JOHANNESBURG – Tech Day (Part 2)
Monday, June 26, 2017 – 15:15 to 16:45 JNB
ICANN59 | Johannesburg, South Africa

EBERHARD LISSE: I didn't want to interrupt you. I just wanted to interrupt you. Anyway, good afternoon. Everybody in the back can sit down. We are starting with the second and/or third session, depending on whether we need a transition break or not.

Matthew Zook approached me a while back. He wanted some data from us for some [maps] he's drawing, so we refused to give it to him. But when he explained to us why he's using and what he's doing with it. We still refused to give it to him, but we thought he could come and bring this to the wider audience. It's quite interesting stuff.

MATTHEW ZOOK: Okay. Thanks a lot. I guess this my attempt for you to give me that data. Anyway, this is – looks way too far advanced. Whoops. There we go. All right.

Okay. We'll keep on going here. Let me just start my timer so I try to keep to time.

The project I'm talking about is something that I've been doing for a long time now. It's been about 20 years. I've always… Oops.

Okay, there we go. I always joked that this is a dissertation project that never ended. If you've ever been in graduate school, you know hellish that actually sounds. This is a project I did back in the late 1990s, looking at domain names as actually an indicator of the information society, the information economy. I've essentially been tracking it ever since, primarily [as] an academic, but I do this work with ZookNIC on domains.

It has become useful in lots of way. Things you've probably seen some of this data, probably the most visible one is in Verisign's domain industry briefs. It comes out every once in a while.

This also has really been supporting a lot of academic work as well. I'll show you some of that at the end because that's really why I got into this, why I continue to do this. It shows up in some things like the GII Innovation Index, which just launched a week or two ago, some World Society, in the Information Society reports, and then some of this work that I've been doing with a couple people, mostly based at the Oxford Internet Institute.

I keep on pushing the wrong button. All right.

Just to give you a snapshot of where we are today, we're about 339 million domains. We might have squeaked over 340. By the end of June, we'll see. You can see the rough breakdown in term of gTLDs, ccTLDs, the new TLDs, and the sponsor TLDs, the biggest ones being biz and info, obviously.

This is the most basic kind of data or the most basic summation that we do. It looks fairly simple in this pie chart, but it's also a lot of work that goes behind here. Just to give you a little more – oops, keep on hitting the wrong button – breakdown, a little more granularity if you want to take a look, you can see how this might be useful, particularly in tracking growth and change over time.

Now, I know this is Tech Day, but I'm primarily a social scientist who studies who computers are used, rather than a computer scientist who studies society. So think of this more of a socio-technical talk rather than a technical talk because a lot of what I do is not particularly robust or all that interesting from a technical standpoint of view.

These are the three main ways I get my accounts for that graph I just showed you, looking at zone file analysis. Also, the ICANN monthly reports are quite useful for this. But this primarily for gTLDs and TLDs and sTLDs.

Then there's looking at the registry reports. I think a lot of ccTLD registries are out here today. The various kinds of counters, press reports, their own analysis that they post online have been extremely useful. It's not possible to do it without this. And then, most recently – well, over time as well – there's directing queries

to the registries, asking simply, “Could you give us some data and let us know?” Hopefully people will say yes. Maybe not.

Anyway, the zone file counts in terms of doing this is pretty straightforward. I’m sure everyone in the room is familiar with this. Because of the ICANN contracts, certain TLDs are obliged to provide access to zone files. It’s just a matter of going in and extracting a complete list of currently active domain names through sometimes some issues of domain names that are moving in and out of the zone and so forth. That’s become less of a problem over time as the files are updated pretty much in real time.

[We] do a lot of sorting/deduping just because it’s one of the things to do to make sure you get an accurate account. But the thing that’s really useful in having these kinds of lists is that I essentially have a snapshot in time and do comparisons and look at zone files – adds, deletes, and so forth – within a particular TLD.

The other thing is actually quite interesting, or that I find quite interesting, is what else you can add to this kind of information because, in addition to the list of domains, you can also get lists of the root name servers for a particular TLD and then, using some other fairly simple tools, you can geo-locate these things based on IP addresses and make something like this.

This is something I did about ten years ago, so it's not anywhere near accurate in current time. This is essentially comparing where these different TLDs headquarters are in terms of where their business locations are and where the root name servers are located for that particular TLD. The white lines are showing this connection between the headquarters' location and then more the technical or the name server location.

All kinds of interesting visualization I might do on this. Of course, you want these things, for a robust network, nicely distributed. I did this also for all the TLDs – ccTLDs as well – but by the time you put all those on, it's just this mass of white lines – spaghetti. Not very easy to interpret.

The other way I get for the zone files is using the dig access for other TLDs, primarily ccTLDs, where access is available. I essentially follow the same procedure for counting currently active domains. It also has the ability to look between snapshots in time and see how adds and changes might be taking place within a TLD.

Now, for ccTLDs, though, the primary way of getting this kind of information is from the registries themselves – literally from A to Z, or in this case, from .am to za. It's appropriate that we're in South Africa for this particular one. There's a lot of these kind of

either counters or monthly reports or quarterly reports that ccTLD registries put up.

I want to make sure that anyone who's associated with this who thinks no one ever looks at this stuff, let me assure you I'm always looking at this stuff. I'm quite appreciative of it. I love to talk and thank people who are actually the ones behind this because this is really, really useful stuff in doing this. And it's a fairly labor-intensive process, again. There are no really good ways to automate this because things constantly change. Once you get it fixed up, it might change two months later. You multiply that over a bunch of ccTLDs and it becomes problematic.

There's lots of other registry reports, like the .au, which has a monthly report. .eu does a quarterly report. They're quite useful. There's some really interesting thing. I have to say, by doing this, I'm probably someone who has gone to individual ccTLD pages regularly on a longer basis than anyone else. There's some really interesting stuff out there, so I'd love to have a conversation with folks about this.

This past year, I've actually been living in New Zealand. I've been talking a lot to the .nz folks there, and they're doing some really interesting stuff with their data. Again, as you know, sitting on top of this there's all kinds of interesting data you might have.

Think about what you might do in order to understand your user base, your slice of the Internet more generally. There's some really, really great resources out there.

The final bit is really just inquiries to registries. We actually in the past two months sent a lot of e-mails out to various contacts we found at ccTLD registries for those that we hadn't had a recent count on, explaining the project, our overall idea, and what we're trying to do with it, and then asked for accounts. I want to say many thanks to anyone who's replied, even if the reply was no. It's nice to have that interaction and know what's going on with this.

Again, one of the reasons I'm here at ICANN 59 is to try to continue and follow up on some of those conversations. If anyone wants to talk about this, please see me after this talk or sometime during the rest of the week.

Now, the question then becomes: "Okay. You've spent time. You've done these various things to get it, but what can you actually do with this kind of data?" I talked a bit about that. It shows up in Verisign's domain name brief. But there's lots of different things you can do. There's a lot of ways you can slice and dice this data. I've done some of this. I've not done as much as I might, mainly just because of time constraints. But there's

lots of interesting things one might do, including this basic history of growth of domain names from 1991 to the present.

Again, it's a very simple graph, breaking down the four different basic TLD types. There's differences within those, obviously. It's quite remarkable to just be able to see the various points in time. There's big events within the domain name space. You have the initial dot-com boom in the late 1990s. That was followed by the bust. You can see the flattening out of the gTLDs, that dot-com hangover taking place. Then you see in the first decade of the 21st century the ccTLDs expanding quite a bit during that time. There was a global financial crisis that put things into a bit of a stop. Then there's the most recent launch of the nTLDs.

It's a pretty remarkable track record. When I first was looking at this, I was checking my notes before this talk. The dot-com zone was 1.5 million domains. I just remember going, "Oh my God. I have to get a bigger hard-drive to deal with this thing." Now we're up to 339 million domains and dot-coms of 130 million or so. It's been a tremendous growth.

So that's one thing you can do. You can slice it some different ways, looking at the share of the TLD over time. Again, all these lines show relative share to the overall domain space. All these different categories are growing over this time period. You can

see that period of time where the ccTLDs really were expanding in the first decade of the 21st century. gTLDs have gone down, and then there's the more recent growth of the nTLDs.

So there's all kinds of different ways one might do this. You can zoom in on regional case studies, country case studies. You can look at IDNs if you're wanting to. You can do all sorts of adds and deletes and look at renewal rates. I already mentioned some of the really interesting stuff that some individual ccTLD registries, like .nz, are doing in terms of looking at this kind of data.

The other thing you can really do is use it as a metric for studying the information and knowledge economy more generally. Again, I bring this up because this is really where my interest lies in collecting this data because it's a very useful metric. It's how I first started on the whole study of domain names.

I just want to give you a brief overview of some of things you might do with this. These are these published reports I mentioned already. The Global Innovation Index of 2017 was just released about a week or two ago. There was the target looking at the Information Society targets. I believe this is being put forward by the United Nations. Look at it that way.

Thinking more specifically and actually building off some research I've done most recently with this, I'm looking at how it

might act as a metric for studying development. The whole tie between ICTs and development is one that's really interesting. I think it's really relevant for the bigger idea of what ICANN is all about, the one world, one Internet – or is it one Internet, one world? – idea for ICANN.

I want to start out with this idea that recently there's been a lot more digital connectivity through submarine cables to sub-Saharan Africa, with this expectation that this would bring a democratization of information and knowledge production within sub-Saharan Africa.

The question is, has that really happened? Just to give you a sense of where things were, this data is fairly old. I think it's 2001/2002 data from the ITU. Just look at a simple metric of how much it costs to have broadband as a percentile of the nation's average income. Africa, particularly sub-Saharan Africa, pops out as a place on the earth's surface that has a relatively high cost of relative of relative to income. This is essentially what these new submarine cables hoped to be addressing with this.

Again, this ties into all kinds of other metrics that you have. Relatively fewer Internet users by country. This is a map called a cartogram, where it's just sorting the size of a country relative to whatever is filling the country. In this measure, it's number of

Internet users. The percentage of people online is there as well. The lighter spots indicate fewer users per capita.

You can also look at it just in terms of domain names as well, looking at where domain names are concentrated. Again, this region of the world remains relatively unrepresented within the domain names space.

There's a couple papers here I'm combining. The citations you can see at the bottom if you want to take a closer look at it. Essentially, we're using domain names along with some other indicators. I think the next one is the GitHub commits. We also used Wikipedia entries. We also looked at a few other things as well, but those were the main ones that we ended up using, just to get a measure of the information economy because, when people are doing this within development studies and so forth, you also use indicators such amount spent on research and development or patents or things like that. We preferred using these kinds of metrics – GitHub commits per 1,000 or domain names – because it gets as knowledge production or content production or those sort of ideas at a much lower level with more granularity. You might be participating in knowledge production without getting a patent and trying to get at those sort of questions.

Looking at this – I'm going to skip over a lot of the details, but for those who are interested, we use a multivariate model/ordinary [least] squares, looking across the globe for these different indicators – the story is really summed up in this graph. In terms of the number of Internet numbers within the region – again, sub-Saharan Africa is marked by yellow in this graph – you can see there's a certain size relative to the world number for Internet users, but it's much smaller when it comes down to these other metrics that we're thinking of as being indicators of knowledge production or involvement in the knowledge economy – domain names, Wikipedia edits, or GitHub commits.

One of the summaries from this that is that, really, contrary to the hopes and expectations, connectivity alone falls short. It's obviously a necessary condition, but it's not a sufficient condition. One of the things we were trying to do was comparing these new metrics of knowledge production to a whole range of other metrics of knowledge production. Patents and R&D were some of those I already mentioned. We were also comparing it to more traditional measures of knowledge generation by academic articles published.

What we found is that, relative to these other things, the sub-Saharan region was underperforming in these new metrics of knowledge production. Again, it's just not a story about connectivity. However important that is, there's a whole number

of other factors in terms of wealth, innovation, capacity, public spending on education and so forth that are really important to help particular places join in the knowledge economy and knowledge production.

Again, I've gone fairly quickly over these studies. I'm happy to talk about them. They're referenced in these slides. But I'm going to end it right there. Thank you very much.

EBERHARD LISSE: Any questions?

Good. Thank you very much.

UNIDENTIFIED MALE: Can I ask a quick question?

EBERHARD LISSE: There is one question.

UNIDENTIFIED MALE: I was struck by one of your earlier slides about a fairly thick slice for sTLDs. Every sTLD I know is vestigial. I was wondering if you know off-hand where those are.

MATTHEW ZOOK: The sTLDs in that are biz and info. Those are the big ones within that category.

UNIDENTIFIED MALE: Sorry. They're the what?

MATTHEW ZOOK: Biz and info. Those are the ones that I put in there.

UNIDENTIFIED MALE: Oh, okay. That's fine.

MATTHEW ZOOK: Maybe they should be listed under gTLDs, but yeah.

KATHY SCHNITT: Hi. This Kathy with ICANN staff. We have a question online from John McCormack. He says, "Matt, have you looked at the IP location of name servers rather than just the root name servers to measure to the development of country markets' infrastructure for ccTLDs and go gTLDs? The more mature the market, the more domains hosted in country. With an early market, more [doms] will be hosted outside the country's infrastructure."

MATTHEW ZOOK: Easy answer is no, I've not done that. I've looked at geolocation for name servers and so forth, but I've not done what – was it John? – John was suggesting. I think that'd actually be a really interesting study. I've just not done that.

EBERHARD LISSE: All right. Thank you very much. Dave Piscitello is the next one. He will speak to us about The Tool Formerly Known As Domain Abuse Reporting Tool.

DAVE PISCITELLO: Thank you very much for having me. I haven't been to an ICANN meeting since November of 2015, so it's nice to see a lot of you. I admire your tenacity.

Let's see here. The Tool Formerly Known As DART. Moments before I was getting on a plane to come to South Africa, the ICANN organization received a cease-and-desist letter from attorneys who represented a company that apparently does some sort of domain protection. They said, "DART is ours." I said, "Well, phoo!" We thought of a bunch of names. I personally had wanted the Domain Ecosystem Reporting Project, which would have been DERP, but no one else agreed. So we settled The Domain Abuse Activity Reporting Project, or DAAR.

A lot has been said about this project. I'm fascinated at how quickly disinformation and confusion propagates. It's faster than any routing protocol I've ever seen.

The project is intended to be a platform for reporting on statistics and behavior and patterns of abuse across TLD registries and registrars. The history of this project is almost as long as my tenure at ICANN. I have sat through so many meetings where people have made assertions. When you asked them, "Well, where did you get that data?" there was silence or a time-out.

Going through some of the literature and looking at a lot of the academic papers over the past five to seven years, we wanted to try to distinguish what we're doing in several ways. Most of these studies that we had seen did some samplings or focused on a particular kind of gTLD or ccTLD.

We decided that we wanted to do a little bit better than that. Most of the studies, especially the academic studies, did not incorporate commercial feeds because they didn't have budgets for actually subscribing to feeds over a very, very long period of time. We wanted to also fold into our studies not simply the same spam lists or other block lists that people relied on but tried to acquire as many what we'll call high confidence lists as possible.

We were very surprised when we began the consumer confidence and trust activity in ICANN to discover that a great number of reputation providers don't have a history of their own lists. So we decided that one of the benefits that we could bring to the community would be a permanent or persistent store of the data that we collect so that we could do some historical analyses on the ecosystem.

Again, a lot of the studies that we looked at focused on one kind of security threat. Our motivation was to provide an answer in response to the Government Advisory Committee. They originally had wanted to look at phishing, botnets, and malware. Then, in Hyderabad, they mentioned in a communiqué to the Board that those were just examples of the kind of abuse that we're interested in. And spam is one as well.

We also, on my team, truly believe that spam is critical because it is the delivery mechanism for so many other malicious activities, especially phishing. The majority of spam today is not spent by individual malicious actors but is transmitted through botnets. So we thought that was in fact an important threat to measure.

Another goal that we have is that we want to be able to present the data to the community without making judgment, without ranking. We want to have our methodology transparent. We

want the information that we use to be accessible so that, like all good studies, someone should be able to look at our methodology, get the same data that we have, use these, and reproduce what we've done with similar, if not the same results.

There is another initiative within David Conrad's Chief Technology Office called the Open Data Initiative. This is a project that ICANN is attempting to facilitate access to data that our organization or the community creates and curates and is of public interest or public benefit. The DAAR Project already uses entirely public open or commercial sources. These are sources that anyone can get. We don't use things that are internal and proprietary. We use zone data. We use WHOIS data. We use open source reputation data. We also use commercial reputation data or feeds. These require a license or a subscription. I'll talk more about that as we move forward.

As there are some limitations and constraints on the way that we actually use some of the commercial feeds because of the contractual obligation that we have with the parties that provide then, we feel that, in any case where there aren't any limitations, the DAAR Project data or reports generated from that data will be published periodically and included in the Open Data Initiative. We'll be talking later about trying to understand how to shape those reports for the community.

Why are we doing this? I think – and our CEO and my CTO agree – that the community will benefit from having data to support the Policy Development Process. Informed policy needs information. Understanding how policies are currently effective or whether there are unintended consequences can best be understood by presenting data, not simply anecdotes.

We hope that what we are presenting will be a database from which people will be able to derive value, identify studies, identify outcomes and findings, and understand whether or not these will produce some incentive to revisit policy or to think of new policy.

The way that we see the DAAR Project data being used is to identify a very broad threat landscape, not just a single threat, that's reported at a TLD or a registrar level for all the TLDs for which we can obtain data. I mentioned that we want to be able to do something more than a day-in-the-life or a single snapshot of abuse. We'd like to be able to track or see behaviors at registrar/registry or entire namespace levels to understand things like flocking behavior and migration, and perhaps in the future understand the relationship of pricing or of hosting with various threats and various operators.

Primarily, one of my goals is to help operators understand how to manage their reputations. These data are not generated by

ICANN. This is the data that the rest of the world uses to determine whether or not they're going to allow or block access of their customers or their users to a domain name or to an entire TLD. So this is not out judgement or our perspective. This is the perspective of the entire reputation data industry.

One of the best ways that I think we can do this is by collecting the data and providing some access or some reporting that can be used in policy development.

I mentioned that we use zone data. We collect the zone data from currently 1,241 TLDs. At the moment, we have just shy of 195 million domains in our database. We use the publicly available methods to collect data. We use the Centralized Zone Data Service for all the new TLDs that are obliged to sign up for that. We also have access to the legacy gTLDs in the same manner that we've been using for many, many years.

Previously, when I first presented this in Madrid almost a month ago at the ICANN DNS Symposium, six country code TLD operators came up and asked if they could participate. And I was very happy to see that they immediately grasped the value of being able to have someone else do this massive collection. In some respects pay for some things that maybe they can't get their own organization to pay for and be able to apply or use the data to assist in their own operation.

I hope that if any of you after this are interested, you'll come up to me and we'll talk about getting some sort of process in place and that's something we can arrange after the close of the day.

We use WHOIS. Saying you use WHOIS these days is like saying you drink Agent Orange. We currently only need sponsoring registrar, so irrespective of many of the outcomes of the GDPR and the RDS conversations. So, if people want to understand the registrar level or the portfolio of registrars from their zones, we just need the sponsoring registrar. We don't need point of contact data. We used published registration data from the WHOIS system.

One of the things that distinguishes what we do from maybe Zook and some other uses of zone data is that we only use names that resolve. Our philosophy is that if the name is not resolving, it's not a security threat because no one can visit that particular website or no one can emit spam from that particular origin domain. No one can direct NSMS or a Facebook comment to a malicious site. So, our counts are always going to be smaller than the actual counts that are published by the registry operators.

One of the things that I'm most happy about, and it took us a great amount of time to settle on is the threat data sets that we use. We collect the same abuse data that is reported to industry

and Internet users. And these are the data that, for example, Spamhaus or [Cerbil], Malwarebytes, APWGs, eCrime Exchange. These are the data that almost all security systems and anti-spam software and gateways use to protect billions of users daily. And so, that's the optic that the rest of the world has of the name space. This is how people see it. And so we want to use the same perspective or the same lens that the public uses, not something that we derived in-house.

We used lists that we felt were curated, had a strong history of accuracy. We wanted to make certain that by the combination of the lists that we had, we had a very global coverage and low false positive rates. We also built the platform so that should we come to a time when we were unhappy with the list, we could drop it and we could add another list if we became really excited about a new list. This is a practical consideration that recognizes that a lot of these lists actually began as research projects. They get funded for a while, the grant goes away, the quality of the lists deteriorates and unless somebody else is willing to pick it up, the list sort of disappears. So we tried to avoid those lists… Did I go too fast? No.

So to emphasize again, we are not an abuse listing service. We don't go out and build spam traps. We don't build networks. We don't sit and apply heuristics to e-mail and come up with our own lists. We use commercial or open data. When we apply

these lists, we're generating counts for unique security threat domains and then we also count spam, phishing, malware, total abuse domains and a cumulative abuse domain. The cumulative abuse domain is going to be a running window of 365 days of history.

So currently it starts on January 1, 2017, because the project began in October and by January we felt we had sufficient data to be able to start considering January 1 as day zero. The tool allows us to export data and auto generates some interesting histograms. We auto generate charts. We can export to Excel or CSV. And then we can take snapshots in day-in-a-life use. There's also a nice tool that allows us to search by an argument so we can enter a TLD or a registrar. And then we can choose a particular security threat and we can choose a date range and we will pull from the database the set of domains that actually match that criteria.

The output will tell us the domain name, what reputation list it was reported from, what it was classified as and currently it reports the creation date and if the domain has been deleted. After some conversations this morning, one of the things we are considering is trying to see if we can also include first observed as another one of the metrics.

So the current set of reputation data sets is here. One is actually missing. I apologize. But [Cerbil] is on this list. These are all lists that we think are very, very high quality and we also have at least two lists for each security threat. So several times people have asked us, "Why are you using so many data sets?" There have been other activities in ICANN where people are arguing about using one or another and not wanting to use many. We did some research and we talked to some people who were working on the ecosystem at Carnegie Mellon University in graduate work. Metcalf and Spring in 2012 actually began a series of articles where they tried to understand how much overlap there was among block lists. And it turns out that there's very, very little overlap.

We actually experimented in-house by running scripts against several zone files for weeks at a time testing domains against 86 block lists. And our results against those 86 block lists were very consistent with what Metcalf and Spring found. So this was, again, trying to make certain that what we were doing was going to give us very quality abuse data.

Do we get all the abuse? Just the simplest answer and so I don't spend a whole lot of your time talking and we have some time for your questions perhaps is, no. No one has a full picture or a composite picture of all the abuse. We, I think, get a lot and

certainly enough to be able to make some very good observations about abuse in the name space.

We did not want to do scoring. We don't want to do ranking. But we want to be able to measure or understand registry abuse in terms other than raw numbers. Because raw numbers for com are in the hundreds of thousands in raw numbers for some of the new TLDs are in the 10s or 15s.

So we have a percent of abuse metric that is calculated very simply as a fraction. We take the numbers of domains that had been listed in a reputation list in a TLD on a given day and then we divide that by the number of domains in the TLD Zone on that day and we multiply it by a hundred. So I'll show you at least one slide that reflects how this works. We do that for registries and registrars.

Currently our data for registries is very solid and we're very confident that we can begin doing some reporting and providing some information for the community. The registrar activity has been slowed by WHOIS collection. It's very hard for us to keep pace with the amount of WHOIS that we have to collect in order to be able to process or associate a domain with public information that identifies the sponsor and registrar.

Just to give you two little peeks at what our system does and what kind of information we have, this is a chart generated

based on the May 31st data of this year. And what you can see here is that while malware and botnets command the control domains still are predominantly registered within the legacy TLDs, spam and phishing have pretty much been distributed across all TLDs fairly uniformly.

This is an example of how we've scatter plotted all the TLDs in the DART System for which we had at least one event on May 31st. And so the dotted red line or purple line represents the mean abuse score of that fraction which was 0.6. The Y-axis is logarithmic . And so you can see that there's a great, great number of top level domains that are well below or below the median abuse. There's a number that cluster just above that and then there are some outliers.

This tells us a lot just pictorially about how this space looks. When somebody comes to me these days and says, “The new TLD space is just awful,” I can say, “Well, no, the new TLD space is not awful.” There may be some TLDs that are having difficulties, for example, there are 780 new TLDs that don't have any abuse at all out of the 1,200 that we have in our system.

So these are the kinds of things that we can do. We've got a lot of other ways to represent data that I didn't want to spend a whole lot of time presenting to you and I just wanted to talk a little bit

about where we're going next and then spend some time getting a feel for how the cc community is reacting to this effort.

We are in beta. It is an internal use administrative console only at the moment. So we have accounts with a provider that allow us to go and take a look at all the data that we're talking about, most of it is in tabular form. A lot of it is hyperlinked and allows us to pivot from name to registrar, from registrar to domain and get very detailed statistics on each.

As I mentioned, we have some expressions of interest from cc operators and if there are others here, you can contact me. I think many of you know me, if not, I have cards. And certainly, others in the community especially Eberhardt can connect you if you can't remember who I am.

What's most important for me, this week, is to express the following thoughts. We have data for you. And so we've put eight months of effort into creating what I think is a possibly very valuable set of data for the community. The question is: how do you want to use it? As staff, we don't want to make judgments. We want to make data available and have you tell us this is the kind of data we'd like you to share publicly or to share it with us as individual operators and in this fashion with these kinds of questions that we'd like to answer. And then we will start to consider how we actually generate those kinds of reports or

export data and then we'll take that into consideration as we go back to our vendors of our data and say, “This is what our community wants us to do is our license to correct one and, if not, we'll try it in the negotiated contract that allows us to do that sort of thing.”

So, I'm done talking. I think I have a few minutes. And if you have some questions, I'd be happy to answer them. If you are interested in getting a little peek at the command interface and see some of the data, I'm not crazy enough to put it up on a screen in an open forum where I'm not wearing Kevlar. So, I'll be happy to show people some examples of the graphs and such, but that would be just a demonstration. If you are a gTLD operator and you want me to share the data we currently have for you on a given day, I'm happy to do that right now. And if the ccs want to participate, just come talk to us. Thank you.

EBERHARD LISSE: Okay. Thank you very much. One thing open carry in South Africa is there but very, very severely restricted. Open carry —

DAVE PISCITELLO: Open carry. No, I didn't say I bought guns.

EBERHARD LISSE: No, but you said [inaudible].

DAVE PISCITELLO: Yeah, I was thinking dragon scale. That's really very slimming and very tight and when the bullet hits it explodes. It's great.

EBERHARD LISSE: And before we open the floor for the questions, I've got two things. All the links of all the presenters are on the agenda. So if you download the agenda, you can click on the name of the individual and you will be pointed directly to their e-mail address.

I have a question, if a ccTLD gives you zone access, preferably a zone transfer access, how can you guarantee that nobody else at ICANN gets access to that data?

DAVE PISCITELLO: So that's a good question.

EBERHARD LISSE: You see, some of us have steadfastly refused to give PTI or its predecessor access to our data and some of us like me still steadfastly refuse to do so. Not only because it's our intellectual property, but also because the way historically ICANN staff has

on some occasion did deal with that. What can you tell me about that?

DAVE PISCITELLO: So, we are working with a company called iThreat Cyber Group. Jeff Bedser and Greg Aaron who are on the Security and Stability Advisory Committee. Many of you probably know from affiliates or from previous experience. We have built a custom extension into their product, and the way that other ccTLD operators have approached this is they said that we will work with a Memorandum of Understanding or some contractual obligation with ICG. Those zones themselves will not be pulled into ICANN directly, they will go into our database that is actually hosted by and owned by ICG. So you would not be giving it to ICANN staff, you would be putting it into database that our ICANN staff can go and use in the same way that we're using for gTLDs.

EBERHARD LISS: As I don't see another question at the moment, I think that contradicts itself a little bit. How can I be sure that PTI is not going to get access to my zone data? Is that guaranteed?

DAVE PISCITELLO: To pull the entire zone?

EBERHARD LISS: Yes.

DAVE PISCITELLO: So we will not know what your file transfer or AXFR is. It's going to be something between you and iThreat Cyber Group. So your issue of not disclosing the zone to ICANN is one that they'll have to ensure.

EBERHARD LISS: Any other questions? Thank you very much. Oh, there is one.

UNIDENTIFIED MALE: It takes a while for us old guys to stand up, you know. So first day, I'm thrilled that this is actually coming to fruition and you've actually got stuff that you can publish. And while understanding that identifying specific TLDs would be a death wish, I'm wondering if it would be possible to sort of categorize the results by category, legacy TLDs, brand TLDs, public new TLDs, restricted new TLDs and stuff like that. Because my guess is we would find significant differences among those groups.

DAVE PISCITELLO: You're spot on. So let me answer a couple of your questions. First, thank you, this has been something I've been working on

for five years, the last two years intensely. So I'm delighted that we have it.

Second, we have the ability to – and I already have generated some charts for internal consideration – we can do an individual TLD. We can do all the TLDs that are operated by a common company. We can do legacy versus new TLD versus IDN gTLD versus ccTLD and we can do individuals. So the question is really going to be up to the community and obviously the contract parties are going to have a say about doing this.

But, one of the bullet items here that I overlooked was the order of reporting. It may be that some people will choose a model where the registries and registrars get an opportunity to take a look at and curate the data by going in looking at their portfolio before we publish. As far as I'm concerned, the outcome is the same because if somebody is going to dump 100,000 bad domains and is not going to start gaming the system, that would be great.

UNIDENTIFIED MALE: [Inaudible].

DAVE PISCITELLO: Yeah. I mean, the goal here is to get the abuse score down to as close to zero as possible for all TLDs. How we accomplish that is

something that everyone else can sort out. But we can give you the data all the way down to the lists for you to go and do that kind of curation. So I'm hoping that the proposition in front of the contract parties is attractive.

UNIDENTIFIED MALE: And since I'm standing here with nobody behind me, on your scatter plot there was some straight line artifacts. Any idea what those are?

DAVE PISCITELLO: I think it's the order in which they were listed in the comma-separated value file.

UNIDENTIFIED MALE: Okay. Yeah.

DAVE PISCITELLO: So one of the things I had on my task… By the way, it's logarithm by logarithm, so if that helps. When I took it off logarithm by logarithm and X-axis was not logarithm, everything shifted in one direction but it was still the same pattern. So I noticed that myself and I have to go take a look at [Y].

UNIDENTIFIED MALE: Okay.

DAVE PISCITELLO: Thank you.

EBERHARD LISSE: We have more time for questions. Okay. Thank you very much, Dave. Next one will be Linda Müller from Knipp. She will talk about mambo which is a commercial domain abuse reporting tool. But we have had, as I said before, on occasion allowed commercial presentations provided they were not too commercial and, in particular, if they offer services free of charge to deserving ccTLDs which I have managed to convince the management of Knipp to do. So go ahead, please.

LINDA MÜLLER: Thanks for having us here or me. Just real quick I wanted to say sorry to the DART Project which renamed itself. We don't consider ourselves as providing DART, that is actually ICANN's terms. But as you renamed yourself, I think you won't be any mad at us at all anymore.

Well, I'd like to talk to you about our mambo+ funds program today. We have an abuse monitoring tool that's correct, but we

do have a slightly different approach than DART has and we will look at this real quick. I think I can do this myself, right? Okay.

So, first of all, just a short introduction from Knipp. There are two people here over there, they're sitting. My colleague, Dr. Michael Bauland, he is a software architect from our company. I'm Linda Müller, I'm a project manager and Knipp [inaudible] has a Germany-based software company that's providing solutions to the domain industry.

So today I'd like to talk to you about our lessons learned and what we did with abuse monitoring and then, of course, go into detail about the mambo+ funds program.

So, the abuse monitoring, the first question you have to ask yourself is: what is domain name abuse and what is actually the domain name abuse you really want to handle as a registry? The topic got really hot in the last two or three years within ICANN and there were several things on the table.

We are looking at spam, phishing, malware and botnets as is DART also. I don't know if everybody's familiar with that, but just a short introduction on that. Spam is unsolicited and [bulk] e-mail. So it's unwanted e-mail that is equally applicable to many recipients. Then malware is malicious software such as viruses or worms. Botnets is a group of computers that is controlled by one single source and is infected by malware. And this one single

source is the command and control server and this could be reached by a domain name.

Then there is phishing which is an attempt to steal someone's personal identity or data such as PINs or TANs by using fake websites. And just so you know, at ICANN there are some discussions going on about pharming. In our view that is nothing that a registry can have a look at, because pharming itself is an attempt to steal somebody's personal identity or data by redirecting somebody and using modified DNS entries. So you have to have a look at the resolution service or the client's implementation. And so for us we look at phishing or malware, but not at pharming itself because it's actually the second step.

So who can do anything about domain abuse? It's actually the whole value chain that is able to do anything about that. So the registries can do something. The registrars can do something. The hosters and all the registrants. As we are here talking to registries, it's most important to you what you can do. The first thing is, of course, very important, you have to monitor your zone. And if you find out anything, you might want to inform the one that is capable of doing something. So you might inform the registrar's host or its registrants or the third-parties such as law enforcement agencies.

And, of course, if you are able to do anything, you could either put the domain on server hold, delete it or suspend the registrar's account. Naturally, that depends on your policy and across ccTLDs that is very different. I have some European ccTLDs in mind, they have very different approaches to that.

So, as we are very familiar with this topic, I don't know if you are, as I said this topic is very hot right now at ICANN and there's a lot going on. First and foremost, there's Spec 11 3b ngTLDs. This specification is their Registry Agreement and this is where probably most of this started from.

So, there is also an advisory that was just published on the 8th of June that is advising the registries what they could do in detail.

Also another working group, the Security Framework. This is registry-related working group where the registries voluntarily define what could be done with abuse and define measurements of what they want to do.

Then the ATHI, it's the Identifier Technology Health Indicator. That is the group that is firstly looking into what actually is health of a domain portfolio. And then in the second step tries to identify or tries to define an indicator about zones and their health. And, of course, there's DART and now it's renamed.

So during our processes of looking into abuse handling, we found that there are some different information and tools needed concerning the information part, of course, you need abuse information. And as Dave already told you, there are various sources. They are free. They are with costs. There are pull or push sources. And, of course, they are required for doing the abuse monitoring and management.

Secondly, and very important also in our view is the zone file because you need to know about the domains you want to monitor, because you could do the abuse monitoring without this information. But, if you don't know about all the domains you couldn't use source where you have to pull data from. So it's optional, but in our view very important.

Then there's other information and, for example, there's registrar and registrant information that could be of interest. This is also optional, but in our view also very important as you might want to have the contact of a registrar directly whenever abuse is detected.

Well, having a look at the tools you need we found, of course, you need statistics, you need the overview, but very, very important is the details. Because if you don't have the details on an abuse case, you can maybe measure what is health in your zone, but you couldn't do anything about that.

Then you might want to do some action tracking because otherwise you don't know what has been done and what was successful. You need the history of your information to learn from the history obviously. And you need an itemized process as far as possible, for example, you need alerts via e-mail or mobile phone.

For your overview, it's important to have immediate and periodic reports. So while our review at all these things we found some lessons and the first thing is you need most recent information which is pretty easy to say, but actually that is a hard thing to do because one problem, for example, was the sources is that it takes a long time for some sources to actually add new data. So you might be informed of an abuse case months or weeks afterwards. And whenever you did the abuse handling it takes some time until the source deletes that information. So you really have to know, in detail, what sources you are looking at, what they are doing. Nevertheless, those sources, if they are using that data or if they're presenting the data for a long time while it's not accurate, your reputation is still, well, negative by those sources.

Then, of course, you need user-friendly accessibility because this is key for successful mitigation. Otherwise, you would probably do that once and then let it go because those different sources, if you really want to do the manual processing of those, that's

pretty hard to do. They are in different formats. They are using JSON, XML, CSV, and this is kind of an endless story if you do it manually day by day. You couldn't filter, you couldn't search, so you need the combination of the sources and you need it in a user-friendly interface that solves your problems with filtering and searching.

If you don't want to have a look at this data every day, you might want to automize as far as possible, so you want to have automated, immediate alerts whenever something is detected so you get a signal whenever you should do something. And of course, you would need the reporting so you wouldn't have to do your summary by yourself.

Then, many benefits come with data combination, so if you combine the data of abuse to other sources or type of data, that is very valuable. For example, if you have the contact information, you can contact the registrar directly. If you have historic data, you might have an indicator from malicious domains even though they are not registered yet.

And the last point, which is for our team, most important is details are important. If you have just a list of abuse domains, that won't help. That might actually help for researchers to have an indicator of health of zones, but if you really want to do the

mitigating, you need details. You need, for example, the exact URL.

I don't know if that is possible to see here. It's probably not. This is just a short look at our interface and we have collected those different domains. It's just a randomized portfolio. And we show you, okay, there's abuse, but that's not the important point.

The important point is that we show you there directly. We show you the URL and you can use that, for example, for detecting what abuse is really there, although, in this case, we also already tell you the Locky distribution site. And then, if you want to do some evidence collection, you need that URL because otherwise, you won't be able to do anything about that.

So now, about the mambo+ funds program. I'll just go on and then we'll do the Q&A afterwards.

So our goal is to make the advantages of an abuse monitoring tool available to registries currently don't have the financial resource to pay for such a service. We did that once with another product of ours and that was, actually, a pretty good experience.

We provide the monitoring for domain portfolios for registrars and registrants, but especially for zones and registries, and in this case, for ccTLDs. We identify the abuse cases for spam, malware, phishing, botnets and additional data.

Concerning the additional data, at the starting point, we were an abuse tool but our main principle with this product is that it's extensible, so we extend that on an ongoing while we are developing it, and currently, you can also look at the DNS information of domains and also ranking information. So there is also a point where you might be interested into that.

You can view historical data, so we keep track of what has been done and what we saw. Our data is stored in Germany on our servers and they're under the European and German Data Protection laws. We will have the immediate alerts for you and also the automated reports.

So, just so you know or you know who's providing that solution to you and you can trust us, we are Knipp Medien und Kommunikation GmbH. We are specialized in software. We've been providing services to the domain industry for over 20 years now. We are providing registry backend that's called Tango, and the domain server structure, ironDNS. And everything is under very high standards of data protection.

So we are ISO 27001 certified. Our own data centers are in Germany and everything is under the German and European Data Protection laws.

So who can apply? This program is for ccTLD registries that are non-profit organizations, have less than 5,000 domains and are in non-OECD countries.

So how can you reach out to us? Well, it's an unbureaucratic process. Just get in touch while we are here. Drop us your business card or contact us by using this e-mail address or go to mambo.plus, which is the website.

EBERHARD LISSE: Okay. Thank you very much. For disclosure purposes, we [inaudible] rents server space at Knipp and we make use of their free ironDNS anycast secondary service and we are very happy with the service, especially the one that we don't pay for, but the other one too.

LINDA MÜLLER: Yeah, well, that's… Sorry.

EBERHARD LISSE: Any questions? Dave.

DAVE PISCITELLO: Hi, Dave PiscItello from ICANN. This is the second time I've seen you talk about this and sometime, I'd love to get a demonstration and bang heads. I'm really glad that you're doing

this. I think the more people that do this, the more we press the information out into the public, the more we put pressure on people who are misusing the space, and so this is great.

You did have a slide that I was curious about. You mentioned registrar and registrant information and then in parentheses, you said, "WHOIS and escrow". Can you explain what you mean by escrow?

LINDA MÜLLER: I'm not a technical person, but I can explain what I think of escrow. Well, we are working together with ngTLDs that are providing their escrow to us.

DAVE PISCITELLO: Okay, so this is not the ICANN escrow process.

LINDA MÜLLER: No.

DAVE PISCITELLO: Okay, great.

LINDA MÜLLER: If the registry wants to share that, that is, of course, welcome because then we have the current data available.

DAVE PISCITELLO: No, that's fine. I was just curious because we can't get to the escrow. I was thinking, "But why can you?" So.

EBERHARD LISSE: Any more questions? Thank you very much. Now we need to discuss. Do we carry on or do we take a break because we have two more presentations? I personally am of the opinion, let's get it over and done with.

The next one would be Francisco Arias. He e-mailed me he was in but I haven't seen him because he's fine, in the back. I see him now.

He's going to speak about an ICANN monitoring API system and somehow, when we were discussing the agenda, an e-mail crossed our paths and the program committee and several members on the program committee expressed strong interest on this topic, which is why we invited him to come and give a presentation.

FRANCISCO ARIAS: Thank you. Hello, everyone. I'm Francisco Arias. I work for ICANN on the GD area technical services, so I, my team focuses on technical issues related to gTLDs primarily.

So this is a presentation of a system we have in ICANN. It's a monitoring system that was built, a [sort of] of some provisions that gTLDs have in their contracts in relation to performance requirements.

This is the agenda for the presentation. I'm going to do a quick description of what the system do, some of the statistics from there and an API that we have in pilot mode right now that may be of interest to some of you. And finally, I have an ad, you may say, from ICANN.

So first, let's start with a brief description of the system. So like I said before, this is a system that we built as a result of the Service Level Agreement requirements that the 2012 gTLDs have in their agreement and some legacy TLDs have also incorporated as they renew their agreements with ICANN.

So this SLA that finds requirements for response time of availability of what we call the critical functions which are DNS, DNSSEC, WHOIS, and EPP. So we built this system based on the Zabbix monitoring platform with some custom plug-ins and code. This is open source and it's available. You can see the link there to the code if you are interested. It's currently based on Version 2.0 Zabbix monitoring platform and we are currently in the project to migrate to the Zabbix 3.0 branch.

In the operation of the system, we use probe nodes that are distributed around the Internet. We try to cover all the regions, but we focus on the regions where there are more Internet users. We try to have the point of view of the Internet user that is trying to resolve a domain name.

So we have, currently, approximately 40 probe nodes. I say "approximately" because the number of probe nodes that are online at any given time varies depending on issues with a certain probe node that has to go offline.

The one important thing to mention about the system is it was designed to avoid false positives, so by default, we consider a service up unless certain conditions are happen, which are, for example, that 51% of the probe nodes that are online see the service as down. And also, we have a minimum requirement of probes that are online in order to consider a service down. In the case of DNS, it's 20. In the case of EPP and WHOIS, it's 10.

We also consolidated our points regarding the availability of service in a rolling week basis. So this is the last seven days, but not a calendar week so it's rolling over time.

So the most important point, here is a graphic representation where you will see the different probe nodes and we have, of course, a central node that collects information and from there, we derive [letters]. In the case of the gTLDs where we, since they

have a contractual obligation to certain SLA, then if they are not compliant with that SLA, we get in communication with them and try to help them through the issues so they can resolve whatever is happening.

In the case of the DNS, this is the query that we do. It's every minute. Its probe node sends a type A query for the QNAME that is listed there. The TLD, of course, varies depending on the TLD that the name server is serving. And we do this to each IP-address of the name servers for a given TLD.

And if DNSSEC is offered, in the case of the 2012 gTLDs, every one of them is required to offer DNSSEC. In the case of the ccTLDs, not everyone offers DNSSEC, but those that offer DNSSEC, we are also checking that name NSEC/NSEC3 records and their signatures are correctly formed and are valid.

We also check the chain of trust that is, that you can go all the way to the root zone KSK. And these are, of course, examples of the typical issues we see. I don't think we need to get into too much of this here.

And some statistics. So this system was designed for the gTLDs, but as we put it in production, we thought, why not we also look at the DNS service of the ccTLDs so that we have some data, and perhaps in the future, we can offer something to the ccTLDs

trying to help improve the [scale and] stability of the DNS since that is in the ICANN's Bylaws.

So this is some data that we have seen on the ccTLDs only. We made a presentation last month in May in Madrid about gTLDs. But in the case of the ccTLDs – by the way, this is data from October 2014 to May this year – so we have seen… another thing that I should mention, this is related to how we are measuring gTLDs in their contract failure that extends for more than four hours is considered a very serious issue and there is something that is called [inaudible] thresholds in the case of the gTLDs and if a TLD passes that [inaudible] threshold, then they could even lose their TLD. They can be taken over by ICANN.

So that's why we are having, we are showing this in regards to failures that last more than four hours, because that's how the system was designed to be. I think we could take data on a different basis, however, this was what we were able to provide on a quick basis for this event.

So we have seen in 273 DNS failures that have reached four hours or more on a rolling week period in this period of time. And from the 295 ccTLDs that are active in the root zone as of today, 60 of them have reached a four hour or more downtime in a rolling week window. And you can see that, also, the number of ccTLDs we have seen in the system that have had at least one

DNS service down event, that is 60% of the ccTLDs and interestingly, there is a slight difference in the ccTLDs that we have seen with at least one DNS down. Even if you look at the ccTLDs from IDNs [inaudible], 70% while the ASCII ccTLDs is 58% that we have seen at least one failure.

As interestingly, there are five ccTLDs that are down most of the time, if not, all the time. And here is a graph of the downtime we have seen over time. The yellow or orange line that is there is the average, so we have an average of a little bit more than eight in downtime incidents per month, incidents of more than four hours or more on the ccTLD space. And you can see there how it varies, sometimes widely from month to month.

So for the system, after we built the system, there was interest from, again, gTLDs to gain access to the information we were seeing in this monitoring system. So with that in mind, we built this API that we called MoSAPI, the Monitoring System API. So this API provides a REST API to retrieve the data collected by SLA Monitoring System in close to real time. We have it in pilot at the moment. We still don't have a launch date for the tool to be in production. Probably early next year, but we still don't have a date with certainty here.

It's important to mention that in this tool, registries can only see their own performance data. They cannot see the performance data of someone else, only their own.

Here are the credentials and what you will need in order to gain access. After the presentation in Madrid, there was some interest from some ccTLDs to gain access to the system. We went about this internally and we think we are ready to offer this access to any ccTLD that is interested in doing it.

If you are interested, you can send an e-mail to this e-mail address – it's our Global Support Center – and they will route this request internally, even to legal. It will go to my team. Just bear in mind that this is a pilot system so we are not making any guarantees in terms of uptime. However, I think it has been very good, the time we have been [inaudible] this.

You should also know that we are going to be processing these requests manually since we don't have, like in the case of the gTLDs in which we have already established credentials to indicate the gTLD, in the case of the ccTLDs, we are thinking of doing this line on the ccTLD contacts in IANA, so probably sending an e-mail with a unique, random code to authenticate that we are, in fact, talking with contacts from the ccTLD. And I think that's it in regards to the system.

What comes next is a [shameful] ad from ICANN and hopefully, [Al] will not get mad at me for this. So my colleague, Dave, talked before about access to the zone files of ccTLDs and ICANN is interested in getting access to zone files of ccTLDs for various purposes listed there.

We're interested in getting the statistics about DNSSEC penetration, IDN penetration, active names in all of the TLDs and, of course, as an input to the DAAR System.

As Dave mentioned, if you have sensibilities in terms of providing zone file access to ICANN, but you are still interested in the DAAR System, you can do it through the way that Dave mentioned it. But if you don't have those sensibilities and you are willing to share that zone file with ICANN, that would be very useful for us and you can let us know that also through the Global Support e-mail address that is there.

I think that's it for me. Yes, that's it. Thank you.

EBERHARD LISSE: Thank you very much. I think this, I would be most interested in seeing what the five domain names are that are more or less permanently down so that we can assist them.

Secondly, I would then privately be very keen into knowing whether you found something for us, .NA, for example. But we'll do this offline.

And secondly, I think this is a very cool idea. And I said, we are not interested or we have sensitivities like [inaudible] access to our zone, but if part of the system, like SLAM again goes without, that is a good idea. If you need some information without getting access to the [total form] such as penetration and so on, just send e-mails.

And then we have Robert from PCH.

ROBERT MARTIN-LEGÈNE: Hi, I'm Robert Martin-Legène from PCH. What Eberhard had said and a little bit more. No, I think it would be very interesting for others than just the registries themselves to see these data because we are all users and consumers of those data.

I think it would be very interesting to know, in my case, whom I can help because I talk to a lot of people. There's a lot of people that are very good to help colleagues, but it is very difficult if we don't know there's a problem and ICANN is probably often a little bit afraid to reach out and fix, let's say, five TLDs. Maybe there could be better ways of seeing all this.

Also, as a DNS provider, we provide DNS services for a few TLDs, or quite a lot, and for me, it would be impossible to talk to my customers to get each registry to give me credentials to log in to see how I provide DNS services for them. Of course, I should know that already, but I would like to know how ICANN thinks I run things because I think what I have seen a few times would be false reports, but maybe I could be persuaded otherwise. So for that reason, it would be interesting for me to see more open access to this data.

EBERHARD LISSE: As far as I'm concerned, you can get our permission to see our data. Maybe you should approach all the people that you serve and ask them whether they're willing to assist you in this regard, but you have my permission you have to see for [us].

FRANCISCO ARIAS: I think as we see things from ICANN's side, we don't see all the TLDs necessarily willing to, for everyone to see their data. So we are sensitive to that and in that sense. And there is some consideration internally to providing some data publicly, most likely, without identifying the TLDs that are, for which we are seeing these issues and this is in regards to the gTLDs, I say that. We are, at the moment, not thinking on publishing anything related to ccTLDs, only to gTLDs.

In regards to giving you access, you as a DNS provider, I think just what you were saying, it occurs to me that maybe something to consider and this is not yet something we were planning to do, but we should probably talk. Maybe if there is a set of name servers that somehow we can authenticate that you control, then maybe we can give you access, have a few [inaudible] name server instead of the TLD, because also, maybe you are not the only DNS provider of certain TLDs that are –

ROBERT MARTIN-LEGÈNE: Definitely, yes.

FRANCISCO ARIAS: So maybe we can think of giving you a [inaudible].

ROBERT MARTIN-LEGÈNE: That would be very interesting. Yes, thank you.

EBERHARD LISSE: Okay, then we have a remote question.

CATHY: Yes, this is Cathy with ICANN staff and this question is from online from M: "Hi, Francisco. Did you compare your results with the data from…" – and this is an address –

"atlas.right.net/dnsmon which is using more probes as far as I know?"

FRANCISCO ARIAS: The short answer is no, we have not compared our data with [them].

EBERHARD LISSE: All right. Thank you very much. And now, we come to the last bit which is a paper, a small presentation that Alejandra and I came up. We'll do it from the chair and you can go with the next slide.

Actually, I [can't] do this. What happened to us was we received an e-mail in early June by some youngling hacker type of person who told us in a ten-page e-mail that we have something which you could describe in three lines. We had three of name servers of a particular domain name we are using [inaudible]. They were all on one free service provider – I'm not going to mention that it was Rapid Switch – and that would create a possible man in the middle attack in a convoluted way because we use these e-mails from this domain names to talk to IANA. If you could get access to this domain name totally, you could basically re-delegate .NA by communicating to IANA.

The threat is, in fact, fallible but it did not affect us and could not affect us because IANA validates, so this is a man in the middle

attack and it would not, IANA's [inaudible] validate would find out that if the mail comes from one of the three compromised ones, it would get the non-resolve and then try again until it gets the one that it can deal with. So it is a credible threat, but it didn't affect us.

The actual thing was that we used these two e-mail addresses as the admin and technical contacts for .NA. They are published in the WHOIS, so there is nothing secret about it.

When you communicate with PTI or IANA, you can easily go through the root management system [RZM], where you have a username and a password which is separate from e-mail, so that's a separate way. Or you can send an e-mail template.

They are looking at that, whether this is still doable, but it's a large number of ccTLDs, five of which are down most of the time, but a large number of ccTLDs don't have accounts on their system and they use the template when they want to make a change.

When you go to RZM, or you send them an e-mail, they send you, the system sends you an e-mail to each contact and if you change the contact, it sends you an e-mail to the changed contact as well. All of them have to agree to it, and then it goes back. If all are in agreement, it's a seamless process and it shouldn't take more than six hours, which on the second

attempt we did, we will come to at the end, it took exactly six to eight hours which is perfectly fine.

So we had registered with Rapid Switch, as a name, as a secondary for na-nic.com.NA and they offer three name servers, so when you add a server with them, you get automatically three name servers within their RapidSwitch.com.

For some other reason, our account went away and we did not, and I have records. I could find out that Matthew Zook, who presented, talked to us in 2004, in 2005, and in 2008, and each time, I told him to go away. Yeah? So we have not got this e-mail. So I don't know what happened, but the point was we got, our account got deleted and these three name servers, because they are connected to the account there, went away.

So if somebody else would have registered with Rapid Switch, this domain name, he would get control of the three out of four domain name servers. Yeah?

Then you can re-list this and then you give the three names that, on Rapid Switch, I give it a different Master, the malicious one, which then would propagate to the three name servers of Rapid Switch and then you have three of four MX records under control.

Then you can attempt a modification. If staff knows you very well and says, “That doesn’t look right,” they will contact you over another mean, but if it’s a routine change or something that looks familiar, if they were to re-delegate .NA, they would probably, if an attempt at re-delegation of .NA was happening, given my history with them, they would probably start alarm birds going off all over the place. But in theory, it’s possible.

Since we are DNSSEC-signed, it would have not worked because I checked with IANA. Their system validates DNSSEC. It is a credible threat for cc or other TLDs who have a similar setup and who are not DNSSEC-signed and to re-delegate a whole TLD to revoke and delegate it to somebody else is probably not going to happen even if it was technically possible because people at IANA would start waking up and that’s not what’s happening. We are not aware that this is going to happen. Nobody talked to us about it. This is a step in a process and it’s not the first step.

When we got the first e-mail to ask whether we are the right individuals, I checked and immediately saw that my name servers are lame and immediately checked the service in minutes before I even got the long e-mail from him, this was fixed. We added two name servers. Fortunately, we, among others, work with PCH and Mark Elkins are sitting next together, and both of them used TSIG to get our zones. So –

UNIDENTIFIED MALE: [inaudible]

EBERHARD LISSE: Of course. So it worked. It took about, it took minutes to add name servers. And I mentioned the TSIG thing is because I worked with them to establish it was relatively easy to generate a TSIG for my own server that is set up. That sits in a different infrastructure in Ireland.

We don't change it on the portal, and the portal regenerates every hour or every two hours, and then it took a little bit for the caches to expire, but it was fixed very quickly.

We then ran every important zone that we used to deal with this through Zone Master and fixed all warnings that we could and all errors. We had a reverse lookups that didn't work, so we fixed them all.

We then contacted IANA and now comes [inaudible] you will probably want to watch my performance tomorrow at the PTI meets ccNSO meeting. We wanted to move one of our contacts out of .com.NA into a different top level. We heard this early on. If you do that, if you get compromised, the amount of work to compromise increases. So we not only did that, we also used different hardware and as we said, we put it on a server in

Ireland so the more you have to compromise, the more effort it takes, the more, the quicker the chances are that you figure out somebody is playing with this.

We first went to an existing name and we then registered na-nic.com without .NA and used that so that it makes sense because the names are similar. As we have all accounts, IANA then said we are going, we held up the request until the time we told them who is behind the whole account and I said I'm not having any of these conditions. This is not a published policy. I don't want it. It's very clear who is behind all the [whole] accounts. We have a long history with IANA. And in the end, they agreed with us that if we are behind, if I am behind the whole account, they work on this assumption, they didn't want to publish it. They wanted to be able to be contacted apparently which is not wrong, but to make it a condition of making such a change is something that we must nip in the bud. We don't want them to force us to make changes only when we have certain conditions such as entering into agreements, giving zone transfer, which has all happened in the past.

So now I want a show of hands. Who, in this audience, is a TLD manager, cc or otherwise? And keep the hands up until I tell you take it down. Not many. Okay, who knows or when this happened to me, knew what an MNAME was, of the ccTLD

managers? I didn't. Who knows the requirement for an MNAME? I didn't and I now know it.

Now, why, this is the RFCs in which it says, why is this, who is this, who has recently checked whether your MNAME is correct? Yeah, we, too, were involved. We obviously checked, but this is important.

Okay, we'll just quickly review an SOA record. You see the MNAME in capital letters. The first, this name is basically one of the name servers where you have your data on. It actually is not really one of them. It's the primary. Yeah? It's only an example. I'm not sure, the timings are debatable. It's not really relevant. This is a more complete record.

What I didn't know is that the MNAME that I have on the top is not necessary behind an NS.2, NS.3, and then MNAME in the middle. These two record entries are not necessary. The MNAME is already fulfilled by staying in the SOA record. It doesn't have to have an S entry.

In your parent, you have three entries for name servers. Sorry. Do I have a laser pointer here? No.

The third name server entry in an NS MNAME is not necessary. It's good to have it, but it's not necessary. I did not know that.

The queries, they go… if primary is looked at, it goes at the MNAME that is listed in the SOA record.

This is where the problem comes from, that .GT had. If the MNAME does not have a Glue record and it's not listed as a Glue record, it can be registered by somebody else and it can get you into serious trouble. It can get you a man in the middle attack. The MNAME can get a false IP address and you're not really aware of it because you're basically thinking the three NS records below the SOA records are the only ones you need. If you have DNSSEC, it will work. [inaudible]

Now, Alejandra is going to take, talk a little bit about their little problem which is related to this.

ALEJANDRA REYNOSO: Thank you very much. This is Alejandra Reynoso from the GT.

So what happened to us was a series of unfortunate events due to Windows Dynamic Update errors.

So on June of 2016, we migrated our GT services to other facilities and then we changed all configurations, including the MNAME. So we also received an e-mail from the same person saying that the MNAME was not resolving and it was not registered and these will eventually, or potentially, could lead to

an active directory vulnerability due to the dynamic update of Windows networks.

So the GT was not compromised at any moment and it was true that what the e-mail said. Of course, I went and checked.

I'll change the slide here. Oh, you cannot see the last thing, but oh, there.

So Windows Active Directory and Dynamic Updates used to manage a number of services in Windows networks. When there is a change at the network, say a resource is added or removed or there is a renewal of the IP address, then the Dynamic Update occurs.

So the thing is when the Dynamic Update is sent and it fails, then it asks for a primary server to then request to this primary server to do the update and it cannot do the update. It keeps escalating over the hierarchy of the DNS until it finds either the primary that will accept the update or fails.

It's supposed to remain internal. This is an internal network that is doing its updates. And the problem was the following.

So, yes, thank you. Thank you very much.

Subtle misconfiguration can cause leaks. That means name collisions, which are private domain names trying to be resolved

outside in the public space. Then DNS queries reach external name servers, or that could be, in our case, the TLD. And then, if the MNAME is registrable, then someone that has that domain name can see the traffic that it's been sent of updates of those internal requests.

And please, thank you. Next one.

So when we got the e-mail, we immediately resolved it. We just changed the MNAME and within a registered domain that is ours. As I told you, when we migrated the services, since we do not do Dynamic Updates, we just placed any name without thinking that this could happen. So it was an unintended consequence.

Notice the MNAME that we have, it does not resolve because we don't want to receive this traffic. We don't want to deal with it and we don't want to handle it, but the MNAME, it's now in a controlled space.

Next thing, please.

EBERHARD LISSE: I personally think that this is not in compliance with the RFC and I must say we have seen, we have presented a few years ago in Singapore about attacks on our name servers and when we developed some [inaudible]-based system, we saw a lot of queries and now know this is MNAME queries. This is Dynamic

Update queries. Yeah, because we have a name server. We have, our primary does this for a number of domain names and we're getting queries for domain names that don't belong there, and this is clearly from this.

The point is that it's not, when you don't [inaudible] RFCs over and over again, you think your three NS entries are the name servers that you are looking at. But, in fact, no matter what, even if the MNAME in the SOA record has aged and has not been fixed, it's not keeping up with the time, they still go to the old one and that is what is not an obvious problem. I didn't know about it. I didn't pay much attention to it. I thought, "Oh, I changed my name servers, finish [inaudible], and that's it." This is a non-obvious error and if you don't pay attention, it can really bite you. In other words, it is not helpful, not really hurtful to read these [inaudible] documents again and again and again.

What I would have, if I had been, if that had happened to me, I would not have put it in a black hole, I would have registered it to a domain name, put it somewhere and just dumped [inaudible]. It's not that much effort anyway.

What we have done and what I think, if you have such a problem, is diversify the infrastructure. We put up a different name server in a physically different location. We registered domain names in a different gTLD. We changed our e-mail addresses to two

different TLDs, so it becomes much more effort. If you want to break this, it is much more effort and you find that in this kind of DNS fraud business, it's all least effort business, it's automatic high volume. We also see that they work during working days only from Monday to Friday and roughly 8:00 to 5:00. Over Saturday, we get very little of this traffic.

But the more simple solution you can put into, the more difficult it becomes to come [inaudible].

Unfortunately, there is no really automated and not really automated system. Zone Master is there, but Zone Master gives you a graphical thing and you can't do it for all your structures, unless you can program in [Perl] and can take the, get the stuff on the GitHub and write your own thing. But I have played with it. I couldn't manage it.

The problem is if you've got 5,000 domain names, you can't go and click on the website. You need to have a tool and then you need to have a tool that only functions if there is a problem, preferably in a graphical manner so that you basically graphs or something to at least let you see, "Oh, there is something that we need to look at."

Otherwise, these things happen. You get an e-mail and you get so many e-mails, you get 6,000 per week from CCWG-

Accountability and from your other work, so it just go in a folder and you don't really act on it.

The point is not so much what happened. It's just something that you, I was not aware that this could happen. I was not paying attention to every single detail. So it's a good thing that we, once in a while – [.KE] exposed when they had a total failure of the DNSSEC – it's also good to, once in a while, show that you get caught with things, preferably if nothing serious happens, so that we all learn from this and we all go, if a few of us go home and look through all their things and run them all through Zone Master and see if we have got a few inconsistencies that we need to fix, that would be a good outcome for this.

DNSSEC. DNSSEC will catch man-of-the-middle attack. My thing would have not worked if it had been attempted, first of all, because the IANA mail servers would have not accepted the e-mail. It would have queried, queried, queried until it reached a validatable mail server and then communicated with me. That would not have affected her problem because it was not mail-related.

.GT would not have been affected. Only somebody would have been able to read active directory updates for a host that was in that thing. Whether that is a good or a bad thing is debatable.

I don't really care about what Windows users do. It's their own fault when they use Windows. But the problem is it's done so, we got this from a university, a big data center, slightly misconfigured. We got a huge number of leaked queries from them and they were very much embarrassed and very polite when we pointed out politely that we saw the traffic from them because they figured out fairly quickly what the problem is.

So the take-home message is don't believe when it's working that it is correct. If it works on the first attempt, it is usually wrong, and as we all know, it's easy to break but difficult to fix. And now, Robert Martin-Legène from PCH.

ROBERT MARTIN-LEGÈNE: Is it question time?

I think I might have misunderstood what you said about the MNAME actually being used as a name server for the domain. I don't think that's the supposed behavior. Only the listed NS entries should be queried for authoritative data. That's what I see in 1035, at least. Has that changed?

EBERHARD LISSE: Nope. 1035, the ones that are listed, I don't want to go through.

ROBERT MARTIN-LEGÈNE: Yes, 1035 is the first one. It says that the MNAME is just the primary server. It doesn't mean that it's an authoritative server you can reach.

EBERHARD LISSE: It's the primary and the NS entry for the one in the NS is facultative so you can leave it away and at least active directory queries it, whether this is right or wrong.

ROBERT MARTIN-LEGÈNE: Well, it says that I make update to the MNAME as [inaudible].

EBERHARD LISSE: The point is you must have it and you must pay attention. Do you have it?

ROBERT MARTIN-LEGÈNE: Yes.

EBERHARD LISSE: And you must pay attention to it.

ROBERT MARTIN-LEGÈNE: Yeah, but.

EBERHARD LISSE: And only looking at the three, you're having it and then not synchronizing it with the SOA is a fault.

ROBERT MARTIN-LEGÈNE: What do you mean "synchronizing it"?

EBERHARD LISSE: If you have three NS [inaudible] and one primary, the primary should be identical to the MNAME. And when you change it, you must also change the MNAME.

ROBERT MARTIN-LEGÈNE: No.

EBERHARD LISSE: That's what, that's how I read the RFCs and I did some [inaudible].

ROBERT MARTIN-LEGÈNE: Well, yeah, okay. You could say that, but it could be a hidden master that's behind a firewall. It would still be the primary source of data with an address, but it might not be one of the authoritative names that you allow to query.

EBERHARD LISSE: They will query it. It's query-able.

ROBERT MARTIN-LEGÈNE: Well, I understand the Active Directory issue where they would send a Dynamic Update. But actually, have you noticed that the updates that you are receiving is just for non-existent system domains because the existed domains, existing domains, they will be something, well, let's say lisa.NA, that would actually be a domain name that exists, so they will go into and look at the SOA of that domain name.

EBERHARD LISSE: I didn't look at that. We noticed a bit of traffic when we actually looked at what's coming over a while and we found that we got a large number of queries for domains that do not belong on this name server, obviously.

ROBERT MARTIN-LEGÈNE: It would be a good place to register that domain for somebody who wants to do drug catching or whatever, and he would start seeing funny traffic maybe. Who knows?

EBERHARD LISSE: Yeah, but that's the point we are making.

ROBERT MARTIN-LEGÈNE: Yeah. I don't think I am willing to agree with your NS and MNAME argument just so easily.

EBERHARD LISSE: Whether you agree or not is not my problem. The point is that you must pay attention to it, and if you don't write RFCs and only read them, and only read them occasionally like me, if you get caught, you fix it the way that you don't get caught anymore and start reading the RFCs again once in a while and you get surprised what you forgot.

ROBERT MARTIN-LEGÈNE: Yeah.

EBERHARD LISSE: Any other questions? Preferably of authors of RFCs in this regard.

Okay, then Theo, thank you very much. Theo Kramer will summarize. Thank you. You can do it from standing here.

THEO KRAMER: Okay. First of all, my name is Theo Kramer. I'm with a company called Domain Name Services. We provide backend registry services.

But I think the real message over here, it's really good to see the attendance over here and I hope that you guys all enjoy the amenities that Johannesburg have to offer.

And in regarding some of the talks, yeah, it was really interesting to hear some of the DNSSEC experiences from some of our colleagues, [ourselves] as well, and good to see that some of the expertise is now being exported from this part of the world to the rest of the world.

Also good to look at some of the services being made available for organizations like AFRINIC to the, to this particular region, and also good to hear about some of the monitoring, some of the continued monitoring that's happening, services from ICANN, from other organizations regarding zone health, domain health, and that kind of thing, I think can only lead to a much improved ecosystem for the domain name word and the arena that we find ourselves in.

And then, you know, just concluding with all of that, Eberhard, especially to you, thanks hugely for championing the Tech Day at ICANN. And yeah, then to the rest of you, enjoy Johannesburg. Thank you very much.

[END OF TRANSCRIPTION]