COPENHAGEN – Tech Day (Part 3)
Monday, March 13, 2017 – 15:15 to 16:45 CET
ICANN58 | Copenhagen, Denmark

BENNO OVEREINDER: …spin off a number of other work. RFC 7258, pervasive monitoring is an attack on privacy. And it gives you some idea why engineers should care about privacy and that pervasive monitoring is bad for the Internet citizen.

With that, there's another draft, RFC 7624, Confidentiality in the Face of Pervasive Surveillance, and it gives a threat analysis and threat model based on existing attacks actually on privacy. So, this is all good read, and it's the background what's going on in the past two years in the IETF.

But wait, DNS and privacy? Don't we consider DNS data to be public? Well, it depends, of course. The zone files are public, but I think for most people, for organizations, for ISPs that run recursive or authoritative nameservers, TLDs or organizations, the traffic arriving and the queries at their servers is considered in most countries to be privacy-sensitive information.

So, just depending on which question I ask or how I frame the question, people will raise their hand, "Is it privacy or not privacy-sensitive data?" So, this has been all written down by

Stéphane Bortzmeyer from AFNIC. DNS Privacy Considerations, it's a good read and it debunks the alleged public nature of DNS.

The DNS data in the zones is public, but the transactions should not be public information. So that's kind of the setting of what I want to talk and continue to discuss. So, if I hear or you see the word attack, think of attack on privacy, not attack on your infrastructure. And these attacks are from the RFC of Stéphane Bortzmeyer, so the DNS Privacy Considerations.

There are many more. I just pick up two because we have worked on that. The first mile or last mile, depending on the perspective you have. For the user it's the first mile, for the ISP it's the last mile. There's a lot of information being shared of the end user to anyone who's listening on the line. Although I'm encrypting everything, my e-mail is encrypted, my HTTPS is encrypted, etc., but my DNS traffic is not. That is out in the plain.

And if with pervasive surveillance monitoring, lines are being listened to, quite some information can be deducted from your DNS queries. It can follow of course with information I asked for, but also if I ask for SOC records or with [inaudible] Open PGP it can also infer which person I want to communicate to. So this is some information you want to protect from surveillance.

On the other side, DNS is also leaking information it doesn't necessarily need to do. So if I ask for the schedule.ICANN.org to

my resolver, the resolver – clean cache – goes to the root, asks for .org apparently, I think. It finds the delegation .org, it asks for .org for ICANN, it gets the nameserver for ICANN, and then it asks for the schedule.icann.

So all the authoritative nameservers here get the full query, and of course, the way the authoritative nameserver can correlate this data with the users much less than at the resolver because behind the resolver are a number of users and things are in cache. But still, information is being shared, and if I run my resolver on my local laptop – which I do – it doesn't matter if I run it behind a resolver because it's directly identified to my laptop.

Or if currently it's being used more and more, it's EDNS client subnet. Part of my – in the query, in the EDNS option – my subnet is shared with the authoritative nameserver because the CDNs use this kind of information to give the optimal server to serve content from a close by server to my laptop or to my session. So more and more information is being shared also by the authoritative nameservers.

There's much more to tell here. I just want to point out there's an excellent IETF tutorial by Sara Dickinson from Sinodun that takes about two hours to really go into this DNS privacy problem, statement, approaches, etc. The URL is at the end of

the slide, and it also tells about the current IETF RFCs drafts and working group activities. So I really want to limit here to these two use cases or attacks on privacy.

The implementation, I want to now make it really practical, so I little bit sketched the problem here. I mentioned IETF Hackathon, but are also working with other people in projects. We worked on a number of these issues here and tried to find the solution.

So first, protecting the first or last mile. This has been a lot of work going also into DPRIVE working group, the DNS Privacy Working Group, and one of the strong factors or strong points which get attention here in DPRIVE Working Group is encryption of your DNS traffic on the first mile. And of course, there are a number of options here.

There's the opportunistic TLS, STARTTLS; [strict] TLS; datagram TLS; confidential DNS draft. That wasn't pursued, but it was more just a DNS-like solution not using existing technology but just to span the design space. And existing DNSCurve and DNSCrypt work, but that's not in the IETF and so it was not really further considered in the working activities.

So the encryption of the first mile settled on DNS over TLS. I just saw Paul Wouters here, but he also proposed something else, using IPSEC on the first mile or to your resolver. But most of the

work we've done and within the working group was on DNS over TLS. And that's good, it works because we can use all the new TLS 1.3 features which are really very handy for the problem at hand.

But also, we need to fix some parts of TLS or of the DNS for TCP/TLS. To optimize the session setup is much more expensive for TCP and TLS, but TCP Fast Open and TLS session resumption solve most of these problems. So other working groups did solve our problem here.

Robust TCP management is necessary because you have to deal with 10,000 or more, not necessarily here at the resolver, but in order of thousands TCP sessions. We can learn from the [ICTP] server and community. And we have to deal with kind of, well, [it's] TCP and streaming, pipelining, and out of order processing, which I go a little bit more in detail here.

So with UDP, you can send out a series of queries and they can come back in different order. With TCP pipelining, TCP, what goes in kind of first in first out, and that can hamper the performance. So, if query A takes a little bit more time, then B and C which are already ready, have to wait until A is answered, and then they can be pipelined back to the user.

With out of order processing, it needs some changes to the resolver that deals with TCP pipes. A, B, and C can be requested,

B and C are already answered and directed back, and A is later replied and sent back. These kind of things are – this is the ground work. It has to all to be done, and it has been done in resolvers. So in the [inaudible] not resolver and Unbound, this is all implemented now.

What about the other thing? We just discussed about encrypting the first mile, but now leaking information with authoritative nameservers. Reducing DNS information leakage, QNAME minimization was proposed, also by Stéphane Bortzmeyer. And what does it mean? QNAME minimization is actually I only share the information which is really necessary, which is relevant to the authoritative nameserver.

So root doesn't need to see the whole schedule.icann.org query, .org is sufficient because it has to give the delegation to org. Org only needs to see ICANN because it has to answer the delegation to the icann.org nameserver, and schedule.icann.org can be directed to the nameserver of ICANN.

This seems to be quite straightforward. It has been implemented, but as usual, nothing is that easy on the Internet with a lot of legacy – not necessarily legacy software, but broken legacy software. And especially also we've seen things – CDNs again – they do nice DNS tricks, and DNS tricks don't work well with our QNAME minimization.

One of the things we could solve directly and we thought we are in a good shape, we put it in Unbound and people were testing it using [and they thought,] "Well, it's a blackhole. I send in queries for this name and it disappears." And it ended up that we asked explicitly for .org and .icann.org NS records, which means the nameserver. But some, I don't know which middlebox it is or it's the nameserver itself, they don't give an answer to NS queries, they just drop them. It's dead. Don't get "server failed,"] don't get "[inaudible] domain," whatever, just silence.

So now, we ask for the A-record of .org to the root, and it's fine, actually. We get an answer because it doesn't have the A-record so it goes, "Well, you should go to .org," etc. And actually, as a coincidence, this might be even better, because if people now are monitoring these kind of queries and they see NS records versus A-records, A-records are quite normal. NS records? "Well, maybe now you are enabling QNAME minimization," and I don't know. So, it was maybe a really paranoid perspective on this, but A-records will work fine with this.

So the deployment. What I am doing [with time?] Okay, I've got ten minutes. Great. So we have implemented all these different parts, so the deployment of DNS privacy and enhanced DNS services can take a start.

For the first mile, we have DNS over TLS. For the authoritative, we have QNAME minimization. Let's go first for the DNS over TLS. Oh, sorry, you need a stub resolver and a resolver that are DNS over TLS capable. So, for deployment of DNS, the stub resolver – and I give this as an example, there might be other implementations also, sorry for not mentioning them – you can use the getdns as a stub.

So, getdns is a project, it's a library. Actually, you can link your application with getdns API library, and then your application will actually have its own stub resolver and can even fall back if necessary to full recursive. But you can also run getdns as a daemon, as a stub resolver, local on your laptop for example, so then other applications can use that as a stub resolver.

And there are all the DNSSEC goodies. DNSSEC roadblocks, DNS64 even with IPv4 can be validated. And it implements also DNS privacy extensions, so DNS over TLS. And Stubby is an instance of the getdns stub resolver with all the privacy options enabled, so you can install that, run it, and then you're in good shape from your stub part.

The privacy enhanced resolvers, there are three implementations we know of. The Unbound and Knot Resolver, they do implement TLS support within their resolver, and Bind has a how-to documentation how you can run Bind with a TLS

proxy. [OSARA] actually is using that in their test setup with nginx and HAProxy.

Mentioning the test DNS resolvers, TLS resolvers, NLnet Labs and OARC are running an Unbound kind of open resolver. SURFnet/Sinodun use Bind, HAProxy or nginx solution, and Daniel Kahn Gilmore use Knot Resolver as a kind of service to the public. So, if you don't want to run it yourself, your DNS privacy enhanced resolver, you can go to one of these open resolvers. They are TCP/TLS only. For references to IP addresses, etc., you can go to dnsprivacy.net.

Good, so for the QNAME minimization enabled resolvers, again you can use Knot Resolver, Unbound, and a future release – it's on the future release list – of Bind. They definitely intend to implement this. Good. Wrapping up. Let's see. So, just for time, I won't go here into details.

This is more for reference, so if you look back to the slides you can find some references here. There are two websites. Currently, they're pointed to the same actually website, dnsprivacy.org and dnsprivacy.net. It's either for the community, non-technical or the corporate users. And I want to acknowledge some of my colleagues here.

Sara Dickinson for giving good presentations, and I lent some of her content to this presentation. Allison and Willem and the rest of the getdns team. So, I'm open for any questions. Quiet.

UNIDENTIFIED MALE: Come on.

UNIDENTIFIED MALE: Go on.

UNIDENTIFIED MALE: No questions? Okay.

UNIDENTIFIED MALE: So, I've – kind of, yes, question for -

UNIDENTIFIED MALE: There is one. There is two.

[ALEX MAYOVER]: Alex [Mayover]. Did you get to implement DNS zero pairing yet for [inaudible]?

BENNO OVEREINDER: I don't know about – I think it's for the getdns.

UNIDENTIFIED MALE: Yes, it's in getdns, but I don't know whether it's [inaudible].

BENNO OVEREINDER: Yes. DKG, Daniel Kahn Gilmore is working on that. Yes.

[ALEX MAYOVER]: Okay.

BENNO OVEREINDER: I don't know the status, actually, but I know they are working on that, yes.

[ALEX MAYOVER]: Okay, cool.

BENNO OVEREINDER: Thanks.

BRETT CARR: Brett Carr, Nominet. QNAME minimization, that's not turned on by default, is it?

BENNO OVEREINDER: No.

BRETT CARR: Do you plan to turn it on by default?

BENNO OVEREINDER: It's a good question, and I think maybe I can bounce this question back to you and to Alexander and to Roy, actually. I went to your presentation this morning. Maybe it influences your measurements, but certainly Roy suggesting to help you out with QNAME minimization at the root, you only see .AT and not the secondary domains, SLDs.

So then it's more a general question to the TLDs I think. What will be the impact of QNAME minimization for TLDs? Because currently, TLDs are collecting data and they can help ISPs and [inaudible] signaling botnet attacks or any other spam, whatever, the visibility for you will become less.

UNIDENTIFIED MALE: Are you trying to start a discussion? Because just a very brief comment on that.

BENNO OVEREINDER: Yes.

UNIDENTIFIED MALE: DNS magnitude, as we presented this morning, would not be affected by QNAME minimization.

BENNO OVEREINDER: No.

UNIDENTIFIED MALE: But surely, Roy wouldn't see numbers that we could compare amongst each other. On the other hand, what I see in the long term is most of the DNS operators are using techniques to report copy of the name server and to statistics on a different host, and that's going to change.

If the nameserver itself is the only piece of infrastructure that actually knows everything about a DNS query and everything else is encrypted, then I think I will see a push towards the implementers of nameserver software to actually vastly increase their capabilities of statistics generation because PCAPS are not going to be useful anymore.

UNIDENTIFIED MALE: [Marco Pissori].

UNIDENTIFIED MALE: Okay, we're taking two more questions. Yes.

[MARCO PISSORI]: Okay. Last time I have seen a presentation regarding the QNAME -

UNIDENTIFIED MALE: Identify yourself for the remote audience, please.

[MARCO PISSORI]: Repeat it?

UNIDENTIFIED MALE: Your name please.

[MARCO PISSORI]: [Marco Pissori,] I said before. Okay. Last time I've seen a presentation regarding the QNAME minimization was at DNS OARC in Texas last year.

BENNO OVEREINDER: Yes.

[MARCO PISSORI]: And I remember that the one who was speaking was referring to the fact that at least in [two] cases he found problems with some

providers, and one was Akamai. I remember that he told that he reported this to them and they eventually fixed it.

BENNO OVEREINDER: Yes.

[MARCO PISSORI]: So, my question now is, in your experience, did you find other cases in which QNAME minimization could produce issues to the resolvers?

BENNO OVEREINDER: Yes. Indeed, Akamai issue was solved at that time already I think more than one year ago. I forgot the CDN, but at least that CDN – so other users were using that in production, and then indeed we found out that some of the CDNs don't answer NS queries. They just drop them. That was the only thing we encountered and heard of, actually, because we don't run our large infrastructure ourselves, but that's what we heard. We changed from NS queries to A-record queries and everything worked fine.

[MARCO PISSORI]: The problem I remember was related to the empty non-terminal zones that were not answering correctly.

BENNO OVEREINDER: That was the Akamai thing.

[MARCO PISSORI]: Yes.

BENNO OVEREINDER: Yes, that's correct.

GEOFF HOUSTON: I just had to react. The issue is, what you regard as stats I regard as unacceptable information leakage, and you can't have both. And if you really want in this post-Snowden world to stop telling the world what names you're going to, then your stats are going to die. And there's nothing you can do about it.

BENNO OVEREINDER: Yes.

GEOFF HOUSTON: Because if you really want to hide what I'm asking for in a terminal name to everyone except that last terminal nameserver, then that's what QNAME minimization gives me and that's what I want as an end user. And I feel sad for all your stats. Actually, I don't. I don't feel sad in the slightest.

I think you were basically latching onto a protocol that was way too loquacious and way too chatty, and you're addicted to it. Tough. And that's just the way it goes, and it's the same with encrypted channels.

So, so far if I get you right, I can now have a secret between me and Google, and the rest of you, tough. That if I encrypt my channel to a recursive resolver, a large one, you can't tell it's me from the recursive resolver out to the authoritative. If you look on the wire between me and the recursive, you can't see a thing.

And quite frankly, with QNAME minimization, the terminal name is basically a very limited visibility of information. So either you have stats and you have a DNS which is a gigantic looking glass into users, or you have a slightly better system that's more user friendly and go do stats somewhere else.

BENNO OVEREINDER: Yes. And we give a little bit back to the persons who run the resolvers and authoritative names. If you enable QNAME minimization, you get [inaudible] domain hardening, so you get a little bit less traffic. So that's a good thing for the infrastructure operators. Good. Thank you for your attention.

UNIDENTIFIED MALE: Thank you very much. So now, Don Slaunwhite will tell us a little bit about business intelligence at CIRA.

DON SLAUNWHITE: Okay. Good afternoon, and thank you for everybody for hanging in this long. I'll try and make this as quick and as enjoyable as possible. So again, my name is Don Slaunwhite. For those who don't know me, I'm a product manager at CIRA and currently working with .ca.

As part of a project that we've been working on for the last few years, we've been building a new registry platform called Fury. One of the mandates of that was to try and move away our business intelligence operation from what we currently have, which is an IBM Cognos and Oracle-based system which is pretty expensive to run and to operate, and move into a more open source environment.

So as part of that maneuver, we've started the work to move to a Pentaho system, and today what I'm going to do is give a little brief description about what we had to do and the Pentaho open source environment itself.

As you could see here, we're going to take a look at the architecture, some ETL, the data warehouse design, reporting

and analytics, and then the testing and change methodologies that we have to do to make sure that everything runs well.

For those who don't know, Pentaho is an open source environment. Right now, there's Pentaho 7, and it has a variety of components. For the ETL and data migration work, there is a component called Kettle. They have a Pentaho Reporting environment, they've got Mondrian for an OLAP server, they have Data Mining in terms of the Weka product, and they have dashboard creation and other tools under something called CTools.

In our environment, we laid out – Hang on, let me press the button. I am pressing the right button, right? Point. There we go. In our environment, we've laid out a BI architecture. For those who are familiar with developing it, there is nothing really abnormal about this one.

We've got our source layer – and for those who had just a good discussion about collecting DNS data a minute ago, we do collect our DNS data as part of the bigdata capture. We also capture all the data that's representative in our registry, and because our new registry is multitenant capable, we're actually pulling in the registry information for any registry that's being operated on.

We're also capturing all of our EPP logs, all of our logging data, our WHOIS query data, and we're using Splunk to pull that all together and to push it into the Pentaho environment. The next layer is what we'll talk about shortly, and that's our ETL later where we do all the data transformations.

So on this layer here, we're using a variety of tools that come with Pentaho. The main one is called Spoon. They have a big thing for kitchen utensils, so it's called Spoon and that's the IDE that actually goes through and you build up all your transformations and your data extraction tools from there.

They use Kitchen and Pan as the command line processors to execute on those transform commands that you have built, and they use something called Carte as their HTTP server, which allows you to parallelize your operations.

So, if you're trying to get faster times for your ETL loads, you simply run more Carte servers and work out horizontally as you scale through it. In doing so, for those who aren't familiar with basic ETL stuff, Spoon is used – as I mentioned – to create those transformations.

Here's an example of a transformation that we're doing right now to generate our ICANN reports that we have to provide monthly, where we're pulling in our information from the total

domain name queries, the different transfers, the different detailed domain records, etc.

The little bit on the left there – which you may or may not be able to see very well unfortunately – is just to sort of show that Pentaho has a great number of import capabilities from bigdata to SQL to relational tables, as well as file manipulation. So at each step of the process, you have the capability of moving data in and out very easily, and it all works really well.

This is an example that we have right now of one of our jobs, which is preforming some of the ETL work. It's a simple thing. You're starting through, you're integrating through. You'll notice that we're pushing all of our error logging into a specific error handling file, and again, Pentaho makes this very easy to do.

Each of these steps is all logged, tracked, and monitored, and that's actually a really nice feature of Pentaho that we found that came out, which I'll show in about two slides from now.

At the lowest level is your actual transformation. Each of those steps we're looking at would run through a variety of work to pull in the data from the registry in terms of the domain registry, the contact registrations, etc., and it pulls it up together.

Let me jump back one. So, when I mentioned about the logging, Pentaho is really good for capturing the execution results of all

of the ETL work that you're doing. This one isn't too excited, because if you're really keen and you have a good eye, you can see that the read/write updates and inputs, they're all zero because nothing was actually run.

But this is nice because you can take a look through, and as you're doing this again and again, you're going to see a common pattern. It makes it easier for you to notify and look at where you're going to see errors in your execution and if you're seeing the right number of records that you're looking for.

In terms of performance tuning, they also provide metrics in terms of time for execution for the variety of steps that are within your ETL jobs. Some of these ETL jobs can become very complicated and long, so this gives you an easy way to visually take a look at it and say, "Okay, where can I maximize my benefit?"

Anybody who's done this sort of work in the past, you'll know that the people are asking for a ten-minute turnaround to load the data and it's now taking you an hour, so having this sort of information up front and available is very useful.

The next level that we'll take a quick look at is our data warehouse. In a typical fashion with a business intelligence model, you're building really sort of two basic models, and the

tools that are available from Pentaho allow you to do that quite nicely.

The first model would be your star schema, and this is your dimensional model where you're going to be doing all the analytics. There's a tool called Saiku which is available. It's not part of the box, but it's part of the marketplace, and the Pentaho marketplace has a whole slew of plugins that you can extend the ability of Pentaho itself.

Some of them are good, some of them aren't, but Saiku is a really good one. It allows you to do analytic capabilities quite quickly. In this particular star scheme, we're just puling in all the fact models that you've seen generated through the transforms in the last slides.

We're highlighting one red area right there, because again, from a Fury point of view, this is something different that we're doing that we've never done before. In .ca, we only had one registry. But now that we're running multiple registries, we're able to isolate those out and in the data warehouse pull it all together. So although we're keeping a separation of data from a registry operation point of view, when we go to the reporting side of things we're able to pull that in together.

For querying detailed information – you want to get contact information or registrant information, etc. – we do have a

relational model that is also used, and this can be used by the report author tool and actually Saiku as well.

In this, there isn't really anything particularly outstanding, but again, Saiku allows you to do all of this stuff and more. So from the reporting and analytics side, this is where we use the variety of tools that Pentaho provides. We're building the cubes for the relational work and we're building reports out of it.

I'll just quickly go through some of the reporting types that you can do and the tools that we're going to use to do them. So, the BA, Pentaho itself has a business analytics server, and this is the area where you're actually going to place your dashboards, you're going to allow people to execute on reports, you're allowed to schedule the reports, do bursting of e-mail reports, and it's all set up here.

So as an end user, you provide them with – as an administrator, you have your own passwords in here and you can set up the data sources, set up all the different reports – and as an end user, you're able to come and consume those.

This is an example of a very simple registry dashboard that we created showing your domains under management. The first line there is the registration, second line is renewals, etc. So it allows people to get a very good visualization on the registry operation.

But when you need more detailed work, they do have something called the Report Designer, and this is a pixel perfect report designer that allows you to pull in the data from all of those areas that we talked about – the cubes, the relational databases as well – and you can essentially build any sort of report that you want.

It is a bit more complicated than some of the other BI tools out there, so when we made the switch from Cognos and we were moving into this, we found that you have to have a bit more technical capability behind and an understanding.

As well, some of the functionality that they had wasn't documented as well as it could have been. But as long as you understand what you're doing, it works very well.

This would be a type of report that we would do. This happens to be a registrar billing report. The nice part of Pentaho as well is that you can develop drill throughs, so if you're reviewing the [HTMI] version of this versus a PDF, you'd be able to see all of the events that are being transacted upon, and if you find one of interest you can click on it and drill down to the next level of report and get that information. You can also drill up as well, but it makes it very nice for both the registry and the registrars who are using these reports.

Again, I mentioned Saiku in terms of doing the analytics, and if you take a look at this chart, you'll see on the left hand side that we have the measure here. We create all of these cubes in the tool ahead of time, so this is what the user would be able to use. They could drag in domains under management and then pull in any dimension that we've already created for them. So it's an easy way of saying, "Show me all the registrations by a registrar for a particular time period."

And Saiku also has all the data visualizations that you would really possibly want. It literally does everything. So these can be used ad hoc. They can be put into reports, PDFs, Excel files, pretty much everything you could think of, as well as they can be embedded into the dashboards.

Speaking of dashboards, they actually do have a community edition for a dashboard creation that would be part of the CTools, and this one takes a little bit more. You have to have some programming background if you're comfortable doing web development, but for a report author, they might need to get a bit more understanding because it requires some JavaScript and some HTML work there. But once it's done, it can be embedded and used in a variety of areas as well.

This is an example where we're actually embedding data directly into the Fury registry portal, so this would be an editing or a

creation of a domain within Fury from a registry operation's point of view. And down at the bottom, which may go away based off the last conversation, you'll see that there are DNS query volumes for that particular domain name as well as the WHOIS queries that are coming in.

So this can be used for the registry to get a better understanding of the usage of their domains and to see where things are happening to look for abuse and other items like that. But another option too would be to have this information available to the registrant themselves and allow stuff to be pushed through so that they could see what is going on with their own domain.

So again, all of this is very easy to do once you have your data pulled in and you've developed this sort of a warehouse. Now, in doing this, there are a lot of moving parts, so I just wanted to quickly go over some of the testing and change management that we use at CIRA from an IM point of view.

Up in the far left in the circle there is essentially a local machine. They're going through the testing, utilizing the tools, building the models, building the ETL. Then it's pushed into Git. We do use Git for our source repository. We do peer review at each step so we have another IM person taking a look.

After the peer review and discussion goes back, we then push it out to a fully qualified QA environment that's identical to our production and we run through testing. We run a regression suite against the entire infrastructure and make sure that we haven't broken anything with our new changes.

Once everything is complete, we can then push that out into our OT&E environment, again monitor and then push it out directly into production. This is really important because you do have a lot of power and capability with Pentaho. It works very well, but again like anything that you build, you should make sure that you're managing it properly in terms of the quality assurance and management.

Over this process, it's taken us about a year and a half to build this, and some of the takeaways that we've got for those who may be interested in using Pentaho for their projects moving up: we found that it works really well. We were quite pleased.

We did do evaluation of a variety of other tools earlier on and Pentaho did seem to meet all our needs, but you really only know how well it's going to work when you start to use it for real. And it does have a great deal of flexibility. You can pull in from regular sources, but it has a lot of bigdata sources as well.

The PDI integration with reporting worked really easily, and the ETL was quite good as well. We've basically found that it does

everything that it does everything that we need. The logging, the framework works.

The only thing that we found that it's missing right now is in the community edition, which is the open source one, there's no real scheduling infrastructure. So because we happen to be running our BI platform right now on a Windows environment, we're simply using Windows scheduler. But if you're on Linux or otherwise, you could do a Cron job on it for this. That type of scheduling [though] on more advanced work is in the enterprise edition if you're willing to be paying for it.

The final takeaways really are that the tools are good. They work really well. They're not as easy to use, so if you're coming from a traditional, more established BI environment, there will be a little bit more of a learning curve for your IM team, but it does work out in the end.

As I mentioned before, the reporting abilities and dashboard creation require some coding, so it's not just drag and drop. And the documentation, although it is available, is not quite as thorough as you might like in some cases.

But there is an active community, and that was one of the biggest reasons that we looked at Pentaho and chose it, was that there was an active community of people who are building not only plugins but information and reports as well.

The last thing I'd like to do is just thank the people who are actually doing the work. This is our IM team in CIRA, it's Jon, June, Shanshan and Namita, and they're the real experts on this. Luckily, I'm the one who came here and [Jacques] said, "Hey, do a presentation!" So I said, "Okay, I'll do a presentation."

But if you do have any questions, I'm going to be providing at the end of the slideshow Jon Coote's e-mail address. He knows everything there is to know about Pentaho, so anybody who's looking at using it or has questions surrounding it, you're more than happy to contact him.

But in the meantime, I'll take any questions here that we may have.

[JACQUES LATOUR]: You won't get away that easy. On the agenda, every presenter is clickable with his e-mail address.

[EBERHARD LISSE]: Any questions? Alright, thank you very much.

DON SLAUNWHITE: Thank you very much.

[EBERHARD LISSE]: So, I haven't seen Bobby Flaim. Is he around? Oh, there you are. Okay. Bobby Flaim is – if I'm not mistaken – a senior special agent with the FBI, and he has on occasion spoken at Tech Day about issues that affect both law enforcement and us from a technical perspective.

As we have heard about the Dyn issue in our earlier, when we were considering this I thought not only do I want him to come and give us a little presentation with his colleagues or himself, I also wanted to have him in a good mood that maybe on the way down to the next ICANN meeting he can stop over [in Namibia] for a day or two and help us get our [CERT] up and running. So, I'm nailing you, right? Or am I saving you?

BOBBY FLAIM: Thank you, [Eberhard], and thank you for having me here.

[EBERHARD LISSE]: [I'm sorry, you have got as much time as you want.]

BOBBY FLAIM: Okay, thank you. Thank you for having me. [Eberhard], thank you for inviting me. It's a real pleasure to come talk to all of you today because when I was here at the last ICANN meeting, I actually went up to a few of you and actually solicited your help.

And this presentation is really the fruits of that help that you have provided.

Are you guys familiar, have you heard of the Avalanche case? Just by a show of hands, anybody? Okay, so a few of you. I wasn't the actual cyber agent who conducted the actual investigation. It actually originated out of a local German police department, it was escalated to the German federal police, and then it became an international, coordinated event which I'll show you.

So I'm going to talk less about the technical details, although the slides will have some of that in there as reference, but I'm going to talk more about the cooperation that we had from all of you, the ccTLDs and also the gTLDs, some of the legal and process complexities and how long this took. I hope you'll find some of it interesting, and if you have any questions, please stop me.

Wrong device. Okay. Just to let you know, this presentation was actually supposed to be with myself and also Greg Mounier from Europol, and unfortunately he couldn't make it. So you'll see a lot of the slides actually have Europol and they actually were a huge part of this.

Just to give you a little – and he put a lot of graphics in here, so I'm going to apologize, but – no, one graphic too many. Okay, so just to give you the depth, Avalanche involved about a million

malicious domain names, and what we did was the takedown of this operation.

Avalanche operation actually occurred right after the Hyderabad meeting that ICANN had, and it was very critical, the timing, because I personally went to several of the ccTLDs right here in this room when you had your last ccNSO session to actually get the cooperation.

So November 30th was the international takedown day, and you'll see how very international it was, because it was coordinated across multiple continents, and about 30 countries were involved.

The primary players were the German national police, Europol, the FBI. I know the Royal Canadian Mounted Police had a big part of this. Interpol also played a big part of it, so you're going to see as we go through this the international nature, not only from law enforcement but how we worked with all of you as well.

Avalanche – just to give you a brief description – is a criminal service/bulletproof hosting, not really a botnet, that has been operating since 2010. If some of you remember the GOZ cryptolocker case from 2014, that was another case that we also worked on very extensively with ICANN, the security team, and also a lot of the gTLDs, namely Verisign, Afilias, Neustar, and at

that time that was actually a very good proving ground for this particular case, because at that time we established a lot of what was done here.

Namely, establishing number one a unified or simultaneous takedown. We also were able to work on our legal processes – whether it's MLATs, whether it's voluntary takedowns, whether it was court orders – and dealing specifically with the security teams at those registries.

So, that actually worked very effective. When we did that, I believe it was May 30th in 2014, and that also was a simultaneous takedown and that helped us with this Avalanche case which we did on November 30th, 2016.

As you see, the monetary losses were quite substantial. We'll go into that a little bit. But the customers who would use the Avalanche communications were basically nefarious individuals who wanted to use botnets for their own purposes, to perpetrate fraud, to get money, ransomware, spear phishing, so on and so forth.

The Avalanche network provides its customers with bulletproof hosting for their malware via double fast-fluxing domains, so that is something obviously that involves the DNS. Changing the IP and the domain name in very succession, which makes it a

very difficult problem to solve. And also, the obfuscation of backend information.

So, as you see, part of the problem with Avalanche was that it would generate what we call DGAs or domain generation algorithms, which are domain names that kind of make no sense. They're very long strings, a bunch of gobbledygook more or less. Sometimes they can go up to 30 or 40 characters, so really not making sense and it's basically a botnet that's actually doing that in the background. This is just an example of some of the domain names that were used.

So the role of law enforcement in doing this was – sorry, I'm just trying to make sure I have my slides in the right order – was going to the top level domain names, and it required a lot of coordination, securing the information exchange between ourselves, doing the analysis. Actually, a German security company had done a lot of the backend analysis originally, and they provided a lot of help. Working with our operational cyber coordination centers within our own national police, and also working very effectively – like I had mentioned – with the gTLDs and also you, the ccTLDs with the technical aspects.

One thing that we also did on May 30th was we set up a 24/7 command post where everyone was manning it simultaneously

to ensure that we would do the takedown at the exact time because that was very critical.

Another thing that was very critical is we had to make sure that with all the TLDs, we had reached out to them and that it was all or none because even if we left out one, that would be a vulnerability which would have the potential to perpetrate this botnet or Avalanche or DGA. That obviously was a big problem, so we had to make sure that we were very coordinated and it was very simultaneous.

As you see here, these are some of the malware families that actually were impacted. You see one of them, GozNym, which is a successor or an evolution of the GOZ cryptolocker malware. And there were a lot of different families, it wasn't just one, that actually were part of this operation.

Now, this over here – as I'm sure you can't see it, but you will have it as part of your reference materials. You passed the slides out, [Eberhard]? So you'll be able to see this, but it's actually very small.

A lot of the servers, as you will see based on the flags, [it went] Ukraine, Canada, Sweden, and Russia. And this just gives you a little snapshot of some of the TLDs. There were 64 in all, affecting 830,000 domain names.

And you'll see the cross between the gTLDs and the ccTLDs, so you see .com and .net were at the top, and that's pretty much based on volume and their impact. But you'll also see number three which is India, the .in ccTLD. That was one of the ccTLDs that we actually were able to work with at the Hyderabad meeting very effectively.

I know that Shadowserver – which I will talk about a little bit – as the registrar of last resort actually went to New Delhi and worked with the Indian government to ensure that that operation was going to work. And that was right after the ICANN meeting, so that actually was very important.

So like we said, the malicious domain names registered at specific TLDs were either seized via MLATs – and MLAT stands for mutual legal assistance treaty, which is a formal way that governments, especially the Departments of Justice, actually request and send information.

So in other words, if I as an FBI agent would like information from Germany, I [can't] subpoena or issue a court order to a German ISP. I would have to go through the German government.

And how that works is – it's not convoluted, but it is a long process – I as the FBI would have to go to my Department of Justice, my U.S. Department of Justice would have to go to the

German Department of Justice, its equivalent. That German Department of Justice would have to go to the German police which would then serve the subpoena, the American subpoena which was converted into an MLAT, to the German ISP.

That obviously is something that has to be coordinated well ahead of time, because if you tried to do it in real time, it's a very lengthy legal process and it would not work in such a circumstance. Therefore, it had to be planned way in advance.

And this operation– the coordination aspect of it, not even the prior analysis or identifying victims – actually started six months prior. So six months prior to November 30th is actually where all the international police agencies such as Europol, FBI, German federal police actually started working on just coordinating the event on November 30th. And that obviously required a lot of coordination regarding the MLATs.

The second bullet that you see there is the malicious “unborn” domains, and this was the vast majority of the domains. And this presented a problem for law enforcement and also the gTLDs, because I think about only 1% of the actual DJAs were actually registered. That means they were actually in the zone file, they were actually live on the Internet.

Most of them, the vast majority were unborn, but because of the reverse engineering, we were able to discover what those

domain names would be and where they would be registered. So that meant going obviously to the ccTLDs and the gTLDs to say, "Hey, look, can you actually register them but sinkhole them?" Which means basically they quarantine them.

That was a challenge, because number one, that costs money. I know for some of the gTLDs, the scale of it and how they would do it presented some problems, so that required a lot of coordination on both of our parts.

And as you see there, the quarantine or the sinkholing of these domains actually went to a registrar of last resort, which is registered with Shadowserver. For those of you who know Shadowserver, they're kind of a security firm. They do a lot of abuse analysis.

But one thing that they have done was they set up a registrar of last resort, and that actually came about because of the GOZ cryptolocker case in 2014. So Shadowserver and the registrar of last resort only came into existence about 2015 and actually was very important in this particular case, because once these domains are registered, you need a registrar to do that. Not every registry had that ability or wanted to do that, so therefore Shadowserver had to step in and perform that role.

We also did a lot of victim identification, and you'll also see how that worked on a worldwide scale. If you look to the left of the

chart on the bottom, you'll see that was July 14th and that's where we had GOZ cryptolocker.

And if you look after November 14th, you'll see the graph go down, and you'll also see in the chart the redder the country, that means the more victims were there. So, that gives you a little bit of a perspective of where the victims were and the impact was in particular, and how we also used the Gameover Zeus – which is GOZ – and how the numbers decreased and how they flatlined after November. If you can see the chart – I'm sorry, I don't have my glasses – you'll see the bottoming, and then it went up a little bit, but then it's kind of at a low level now.

Insofar as the perpetrators or the actors, they were both Ukrainian gentlemen and they've both been arrested. You also had some major players who served as customers. They've been arrested, some of them have been extradited, so that is still an ongoing process. There are some issues. I think one of them may have been arrested and released, and I don't know what the actual status on that is, but that was a very interesting aspect of this case.

Both of the Ukrainians who we said were the main perpetrators were actually taken into custody and searched by Ukrainian law enforcement. You'll see Kapkanov – aka Firestarter – fired with a

rifle actually when they went to his house in the Ukraine. He actually greeted the police by firing a few rounds at the police.

That presented obviously a challenge, and you'll see what they found, $70,000 was found stashed in his closet. A number of items were seized and forensic analysis was actually done.

Germany also executed searches on Avalanche customers we had mentioned. German authorities took one of them actually into custody. And also on December 25th, one of the customers actually was arrested at Newark Airport and has now been extradited to Germany.

And just to show you, this is the main – this is Firestarter, so this is Kapkanov, and these pictures actually came from the Ukrainian police. Like I said, a very international aspect, and the Ukrainian police were the ones who did the arrest of the two alleged perpetrators. So, this is Kapkanov, and this is the gun that he used to fire at the police on November 30th.

This is his house. You see the police in the Ukraine getting ready to do some SWAT action there, and this is him on the ledge trying to escape. So just to give you some graphics. Everyone wants to see how arrests are conducted. This is how it went down actually in the Ukraine for the main perpetrator.

This is also a list of all the agencies and organizations that were involved. You can see it was quite a number. You see at the top was the U.S. Department of Justice, the U.S. Attorney's Office, but you also see the Public Prosecutor's Office of Verden, and that actually was the local German police and that's where it originated.

You can also see Europol, other police in Germany, the FBI, investigators and prosecutors in 40 jurisdictions. You see Singapore there. When I was here in Hyderabad, we went up to .sg, we worked with their ccTLD operators and they said, "Okay, this is very good, but we actually have to have our own police know about this. So can you coordinate with them to ensure that we are acting under the right authorities and we won't have any problems?"

So we had our FBI representative in Singapore actually work with the Singaporean police to work with .sg, to ensure that that went well. That is just one example where these particular jurisdictions worked with their ccTLDs to have this operation become a success.

You also see Shadowserver there, and the Shadowserver Foundation is actually the one who is behind the registrar of last resort. Another important player obviously was ICANN right here, their security department. Dave Conrad, John Crain, Dave

Piscitello. They actually were very critical to ensure the smooth operation of this.

ICANN actually has the expedited registry – I apologize, I don't know the acronym off the top of my head – but ICANN basically has a way to waive particular registration fees for cases like this. I think it actually started or developed out of Conficker in 2003. It was used in GOZ cryptolocker in 2014, and it was also used in this case. That actually helped a great deal with some of the financial aspects, the coordination aspects, and the legal aspects of this case.

These are just some of the results. Five arrests in four countries: U.S., Bulgaria, Ukraine, Germany; 37 searches in seven countries; 39 servers in 13 countries, 221 servers taken offline. As I mentioned earlier, there were 64 TLDs which is kind of split evenly between gTLDs and ccTLDs.

About 800,000 domains in 26 countries were either blocked, sinkholed, and we also had the victim remediation which is where we notified the victims, we coordinated with [CERTs] to ensure that they knew they were victims and what they can do, and again, a lot of it was awareness raising. This is just one of many different presentations that we – myself or Europol or the FBI – have done on this particular case.

And this was I guess one of our biggest successes to date, and it was really a culmination of a lot of different efforts and also working on prior cases and becoming more effective and having a lot of lessons learned where we are able to be more effective in going after large criminal enterprises that are operating on the Internet. So, this is a very good example.

You can see some of the other statistics and some of the other things that had an impact on the investigation: 5 million infections annually in 180 countries, so this was obviously a very big case for us. We're hoping that the impact and the effects of it are going to be long lasting, and it actually provided a great prototype, so we're hoping that in the future, this can serve as a model on how we cooperate. [Apologies.]

Legal challenges. Like I mentioned before, a lot of the legal challenges – and one of the reasons I reached out to some of the ccTLDs here – was because there were no MLAT agreements with some particular countries, so that is why we had to get some voluntary cooperation from them. That was an important aspect.

The legal consideration – as you know – in fighting cybercrime is that with different jurisdictions, it's very hard to get those legal orders or get information from those legal orders across borders. So again, the coordination was paramount. And like I

said, it took us six months just for the coordination of the takedown alone, not even the investigative part, so that was very critical.

And unfortunately, since there's no collective – there are cyber treaties, the Cyber Convention of Budapest that's been passed by many countries – there's still a very hard time in getting digital evidence across borders, the transnational, the transactional aspect of all of this, and the fleeting nature of digital evidence all of this makes all of this very difficult. But in this circumstance with a lot of preparation, it actually worked out very well.

So again, it takes a network to defeat a network, and that includes all of you. It's detect, disrupt, deter, and prevent, and that's basically the lessons learned. We thought we were very effective, but we were only effective because we worked with all of you and the community: the gTLDs, the ccTLDs, and ICANN.

And this is really a great example of the public/private partnership that we try to foster hear at ICANN, because a lot of agents, a lot of law enforcement, you don't see a lot of them here, especially the cyber agents. They're very wedded to their cases and the very specific actions and operations, and a lot of times we fail to see the larger ecosystem.

So by being present here, I think it was very good where you have those personal contacts, you have the ability to make those connections very quickly, which are very important in cases like this. So I think that is the end of my formal presentation, but if you have any questions or anything at all, I hope I can answer all of them.

[EBERHARD LISSE]: Thank you very much. Go ahead, please.

ASHELL FORDE: Hi. My name is Ashell Forde. I'm from Barbados. I am an ICANN newcomer and a fellow. Can you hear me okay? Is it better? So my question is about the TLDs. I wondered if there were any operators that don't participate in ICANN or aren't active that you were not able to reach when you came to ICANN, how you went about contacting them, working with them, and were you successful if there weren't parties that are actively in ICANN?

BOBBY FLAIM: Yes. That's a great question, and there is one particular example because one of the ccTLDs was not here at ICANN, and that was one of the ones that we needed to reach out to. So through [Eberhard's] help and a few others actually in this room, I was able to get the connection with that one particular ccTLD where

we were actually able to reach out to them and become effective.

So yes, even if they are not here, whether it's the ccTLD or another partner, being here at ICANN, having the connection with everyone, someone always knows someone. The six degrees of separation, that actually proved to be very effective. And this group in particular was the reason we were able to do that.

PABLO RODRIGUEZ: Good afternoon. For the record, my name is Pablo Rodriguez from the registry .pr in Puerto Rico. In the past, we have experienced difficulties and specific challenges of hacking and so on, and we have reached out to IC3 as well as we have become members of the InfraGard.

However, I would like to know if there is a program in which we could reach out to you or to the FBI and create programs in which we can help other people in our jurisdiction as well as the Virgin Islands where we can help out and teach merchants on how to protect themselves. Because many people don't know when they are attacked, where could they go and what can they do?

So we would like to do something in that spirit, especially since next year, March 2018, we're going to be there. ICANN is going to be in Puerto Rico, so that perhaps could be a good opportunity to do something like this.

BOBBY FLAIM: Totally. I think that is absolutely a great idea. And for those of you who don't know, InfraGard is actually an FBI program in which we liaison and work with our local community. We have a chapter in each one of our field offices. And in Puerto Rico and San Juan, we do have an FBI field office, so we do try to work with the community.

But that is a very good idea. Maybe that is something where we can have a law enforcement ccTLD training or awareness or something like that. But certainly willing to work with you in preparation for the March meeting.

[EBERHARD LISSE]: Just a little tidbit on the side. One of my colleagues from the Asia Pacific region told me that they were also involved and cooperated in a takedown, and when they looked, they found that the registrar used had very few domains, roughly none other than the ones that were involved.

The registrar had registered two years before and had never registered a single domain until they started to use them for these nefarious purposes. So that's maybe something how far they plan ahead. It doesn't really affect us from a technical perspective, but I found this thing very interesting.

Are there any questions from the remote? Okay. And again, as I was saying, we are trying [in Namibia to start a Namibian CERT]. If anyone who is doing stuff like this is interested to come two or three days before the next ICANN meeting which is in Johannesburg and travel through wonderful [Windhoek], just let me know because we really would like to have on the Friday before do a workshop where we can get our local law enforcement, which is starting to get involved, and they have identified a young person who can read and write to do this. So if anybody is interested, please get in touch with me.

Otherwise, thank you very much Bobby for your time, and that would be – give him a hand. And that would just leave [Norm Richey?] to close the proceedings as we normally do.

[NORM RITCHIE]: Sorry, I'm standing between [you and beer]. It has been a long day. Eberhard has asked me just to share some of my views on the session today. First of all, I was taken aback when I walked in

a bit late this morning and there are no seats. So, once again, Tech Day has an over capacity crowd, which I think is great.

If you go back a while ago, it was just a small group, everyone knew each other anyway and spoke on a regular basis. But I look around the room now and there are a lot of new faces here, which I think is wonderful. And maybe next time, we can get a bigger room.

First of all, Eberhard, thank you. I think it was a great set of presentations today. I think you did a really good job putting the agenda together. What I took away from today, probably two things. One, I saw a lot of presentations on measurements, on DNS, on [CERTS], on business analysis, and then with Jay Daley actually going a level deeper where he's looking at use cases for those metrics before he actually does the metrics, which I think is a great way of doing things. Too often, we actually just create metrics and try and find the use afterwards. So I thought that was very insightful.

I really liked the Dyn presentation. I think it takes a lot to come up here and talk about incidents that happen and how you dealt with them. In particular, this was pretty sensitive, so a lot of kudos for that.

And then Bobby's presentation on Avalanche, that needs to be celebrated. The community working together like that I think is

really good, which really comes down to my last point, and it's kind of like an underlying theme, that there are some grumblings out there that change is coming.

And Andrew alluded to this, is that we're in the new, post-Snowden world for privacy, and that needs to be resolved. We have ever increasing levels of abuse happening, and that needs to be resolved. And if we don't, someone's going to resolve this for us.

And quite often though, the solutions to address privacy and the solutions to do better research and work on security collide with each other. And that also has to be resolved. So, that's really my takeaways for today, and thank you.

[EBERHARD LISSE]: Thank you very much, everybody.

[END OF TRANSCRIPTION]