of work it used the phrase, "assess individual RSO service quality" and from our discussion in our previous meeting, there was sort of the sentiment, I think, that this would be better phrased as a "correctness" metric, and so that sort of aligns it a little bit with our previous discussions about the correctness of the system as a whole and there is a long proposed metric here which is how you might do that, which is send a number of different types of queries, these are different queries for different records, which all happen to be signed in the root zone and some of these queries asking for the DNSSEC signatures to validate them, and then you send a report on the number that are good or not good.

There is a comment in here about handle a situation where you might have a probe that's behind something that's interfering, or the probe query goes to something other than an "actual root server" and so you may want to also include some kind of hostname.bind or NSID measurements to try to detect those types of cases, as well. There are a couple of comments in the Google doc already on this one. Any further comments about service quality/correctness for individual operators? Go ahead Robert.

ROBERT STORY:

One of the comments that I had that I almost put in the doc, but I didn't, is when you get in to talking about middleboxes, another issue that comes up is when you have large packets and fragmentation. And so trying to figure out where the fragmentation is happening, if it's an issue

with the RSO or the path in between, is something that also needs to be figured out.

DUANE WESSELS:

Yeah, I think that's a good point and something that is in the start of the document, it talked about if we were to do some of these measurements using off the shelf resolver software, that resolver software would automatically, it has built-in retry logic for handling those types of cases, which could be good or bad. It's good and it sort of works around them, but it may be bad because it hides them from you if you want to know that they are there. Wes?

WES HARDAKER:

A similar problem, actually, Robert brings a good point to follow on, that there is also a possibility that even things like MPU availability and things like that is beyond control, we've discovered that recently with one of our installations where packet sizes in some places is different. So if we want to declare that as a metric.

DUANE WESSELS:

So, you're proposing a separate metric to measure MPUs somehow?

WES HARDAKER:

Well, to measure, do we want a metric for the packet size that the system will support either individually or as a whole.

DUANE WESSELS:

That's a good idea, we'll take that back. Alright, so in the last couple minutes here I just want to point people to the end of this document. This always happens. We sort of leave the BPQ discussion for last, and then we never have time to really talk about it. This is something that I feel we're struggling with in this work party to the extent to which we want to talk about. So, in this Google doc we have some of our discussion comments, and I also copied in the places where BPQ appears in RSSAC037. So again, everyone please take a look at that and if you have any comments on that, enter them into the Google doc.

And then after that there is a couple pages of what I would call Ancillary Information related to this work party, which is responses we got from the root server operators to the questions we sent to them. There is a list of protocol compliance metrics that we may want to include in this that are not already covered, and then there is some text from the gTLD Guidebook which talks about how gTLDs are measured, and it's actually quite simple, it's all just latency based stuff, and it's got the numbers for that. So, take a look at that, provide comments.

Staff is going to find out if we'll have dial in for the RSSAC workshop in a couple weeks where we're going to spend a good part of a day, a day-and-a-half, maybe, talking more about the metrics stuff. So, if interested, please join us if you can. Any final comments before we end the meeting? Did I miss anything, Steve?

STEVE SHENG:

No, sounds good. We will send out a Doodle poll for the next call, probably after the RSSAC workshop.

DUANE WESSELS: Right, so that will be about a month from today, right?

STEVE SHENG: Right.

DUANE WESSELS: Okay. Alright, thank you very much everyone for participating today.
We'll see you next time.

[END OF TRANSCRIPTION]