KURT PRITZ: Go ahead and start the recording, please.

SUE SCHULER: Thanks, Kurt.

KURT PRITZ: So anyway, thanks for joining the call. I know they've been somewhat intermittent. Samaneh's on the call. Welcome, Samaneh. And I think there's … You can see the agenda there. Samaneh and I spoke very briefly before the meeting. So, she has to report some suggested edits to our report after ICANN's reviewed it. So, we'll talk about that and then the plan for publication as I see it. And want to get your feedback there, too. And, too, Samaneh's prepared some materials with regard to our recommendation three, which has to do with persistence. And then we'll briefly discuss our approach to the upcoming ICANN meeting. Is there anything else anybody wants to put on the agenda?

Okay. Great. So, let's just get into it. Samaneh, welcome to the call. Thanks so much for the work you've done so far and joining us again. So, I understand there are a couple edits to the report that you, on behalf of your colleagues, we like to suggest. So, I'd recommend that Sue, maybe, puts the report up where we can each stare at it and you can take the mic. Or, Samaneh, if you'd rather scroll, we can give you control.

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.

SAMANEH TAJALIZADEHKHOOB: Hi, Kurt. Thanks for the intro. And hi, everybody. Nicole, I could maybe share my screen. There are just, I think, one or two points I would like to point out. I have it highlighted and the report open. So, I think that would be easier. Is that okay with you?

KURT PRITZ: Yep. You're already empowered to share.

SAMANEH TAJALIZADEHKHOOB: Perfect. Are you seeing my screen now?

KURT PRITZ: Yes. We are.

SAMANEH TAJALIZADEHKHOOB: Perfect. Okay. So, yes. I will start by pointing out what were the … I think there were one or two things that my colleagues within OCTO pointed out from the recommendations. Please just note that for now, we only distributed it within OCTO. And we are hoping that soon, we can also discuss it with our Legal and other teams.

So, I'll jump right to it. In finding one, there is this part that the recommendation talks about—talks about "reputation lists are primarily crowdsourced." Now, I know what it means because we discussed it. But I would recommend to change the wording slightly because I'm afraid others who are not aware of the context of the discussion—of the discussions among us—might get it totally wrong.

The point about it is that from the lists that we are using—the reputation lists that we are using—that most of them are very well-known in the industry. Only a few are actually crowdsourced. And even the level of crowdsourcing is different. Some allow anybody to participate and to put input. Some only allow certain designated members to do it.

So, I would suggest either to make that distinction or maybe use the wording … Instead of "primarily," which hints that most of them are like this, change the wording to "some of the lists" are like that some are some are only by providers themselves, etc. So, that would be one thing I wanted to discuss with you guys. Maybe I'll pause to hear your thoughts and then I will—

KURT PRITZ: Yeah. Pausing is good. Any comments from the crowd? Rick?

RICK WILHELM: Thanks, Kurt. So, Samaneh, could you talk to us about which of them are not crowdsourced?

SAMANEH TAJALIZADEHKHOOB: For instance, abuse.ch, where we get the majority of the botnet data and also some of the malware data. It doesn't allow for any member to just input domains. It goes through a certain process of … Members can put their recommendation but it goes through a process of the admins of abuse.ch extensively checking, as opposed to some that just—that you can input.

For instance, if I'm not mistaken, I think PhishTank is like that, that you put your … Anybody can input their domain and they don't have extensive check. Or, for instance, as far as I'm aware … This might be a bit outdated. I have to check again. My information is from like a year ago. But Anti-Phishing Working Group also doesn't allow public to provide domains, as in they have also certain measures in place.

KURT PRITZ: Any response for her?

RICK WILHELM: Yeah, Kurt. Thanks, Samaneh. I think that … But even the first thing that you described was crowdsourcing. There's vetting on back of the crowdsource. Some of them are … The reputation lists that DAAR uses—which is why I think that I'm comfortable with this—they are primarily crowdsourced, the ones that DAAR uses. Some of them are vetted and others are unvetted. And I think that you did a good job there of highlighting the distinction between some of those like PhishTank that are unvetted. And some of them like abuse.ch, based on your description, is a vetted. But it's still crowdsourced.

So, we can look at the possibility of doing an edit there but I think that the fact of the matter, what it says in the rest of the sentence there is that it … and sometimes it might be based on review and [computation] by the reputation list provider itself. But the bulk of … I don't think it's really disputable. But we're happy to look at the

data. But the reputation lists that DAAR uses are primarily crowdsourced, in terms of their sources.

So, maybe what we need to do to get—if you'd like some more nuances … If you could give us—go through that list of what the reputation lists are and how much crowdsourcing they have versus vetting, etc., that would probably be helpful, if you'd like a more nuanced comment. But I think that from the standpoint of the group, from the standpoint of what the words say, I think that we can stand by what it says. Thank you.

SAMANEH TAJALIZADEHKHOOB: I see your point. Thanks, Rick. I see your point. But that is where I'm not sure. Again, I haven't looked, myself, at the whole list to be able to be very confident in saying that the bulk of the list is not crowdsourced. So, I am not confident to say that. But I am not very comfortable with saying that the bulk is crowdsourced. I have the same comments from my colleagues within OCTO. So, maybe, as of now, I can just—or you or I can just check the list and see at least … If you want to stay with this sentence, then we can cross-check that—if the bulk is actually crowdsourced or it's primarily crowdsourced.

In addition to that, I agree. Even if the bulk is crowdsourced, which we can check, I suggest to add other details to it that … Indeed, like you said, there are different levels of vetting process on it—going on within crowdsourced lists. But as of now, yeah. Maybe I'll just highlight this to be checked. That would be my suggestion.

KURT PRITZ: Thanks, Samaneh. I have something to say but does anybody else have a point to make on this one? So, I think this. One is, we're trying to make a broad point here. So, I don't think, right at the top of the report, we want to dive into some level of detail on individual reputation lists or different providers. But we're trying to make this big point. And I think the wording is fine but I think the point could be just as well-made if the wording is softened a little bit.

And I also think that if Samaneh and the ICANN team have pointed this out, then others will, too. And this is not the sort of thing we want people arguing about in the report. We want people to get into the meat of it and not start off with a criticism. So, maybe we could just change that first clause into, "Reputation lists use different sources in formulating their reports: many are crowdsourced, some of the crowdsourced data is vetted, some is not." And then, just go ahead with the rest of the paragraph as is.

So, we could just soften that first clause to take out some of what might be controversial. And that would be better for us in the report, to avoid that sort of controversy. Is there any comment to that? I've got so many different screens going, I can't see hands. Any hands? Any type in the chat?

SAMANEH TAJALIZADEHKHOOB: I'm okay with your recommendation. I just know because every time we present this work, there is so much misunderstanding—not about this specific issue. Actually, I never had a problem with this—but about whether ICANN itself is making reputation lists, are the reputation lists from third-party

providers, etc. And since we are working in this—one of the main purposes of this recommendation is to make everything more clarified in the reports themselves—I thought it a good point not to create more confusion for a part of the community that know very little about these RBLs.

KURT PRITZ: Thanks, Samaneh. Go ahead, Rick.

RICK WILHELM: Sorry. Thanks, Kurt. Samaneh, would it be possible for you to give us a list of what the sources are and their level of crowdsourcing? The list is public but the level of crowdsourcing, I think that that says something. Because we could certainly talk about adding a qualifier, that there be different levels of vetting in this. But I think it is a fact. And this is something that … It's not just me that has brought this point up but others in the group, that a lot of this data is crowdsourced and it is primarily crowdsourced. And I think that's a factual statement.

So, while we might put a few words in there to soften it a little bit or something like that, I think that fact of the matter remains. And so, if you could provide some data that describes the level of crowdsourcing, maybe that would be helpful for us. Thanks.

SAMANEH TAJALIZADEHKHOOB: Sure. And don't get me wrong. I'm not suggesting … If this is true—if it is primarily crowdsourced—I'm not suggesting to present it otherwise. My point was that it might not be true. And

indeed, I will take a look and then we can discuss it later, or add it, or change it. Anyways, it has to go through another round of internal ICANN. So, maybe, if it is different, we can edit it then.

KURT PRITZ: Right. Let's go on to the next thing. Thanks.

SAMANEH TAJALIZADEHKHOOB: Yeah. So, the next part … Actually, the only other comment was about recommendation three. I will discuss this when I'm going through the work that I've done about it, if that's okay with you guys.

KURT PRITZ: Yep.

SAMANEH TAJALIZADEHKHOOB: Yep? Okay. So, I will just read it out again, for those who cannot remember. Recommendation three talks about creating a measure of persistence of reported abuse activity. It recommends using—measuring average length of time a domain exhibits reported abuse activity or/and describing characteristics of the population distribution, such as standard deviation and the shape of the distribution, I guess.

So, the point that I have, and also my colleagues made here, is that it is a bit unclear what we mean by "population" here. What are we talking about? Because we have data over a long period. We have data of individual domains but we also have data of …

We can also aggregate it over a gTLD. We can aggregate it over a period of time. So, the difficulty, and also maybe the interesting part of this recommendation, is that the level of analysis that is recommended here is a bit unclear.

Now, all of them are fine. We can do all of them. But then, I think it could be more clear in the text, what the suggestion is about. Do you guys have thoughts about it? I think we can also discuss it … We have discussed it previously but it is somehow not added yet or maybe was not clear, either.

KURT PRITZ: I can comment on this. But would anybody like to make a comment first? So, yeah. So, this is my writing which is why it's so unclear. So, it could use some more detail. And it was short, just in the interest of keeping it short. But the idea is that if the average mean time—the average half life of a domain or whatever—before the abuse is mitigated is, say, four days, we don't really have any idea if that's good or bad. Some people will say, "Four days? That's great!" And other people will say, "Four days? That's horrible."

But what really matters is the distribution of that—the four days. How is the four days calculated? Is it a normal distribution, around four days? And this is for all TLDs—starting with all TLDs, all abused domains. Or is it bimodal? For example, is 80% of them mitigated within one day and then 20% of them mitigated in 10 days? Which indicates another sort of way to address the problem. That 80% is excellent and we have to see what's going

on with those other 20% and see if they're false positives or whatever the heck they are.

So, we thought that just—or at least I thought just the mean was not a particularly meaningful measure but only when you look at the distribution of all the abused domains. But possibly in answer to your question, we were thinking that this was just the total population of all abused domains.

SAMANEH TAJALIZADEHKHOOB: Yeah. I understand.

KURT PRITZ: Samaneh, before you answer, I want to check with our team and makes sure that my thinking is in alignment with others. Okay. Go ahead, Samaneh.

SAMANEH TAJALIZADEHKHOOB: I think I understand your point. Indeed, mean would be insufficient to describe all of the corner cases of the population. But for instance, we will see … We've discussed it also before that we will see how this metric would evolve, whether we mean to fit it within the monthly report or just the one-time analysis. The issue I see with it is that … So, we could, for instance, map out—and I will show you this in a few minutes—the distribution of the amount of time—let's call it "persistence metric—" for a domain over time or for all the domains that appear within a timeframe that we define.

But then, in order to bring … Then, we have so many domains and we have so many months. In order to be able to bring it to the monthly report and have it at the same level of analysis as we have in the report, which is gTLD at the moment, we need to have a metric that is aggregated over gTLD. And so, that would result in a distribution for each gTLD which still is not feasible to have 1,200.

So, I would say this is something to be discussed—first of all, how feasible it is to create such a metric and how feasible it is to bring it to the monthly report or some other kind of analysis. Just wanted to point out that … Do you guys think such a language is needed to be added to this recommendation—that this is maybe not for the monthly report or maybe some language about the level of analysis—to be able to capture this difficulty that we have or that we'll face in any kind of analysis that we want to do with this metric.

KURT PRITZ: I have a comment to that. But anybody? I kind of think there's a good segue into what the next steps are in the publication of the report. I think, like we've said in the past, the report isn't the end. It just a waypoint. And how we go … There's several ways of going about this, and figuring that out could take some weeks, and that's why I think it's perfectly fine to … If the group decides to, I could easily add some more detail to describe why this is or put in a footnote explaining why we're interested in this that lays it out for ICANN a little better and provides a little bit better direction on where we're going.

But we think that the next steps will be us, the DAAR group, or the remnants of us who hang on, will continue to work on improving the report. So, we want to make the report clear to you so you understand what the recommendations are but not necessarily work through all the details of how that's to be done. As you were talking, a dozen items were raised in my mind for interesting ways to approach this and a much longer discussion. But that's not the place for this.

So, I think what I'd recommend to our group is that we might put footnote here or flesh this description out with another sentence, telling ICANN we're interested in looking at the whole gTLD—or the whole population for this sort of thing—or why we're interested in this data, an explanation for that.

So, those are two possible edits to the report. One is we're looking at the whole population and two is the reason why we're interested in this. Is there any support or nonsupport for augmenting the recommendation in this way? Or alternatively, we can leave it as is, knowing that we're going to continue to work on this. When I was a young manager, I learned to use silence as a management tool. Thanks, Rick! Go ahead, Rick.

RICK WILHELM: Thanks, Kurt. As far as this goes, I don't think that we've got … I agree with your assessment, that we were trying to understand … When we see things that have abuse characteristics, are they … How do they … What does that look like? I think that you did a good job describing it.

And so, our use of particular statistical terminology here was meant to be illustrative. And we likely could have just stopped the sentence at the end of the word “distribution,” right? Because what we were trying to understand is, for example, is there a bunch of names that just pop on and off the list or do they—and then some of them hang on the list for a very, very long time? Or do all names pop off the list very quickly? Or do all of them hang around forever and it’s really a static set of names?

And so, I think … The audio’s kind of choppy for me. But I think maybe just elaborating what you were saying, that really, our notion of using the statistical terms was not meant to be specific, or declarative, or a statement of requirement—just a way to characterize the population distribution.

KURT PRITZ: Right. We’re not intending to be prescriptive here.

SAMANEH TAJALIZADEHKHOOB: Yeah. I understand your point. Okay. My only concern is that once … Obviously, I understand what it means and we are going to work together to develop this. So, I don’t see any problem in that area. Maybe because I am more into the details of the work, I am concerned that this will create expectations of why we don’t have such a—why we are not including such a method in our monthly report, whereas even including what you just described, of this characteristic of the population of domains, is difficult to include at that granularity in the report. But I guess it can just be there and we’ll figure it out later. I am not—

KURT PRITZ: Yeah. And I'd be fine with actually shortening it to give you some more freedom in figuring out how it would work out.

SAMANEH TAJALIZADEHKHOOB: Yeah. Maybe that's a better …

KURT PRITZ: So, go with Rick's description there?

SAMANEH TAJALIZADEHKHOOB: Yeah. I would think. That will give us more freedom to decide what we are going to do—if we are going to include in the reports or just other shapes but the persistent metric. Yeah. I think that's a better suggestion.

KURT PRITZ: Okay. Thanks very much, Rick. Anybody else? So, I think before we get into the actual work that you've done, which is the more important part of the meeting, I just want to use this as a segue for our plan for publishing the report. So, we, just to be as frank, or identify our goal … So, we are very appreciative of your review of the report. And the two points you've made are really valuable and we appreciate the lightweight approach you've taken to it.

We have to be careful that this is a Registry report and not a Registry ICANN-issued report. I think it's more powerful for us in the community to know that this was an independent review of

DAAR. So, I think we don't want a final buyoff or approval from ICANN on the report. Our plan is to go ahead and publish it. But we do want to ensure that we're not getting in the way of any work you're doing. Or, like you pointed out in this last example, some of our words might set up expectations and have others taking up your valuable time with talking about that, rather than improving on the report.

So, our plan for publication of the report—and we're leaving it up to our Registries Stakeholder Group leadership to figure out the means for that. But as we've discussed in the past, we don't see the issuing of the report as an earthquake event, just a work in progress. So, we intend to publish it as a waypoint and as a signifier—as, "This is the work the work that's been to date and we're going to continue to collaborate with ICANN in this really positive way going forward."

And so, right now, this report is with the RySG. And we'll make some tweaks to it based on your recommendations here. And then, we'll let you know. We'll keep you informed so you're not surprised at all by its publication and you have a copy of it to share with the rest of the ICANN staff before it's finally published.

SAMANEH TAJALIZADEHKHOOB: Okay. Yeah. That sounds good to me, as long as … And I mentioned this in the brief talk I had with Kurt before the meeting that John sent apologies for not being able to participate. And the main request he had was to be able to discuss this internally before it goes published.

KURT PRITZ: Okay. Great. Any more comments about these edits or the report? So, I would … I'm going to send this report with the two edits in it. And we'll get comments back from the list. Also, Samaneh is going to try to, within a few days, maybe, turn around some additional information about the degree to which reputation lists are crowdsourcing and that'll help inform any final edits.

SAMANEH TAJALIZADEHKHOOB: Sounds good to me.

KURT PRITZ: So, with that, let's go to the more interesting stuff. And Samaneh, you've already got the screen.

SAMANEH TAJALIZADEHKHOOB: Yep. So, I trust that everybody can see my screen. So, the work that you will be seeing in this screen are just my thoughts. And in these notebooks, I just think of an idea and I go ahead and do it so that nothing is final. And I would like to use this opportunity, like before, to just brainstorm with you guys, and get feedback, and we'll discuss with each other what does it mean, what I have, and how it can be improved or be different, etc.

So, we talked about the persistent metrics for the domains that are reported. I thought about it and even the definition itself, I thought, could be grasped in two different ways. One way is to define "persistent" as the number of days or the amount of time that the

domain remains on blacklists. So, we could … If we track blacklists over time, which we do on daily basis, we could just see how many days the domain remains there.

Another definition would be for what amount of time the domain is distributing malicious, or hosting, or contains, or distributes malicious content. The two might be the same but might also be different because each blacklist tracks their domains differently. And some of them have additional datapoint that is called, for instance, the date that they capture an entry, a datapoint—let's call it "domain" here—and the date that the content—that they remove it from the list.

By that definition, it means that when the remove a domain from the list, it means that they are tracking every once in a while and they know that the malicious content is not up there anymore or the domain itself is down. So, they remove it from the list. Whereas for the first, if we only check by ourselves, manually, if a domain or when a domain is removed from the list, it could also be that the blacklist provider is not tracking it anymore and not necessarily that the malicious content is not up, etc.

At the moment, we only have access to the first. Actually, I made a note here. Maybe I—part one and two. So, for the moment, we can actually only measure this. We have this information for all of our reputation lists. And this, we only have for some. So, I wanted to make a clarification of what "persistent" means in this analysis that I have done. Are there any thoughts or questions about that? If not, I will go ahead and start with the analysis.

KURT PRITZ: Go ahead, please.

SAMANEH TAJALIZADEHKHOOB: Okay. So, this is … I am not using the whole abuse data because that would only reduce processing power. For the sake of discussions with you guys, I am just taking a sample of data. And whatever we come up with, then I can extend it to all of the data.

So, this is phishing from a reputation provider called PhishTank. And I am only taking a month of data. So, when I plug this data … So, basically, I count the number of days that each domain is on the list for one month—for month July. And the distribution that shows is basically this or this. Both of them are the same data. This is just a bit smoother.

What you see here is that there are majority—or, well, a lot of domains that remain on the list between one and five days—between one and four days, I think, if I want to be precise. And then, we see an exponential distribution. So, it goes down as … So, basically, little domains are up for more than 15 days. And then, there are some that are up for the whole month.

My guess was that this is an original exponential distribution. So, we should not … Actually, we should … This is an exponential distribution that is capped because I only took one month of the data. And if I look at longer period, this part would be smaller.

Then, I took a look at all of the PhishTank data that we have collected, and indeed—or we had collected on this internally because I am not … I am using the ICANN internal data, which is similar to the DAAR data. But I didn't have any good access to

that. So, indeed, if I look at that, then you see that this peak becomes smaller in comparison to this. And if I increase the amount of time … So, this is 80 days within 2020. If I increase this, then these would be even smaller. This is an effect of capping the data with certain timeframe.

So, basically, what we see is an exponential distribution, which is sort of expected because that is what … Also, I have done work in this area before, for botnet, command-and-control domains, and the amount of time that they survive. It's called "survival [feeds]," in which majority of domains die because the content is crucial to be taken down. And some of them remain on the list.

KURT PRITZ: Samaneh, can I stop you for a second and ask Rick to make a comment?

SAMANEH TAJALIZADEHKHOOB: Sure.

RICK WILHELM: Thanks, Kurt. And this is very brief. If you could scroll up a little bit to the previous set of graphs. It's 25% of the scrollbar in the notebook. Yeah. Right there. When we were in the … Just to tie it back our previous comments on the second point of feedback, when we were talking about the characteristics of the distribution, this, at a qualitative level, is exactly what would be the sort of thing that's interesting—right, what we're seeing here. Like you see there's this big clump at the beginning. Then, it tails off. And then,

the question would be what's going out with that big spike at the end—that sort of thing.

So, I just wanted to relate and bring some color to our comment from the previous topic. This is exactly, I think, what we were thinking of. I'm happy to take others' comments. But when I saw this graph pop up, I was like, "Oh. That's what we were looking at." There's a bunch of them that come and go very quickly. Then, it tails off to about 25. And then, there's this pile that's 30 plus. And so, what's going on with that. And so, that was what we were getting at. I'm happy to take others' comments on that, too. Thank you.

KURT PRITZ: I agree. One question I have, Samaneh is … We thought this data would be taken—and maybe it is—but we thought the data would be taken when the domain comes off. So, where you're showing this for one month … The lifetime of a domain can't be measured until it's no longer there, correct? So, I don't …

SAMANEH TAJALIZADEHKHOOB: Sorry. I think I didn't fully … You said the lifetime. You mean when the domain is up on the list and off, out of the list?

KURT PRITZ: Right. So, in this first bar, where it shows a zero to two days, or whatever that is, is that just … Were those domains already in existence at the start of the month? Or are these domains that were registered during that month?

SAMANEH TAJALIZADEHKHOOB: Yes. That is also … Actually, that points into another difficulty which this kind of analysis has, which is no problem. I'm actually happy to go through all these corner cases. And I think these are actually interesting pieces of this analysis. But the thing is that when we talk about lifetime of a domain, we have to have a starting point, right? Because first of all, the domains appear and disappear. They go off for a while and they come up again, right?

And that we could … Let's say that a domain can be on the reputation list for a very long period. How we define it matters. For instance, we could say we cut it at … We say, "This is our starting point," in terms of time, right—point in time. Let's say we say we are looking at all the domains, from January 2020. And then, that would be t zero. And then, from then, we say, "Okay. When all the domains that appeared on the list from t zero went off the list." And we measure lifetime like that. And I'm happy to hear if you guys have other opinions or thoughts. Also, to—

KURT PRITZ: Yeah. I'll speak for Maxim, who's typed into the chat. But you can follow up, Maxim. So, these are the dates when a name comes off the list. One of our questions will be … This is why this is a long, iterative process. Our next question is going to be when were they actually mitigated? When were the names, maybe, taken down or some other action taken—put on hold—versus when were they taken off the list? So, that might be the next dive into the data. Did I say that right, Maxim? Yes.

SAMANEH TAJALIZADEHKHOOB: So, in order for me to understand what you just mentioned, you suggested that to be able to measure lifetime of the domain, that we just check whether they come up again, if they go out of the list?

KURT PRITZ: Yeah. So, there is … A reputation provider flags a domain. A registry and registrar take action. The domain is put on server hold. And then, sometime later, the reputation provider takes the name off the list. And so, these don't necessarily indicate how quickly the registry and registrar reacted. These indicate how quickly the reputation provider reacted. And I understand there's … Clearly, I understand the difference between the two, and so do you, and that the data for the one might be much harder to get at than the data for the other. But I think that, as we work on this, that might be the next level of analysis.

SAMANEH TAJALIZADEHKHOOB: Indeed. No. I fully agree. And that is what I tried to explain in the beginning, that these two different things … And what we are seeing now, exactly like you pointed out and Maxim also, is just the reputation lists' reaction, which is different, even from list to list. And I want to also be very careful with articulating this towards the community, as in lifetime of the domain, because the two are not equal. And then, we will get a lot of reactions that they are not equal.

And I agree. They are not because we cannot. Like you said, we cannot be sure how much time is between when the domain is actually taken out and when the reputation lists react.

And to also respond on Rick's comment, thanks for this. It's very good to know that this is what you guys had in mind. I noted this. The only thing is that also, at the moment, I am showing you an analysis at the level of all of the domains. So, I am not making any distinction between any TLDs. This is just the overall distribution of all the domains in this list. Of course, in the report, we can also just talk very generally and qualitatively about all domains. But as soon as it gets at that level of gTLDs, then it will bring a lot more complications, which is fine. We can grasp it later together. Are there any other questions or comments?

KURT PRITZ: No. Go ahead, Samaneh.

SAMANEH TAJALIZADEHKHOOB: So, yes. So, one of the interesting things I wanted to discuss with you was this effect—this cap—which will get smaller as I expand the timeline of the data. Now, one point for me to discuss with you later is that what kind of timeline do you guys recommend? Do we take a month? Or do we talk about a year, etc.?

Another recommendation was aggregates from the distribution—so, things like standard deviation or median. For instance, the median number of days that a domain is on the PhishTank is 12 and the standard deviation is also 12—interesting—which is for a

month. This is for the month's data—so, 12 days, which is not … It's actually surprisingly high. I was surprised by this.

Again, this is only for one list. So, as soon as I start to talk about this, it matters from one list to list. Later, I checked this for, I think, three months or so—maybe less than three months—of PhishTank data. And in that case, the median number of days will become six and the standard deviation is [27].

I also looked at spam data—spam data from Spamhaus. And interestingly, we see a slightly different distribution. Maybe if I have the histogram of this it will be more clear.

So, here, we see the data points are more. So, you see little when I have the histogram. But what is different from phishing to spam is that in case of spam, apparently the majority—and a big majority—is taken off the reputation list very quick, in less than, I would say, five days or so, looking at the distribution. Here, you also see that there are some bounces here and then it becomes very quickly flat. There are also some bounces here at the end but they are very minor. Did I take the …? I can also just quickly look at the median value for this. The median value is one day, for the amount of time that spam domains remain on the list.

I would say that the type of security threat also reflects a bit on whether the registry or the registrar is the one that—sorry, whether the registrar or the hosting is the one that is taking the action. I'm not sure if these points can be made in the report because then it will create a whole new discussion of who is responsible, which was are not addressing in the DAAR reports. But for instance, I would say if we say that phishing domains are

mostly newly-registered domains, they could be more on the registrar side than the type of security threat that feeds on compromised websites.

And I'm saying this to say these could be some of the reasons why we see different distributions, in terms of amount of times for spam and phishing.

That is the analysis that I've done so far. The next steps that I had in mind was to dig in at the level of gTLDs, to see—to, for instance, highlight top-end gTLDs, in terms of amount of time, and dig into those to see what is different in those gTLDs in comparison to others that had less median time. Again, I'm just explaining my thoughts. I'm happy to know if you have other ideas—other things that you want me to explore, that you think are more interesting, etc.

KURT PRITZ: So, please raise your hands, you guys, with questions because I feel self-conscious here. But on the one where you were surprised at the size of the median or the mean—it was either six or 12 days—does that include the bimodal distribution? Does that include all the days at the right end of the chart?

SAMANEH TAJALIZADEHKHOOB: Yeah.

KURT PRITZ: And so, it'd be interesting to know what the median is, or mean is, if you whack that off. I'm not asking you to do this now. So, this is … I echo Rick's comments that this is really interesting and great and raises a dozen questions in my mind so even more in others that are smarter than me. And anyway, it's just indicative of the additional work to be done.

I would say, for me, maybe the more interesting thing, before we dive into individual TLDs, might be to look at what's out there in the 80-day range or what's that residue number that's left and why is that there—to dig into those because I think that would … If this were displayed publicly eventually, that's the thing that would grab everybody's attention. Why the heck are these domains still in existence? So, rather than focusing TLD by TLD, I'd be more interested in focusing on what those long-lived TLDs are. But again, I'd really be interested to see what anybody else thinks. Rick?

RICK WILHELM: Thanks, Kurt. Regarding, the long-lived ones, it might be that they're misclassified as being problematic. It's a possibility because there are, sometimes, when you look at these things … And I know that as I've gone through abuse data from time to time, based on stuff that pops up in feeds, stuff that sits there for a long time is maybe inappropriately tagged and it's just hard for it to come off. So, that could be one possibility.

But I agree with Kurt's question about, "Well, why are these things hanging on if they really are that problematic?" And so, it might be

that they're not being taken down or it might be that they're not harmful and that they're just miscategorized—possibility.

SAMANEH TAJALIZADEHKHOOB: I agree. And you make interesting points. I have to … Some of this is actually hard to check, right? I have to think of it. But it would be hard—difficult to check whether they are actually miscategorized. But I will dig into some of the details of this. Maybe some interesting things would pop up. I had an idea but I forgot. Well, you guys go ahead. I will remember it.

KURT PRITZ: Any comments from anybody else?

SAMANEH TAJALIZADEHKHOOB: I do think, while I also started with this analysis, that this, by itself, could be just … This is a whole new domain of analysis that has a lot of meat to dig in. And it could be just a type of different white paper or research paper published, just about that. And, of course, it could be repeatedly, also. But it has more to say than just the monthly reports.

KURT PRITZ: Right. I agree. And I think that the DAAR community is sophisticated enough that you could develop this some more and say, "This is where we're doing research now," without having to be done, done, done and get everybody's interest and approval that you're taking the report in a direction where it's going to

provide more meaningful information. So, I think it can be done. You're right. It's a paper. And then, there will be papers after that. And I think you can take the community along for the ride, rather than try to wait for a perfect product at the end. So, I'm cognizant … So, thank you very much, Samaneh. This has been a terrific meeting.

SAMANEH TAJALIZADEHKHOOB: Thank you.

KURT PRITZ: So, I'm going to pause for a second to see if there are any other comments. So, listen. We only have two minutes left and I've completely forgot about Sue's need to report to the RySG or the ICANN Meetings Team about our plans for the next ICANN meeting and whether we want to have … I guess there's a choice whether we want to have a public meeting—so-called public meeting—at that meeting or not.

So, in the interest of time, my recommendation to this team is that James, in his next report to the RySG … There's an RySG meeting tomorrow. So, we could bring this up. I think this DAAR discussion is one piece of the DNS abuse puzzle. So, the RySG should have an integrated approach to how it approaches the meeting. So, I'd like to hear anyone's comments about how we should approach the upcoming ICANN meeting and whether we should do something different, other than bring it up for discussion in the meeting tomorrow.

SAMANEH TAJALIZADEHKHOOB: Maybe I should just announce that we are happy … If there would be a public meeting with us and you guys, we are happy to participate. I think it's a good idea. But whether it should be done or not, maybe it's more on you guys.

KURT PRITZ: That's great feedback. Thanks, Samaneh. Anyone else?

SAMANEH TAJALIZADEHKHOOB: Kurt, just before finishing and before I forgot, I also had a point that I forgot to let you know in the brief discussion we had before, just for you guys to discuss later when you meet, is that we need and we want to have feedback from you on the ways you recommend to include ccTLDs in the DAAR reports. So, that's just a topic for future discussion.

KURT PRITZ: Thanks. Sue, could you tell me what you just typed in? You have a meeting penciled in at this time on the 14th of …

SUE SCHULER: Of October. Yeah. So, I have to have this done by the end of this week. So, I am penciling the DAAR group in on Wednesday, October 14th at 10:00 UTC for ICANN 69.

KURT PRITZ: 10:00 UTC? That's not this time.

SUE SCHULER: No. It's not.

KURT PRITZ: But you said "at this time." You said, "at this time on the 14th." And I said, "Good."

SUE SCHULER: If we do it at this time, it will be off-schedule because the meeting's taking place during the Hamburg time zone and it does not extend to 15:00 UTC. So, we could still be on the public schedule if you decide to hold the meeting off their recommended time zone. But just so you know, people will be up, and awake, and working during those times. So, having it off-time, you may lose some attendants.

KURT PRITZ: Not me.

SUE SCHULER: I understand.

KURT PRITZ: Right. Are there any other suggestions about how to approach the ICANN meeting? Thanks for your comment, Sam. I tend to agree. Okay. So, Sue, we'll try to settle this during the RySG meeting tomorrow.

SUE SCHULER: Yep! Thanks.

KURT PRITZ: All right. Samaneh, thank you so much for the work that's been done. And thanks for your participation, everybody. Have a great rest of your day.

SUE SCHULER: Okay. Thank you. Julie, we can end the recording.

RICK WILHELM: See you, all.

SUE SCHULER: Bye, all.

[END OF TRANSCRIPTION]