HELSINKI – Tech Day
Monday, June 27, 2016 – 13:30 to 18:30 EEST
ICANN56 | Helsinki, Finland

EBERHARD LISSE: Even if it's shorter, it was I think quite good and of course they sponsored our lunch. We got us a really nice program today. We'll start with the usual host presentation. We always invite the hosts to give us a presentation on what's going on with them and give us insight into whatever research they're doing, if any.

Juhani Juselius is from .fi is going to start then.

Mats Dufberg from IIS is going to give us another presentation, a follow-up on Zonemaster which is their tool that replaces the French tool that you can check DNS health of your domain of your TLD or whatever.

Patrick Wallström also from IIS is going to talk us about some changes in their registry. They are now allowing AXFR will give the zones away and they run two zones at least .se and .nu.

Then Jay Daley will speak about their new registrar portal.

While Arends will give us some insight about the recent DDoS attack on .ug which is probably Uganda, if I remember correctly.

Note: The following is the output resulting from transcribing an audio file into a word/text document. Although the transcription is largely accurate, in some cases may be incomplete or inaccurate due to inaudible passages and grammatical corrections. It is posted as an aid to the original audio file, but should not be treated as an authoritative record.

Norm Ritchie from the Secure Domain Foundation will talk about their tool, Luminous.

Then Garth Miller or Mustafa Rifaee, neither of whom I haven't seen yet, will speak about CoCCA tools which has got since recent of late a hook into the Secure Domain Foundation so at registration it can be checked whether the name servers of the domain or objects connected to those have been used for nefarious purposes.

Then come two of my favorite topics, What3Words – Gary Gale will talk about their project. What3Words is a very cool way of dividing the word into squares of nine square meters and addressing them which offers a lot of potential and a lot of possible ideas on how to use this.

And then Andrew Settle is going to talk about the JAKU Report that's basically a report on DDos attacks, what they do, how to detect it. And I'm wondering why he's wearing a tie. Yesterday he wasn't.

We are not strapped for time, so if there is question time, we have a little bit space at the end. If we're early, it's fine. If we're late, it's also not a problem. So feel free, after each presentation, to interact.

Kim, or who is running the presentation, can you go to the next presentation? Does this clicker work? Excellent. There you go.

JUHANI JUSELIUS: Yeah, it works. So thank you for the invitation here and welcome to Helsinki, Finland. I hope you have enjoyed your stay so far.

My name is Juhani Juselius. If there is someone who doesn't know me yet, well, I guess there are not so many of you here because this is my 28th ICANN meeting. So I have been sitting 27 times in these meetings somewhere else than in Helsinki so I'm really proud to see you here this time. And actually my wife was quite happy as well because this was the first midsummer weekend in 10 years that I was staying at my summer cottage for the whole time. So I'm usually traveling somewhere during this midsummer weekend and it's not that nice for the family, but this time I was here.

But yes, so I have some slides about .fi. There are not so many country codes that are run operated by the government, but we are one of them. So .fi it's run by Finnish Communications Regulatory Authority and we have been doing so since 1997/2002. And I have two different years here because we took over the operations in two steps.

In 1997 or actually in the mid 90s there were some [server] disputes between domain name holders and those who wanted to have them, and in those days it was a company for the profit that ran .fi and it was quite obvious that there was kind of a mismatch between those commercially interest and neutrality demands. So it was our Minister that decided that we should take over those administrative role of granting or selling domain names. And so we did it and then after five years in 2002 this technical operator of .fi, KPNQwest Finland, went into the bankruptcy, so we better have to take over those technical operations as well. So since 2002 we have been a real registry. And actually I stepped in in 2004 so I've been responsible for .fi for 12 years now or so.

We have about 390,000 domains at the moment and with quite nice annual growth at least for a [mature] ccTLD, so it's somewhere between 5% and 6%. Our current price for one year registration is $12 Euros. We are not for the profit, of course, and we try harder to get that price down.

We have about 12 people who are full time equivalent working for .fi domain names. We have seven full time people and the rest of the people there are [quite on] shared resources – I don't like the word resources – but I share people with other functions within our agency.

And at the moment the .fi registrant must be a Finnish company or Finnish association, public authority here in Finland, or a Finnish private person. So we have very solid link to Finland at the moment. And we're selling domain names directly to registrants and through registrars. And at the moment we have about 3,000 registrars.

We also run dispute resolution service for .fi, so we handle those possible disputes. And actually it's not possible but we have about 100 – 200 cases annually.

This is our technical platform. As we are quite small team so we have tried to automate every possible operation we do. So we have this registration system which contains all those business logic or user cases we have. The system it's actually tailor built for us as a turnkey solution and we are currently – actually, we got the first version in 2003 and now we are just about to implement the fourth version of it.

This registration system it has a web interface for registrants and registrars. You can find it at the domain .fi address if you want to have a look at it. We have a web service interface. It's EPP kind of interface for registrars. We also do have an open data interface for those who are interested in about the data we have.

The system also serves as a master database for domain related information, domains, and WHOIS information, and so on. We have this traditional WHOIS interface as well, and the system it serves our own administrative tools at the same time. So it's very important system for us. We are hosting it by ourselves, and every 30 minutes we transfer the whole zone to our name servers, and our primary name server is fully redundant and it's hosted by CSC which is a Finnish company. I think it's called IT Center for Science in English.

We have bundled some other services together with this primary server. So we are, for example, doing some checks that this zone we got from our registration system that it's complete and it's unchanged after this transfer. And we also do this .fi root signing on the same system and it's in two locations.

We have secondaries, and we mainly buy them as a service and we have actually four Anycast Clouds, one we buy from NetNote, and we also buy a Unicast service from NetNote and actually that name server is the oldest one we have. So it was in 2005 when we just wanted to have a name server with IPv6 connectivity and only NetNote was able to provide it for us at that moment. But we have kept it in our role so it's serving well.

We are buying an Anycast service from [DNIC.d], a third Anycast service is from Packet Clearing House, and a fourth Anycast service is from Community DNS. We are also having two Unicast name servers that we are administrating by ourselves. One is hosted by ourselves in our premises. It's fully redundant as well and it's based on not that typical name server software, but please do not ask the name from me. I'm not the technical brains here.

And our brand new name server is Unicast as well. It's fully redundant, but it's based on a Finnish [inaudible] in physics. They have two nodes and we have server in both of them.

So we are fully compatible with IPv6. We have implemented DNSSEC fully so I think we did it in 2012, but for some reason, which I don't know, we have only 329 signed domains. So if you have good advices what we should do just to boost it in Finland so just let me know.

We do support IDNs as well. So we launched three national characters in 2005 and a year later we added some characters used in Lappish languages. So we're supporting all those official languages in Finland but for some reason it took about 10 years when we got the first registration that used those Lappish characters.

I checked our statistics. Last month 22% of inquiries came from IPv6 network to our name servers.

Okay, I think that was all about our checking out platform and as I have this chance to be here, I want to introduce you to our big change that we are going to have September 5th because we have been working for this change for the past I would say nine years almost, at least eight. So it's a big change because at some moment we realized that we have to change, the industry around us is changing and we can't stay in this comfort zone we used to live and we have to do changes as well.

On September 5th we are going to implement registry/registrar model so we are going to stop serving direct customers. There are not so many of them any more left, so we had about 20% of domains were registered directly from us and we have been touch with all those direct registrants and I think the number is quite low at the moment. And the other big change in two months' time is that we're going to stop asking for local presence. So after September 5th, anyone anywhere can have .fi domain name without any restrictions.

And we are also giving up our minimum age limit that we have for private persons at the moment, so I think we are the only not for the children domain name ccTLD at the moment. But that's gone as well. And domain parking will be accepted so we are not going to ask for name servers anymore.

At the moment it's illegal to buy a domain name in purpose to resell it further, but that will be over as well so it will be legal. We have a certain list of those names that cannot be registered at the moment. For example, other TLDs are on that list so for example info.fi is on that list. It cannot be registered at the moment.

On that list there are also some offensive words and words that might lead you to commit a crime or that kind of words, and also words that represents the form of the company or corporate are on that list. And we are going to release that list for registrations in this change as well.

I mentioned to you that we're having web service interface at the moment, but we are going to replace it with standard EPP and as standard as possible. We have worked hard to keep it standard, but of course there might be some variations. But at least EPP will be available.

We are getting some new rights to improve IT security and as we're moving to a registry/registrar model, so we are also setting some new IT security-related requirements to registrants but anyway, this accreditation process to be a registrar it will be very easy.

Yes, so actually I think this was my presentation. Do you have any questions?

EBERHARD LISSE: Thank you very much. There is no question. There is one question from Jay Daley. He can come to the nearest microphone. Please.

JUHANI JUSELIUS: Yeah, I know that New Zealand is located quite far away from here, but you can step closer.

JAY DALEY: Why is everybody irritating New Zealand? This morning Geoff Huston was giving him also a hard time.

JUHANI JUSELIUS: It's a lovely country.

JAY DALEY: Thank you. I believe it's jealousy in Geoff's case.

Can you tell us more about your open data interface, please?

JUHANI JUSELIUS: Sorry, which one?

JAY DALEY: Your open data interface.

JUHANI JUSELIUS: Yes, so a few years ago we just decided that way to hide this basically WHOIS information – it's public anyway if you know what to look for, so if you just type the domain name you'll get the answer this WHOIS information – so a way to keep it secret. So actually nowadays we just have this open database and anyone can access it to have all our WHOIS information, domain names, and holder information except if it's a private person and they have opted out to give their contact information.

JAY DALEY: Okay, thank you.

JUHANI JUSELIUS: It might boost some new businesses but while I'm not sure if there is any but at least we give this chance to people just to have new opportunities.

JAY DALEY: Hello again. Welcome to Finland.

JUHANI JUSELIUS: Thank you.

JAY DALEY: On the subject of releasing previously restricted domain names, whatever they are. Whether they are ccTLDs, gTLDs, or strange words, do you have any allocation mechanisms in sight or any resource problem mitigation strategy?

JUHANI JUSELIUS: Actually we have just one method and it’s first come first served. So we’re going to open those registrations up actually September 7th at 10:00 a.m. Finnish time. And well we know it’s a challenge to our resources, I know there will be a competition but we have done our best just to provide enough capacity for this kind of land rush or whatever it is.

EBERHARD LISSE: Should it not be more like 4:00 to coincide with the sunrise?

UNIDENTIFIED MALE: [inaudible] the $12 Euro that you mentioned, is that the price you charge to customers or what you charge to registrars?

JUHANI JUSELIUS: It’s price to the registrars, so registrars can have whatever price they want, but it’s the same price for direct customers and registrars at the moment.

EBERHARD LISSE: Any further questions? Please feel free, we are not strapped for time. There is one more.

NARENDRA NATH: I want more details about the new rights to –

EBERHARD LISSE: Can you please identify yourself for the remote participants?

NARENDRA NATH: I'm Narendra Nath. I'm from the Department of Telecom from India. So the new rights you were talking to improve security. [I wonder] if you could be more specific.

JUHANI JUSELIUS: Actually, that's written in the law and at the moment if police or a judge ask us to do something, well we probably will do it but it should be a legal request. But in the future, actually we as an agency, suspect that some domain names used for IT related crimes we can with our own decision to monitor the traffic, we can change DNS settings, and we can even suspend the domain name. But of course, all our decisions can be appealed through the court but we can take those actions if they are needed as a quick response to that threat. It might pose –

NARENDRA NATH: So you’re getting those new rights through some legislation or something? Or how you getting those rights?

JUHANI JUSELIUS: Yes, by the legislation.

NARENDRA NATH: There’s some legislation that’s planned. Okay.

[MAX FREY]: [Max Frey], Global Village. Have you considered doing a discount for DNSSEC enabled domains? I understand it has been part of the success of .se.

JUHANI JUSELIUS: Yes, so we would like to have some discount for those prices that those who are using DNSSEC, but actually our law doesn’t allow it. We are not for profit and we should price those domains according to those products and costs, and actually production cost for a DNSSEC signed domain is higher than average domain so we should have a higher price.

[MAX FREY]: That’s why the registrars don’t do it.

EBERHARD LISSE: And the last question now.

ROBERT [MARTINSEN]: This is Robert [Martinsen] from PCH. You will be allowing domaining which I'm sure some people are very happy to hear. I am not particularly. I'm wondering why, is it that you don't care or is it that you think it should be handled somewhere else or if I take somebody else's domain name, what happens?

JUHANI JUSELIUS: Well, I think that why we exist is that we should provide as many domain names as possible in real use. So we try to sell or to grant domain names for those who are really using them. And I know that domainers are not doing that but we are doing marketing [inaudible] force. We are trying to find user for .fi domain names and if some domainer believes that they are doing it better than we do, so please do it. So if they can find real customers for domain names for real users, I would be just happy.

EBERHARD LISSE: Okay, thank you very much.

JUHANI JUSELIUS: Okay, thank you.

EBERHARD LISSE: Next speaker will be Mats Dufberg from IIS. Yes, yes. We need to present from this box because we have a remote audience and they can see it from there. Have there been any questions from the remote audience? Okay.

MATS DUFBERG: Hello. I'm Mats Dufberg from IIS. IIS is the registry of .se. This presentation has been made by me and my colleague at AFNIC. AFNIC is the registry of .fr. Sandoche is somewhere here in the room, too.

So, yes, Zonemaster. If you haven't heard of it, it's a DNS delegation verification tool. The purpose of it is to report errors in delegation or maybe report that there are no errors. It can both test a delegated domain but also something that you want to try before. So if you set up your name servers, you want to make sure that this will work if I change the delegation, or this will work if I create a new domain with this configuration.

And the Zonemaster test we don't invent what we should test. We look at the DNS standards and the best practices of DNS. And Zonemaster, it comes complete with a web GUI and CLI.

This project it started October three years ago and it's effort between IIS and AFNIC. And the reason why got together was that IIS had an old tool, DNSCheck, which is still running and AFNIC had ZoneCheck which I don't think is running anymore. And both of them were old so we needed something more modern to replace the two tools. And AFNIC and IIS decided to do it together. And Zonemaster has been available since February last year.

It's an open source program mostly written in Perl. There is some Java Script too in the GUI and the documentation is open available, so Creative Common license on the documentation and it's BSD license on the program for those interested in licensing. So it's open. You can do what you want. You can modify it or use it as it is.

For each time you test a domain, of course, multiple tests are drawn. And for each subtest, there is a fail, a warning, a note, or okay. So you can see exactly where it fails or if it doesn't.

Right now the output is in English, in French, or in Swedish. But it's possible to add more translations, and that is of course welcome. We would be happy to integrate translations into the program from other languages.

The program is based on open requirements and open specifications. A Centr workshop was created to create common requirements of a DNS delegation and they based that on the input from the Zonemaster work. And the result of that is a

Internet draft that is now available at [ITF] and I have a link to that in the end of the presentation.

Zonemaster is based on documented test specifications that are also available. And there's a link in the end where you can see it. So what Zonemaster does is not hidden in the code but it's documented. That's very important.

There are 60 test specifications that are split into nine categories and you see them listed here. And if there is an error or warning, then you can see what specification that has created that. So it's easy to go back and understand what the problem is.

The program architecture is as detected. It consists of four building blocks separated with clear APIs. So in the bottom, we have the engine which is a Perl library and that is the part that performs all the tests. The CLI is an interface for command line usage, and then we have a backend which handles queries from the frontend – and I will say more about that – and then the frontend which is the web GUI. So if you log in or if you use Zonemaster you could either log in and use the CLI or you can go to the web and use that. And the web is of course meant for public usage.

This tool is designed to be available for any users of a operator, a registry, or anything that want to provide their customers or users, the community, a way of testing their domain.

The engine is a Perl module as I said, and it's found on CPAN. Translation to different languages is done there and through the CLI and backend you can select what language to use via the API. And the structure of the engine is clear so it's easy to add new tests if someone wants that, or extend it by us or someone else. And if you want to use the engine as your own library into your programs and test something, then it's very possible to do that. So you don't need all parts if you don't want. You can use the engine only.

The backend handles the queries that come in from the web user through the frontend. So that's a daemon running and it has a database that stores all the queries that comes in. And storing it, the purpose of that, is to be able to retrieve it later. So when you go through the GUI, you can look at the old queries for the same domain name. You can see what happened last time when I tested. Did I have that error then, too? Or last I know that there were some errors, are they fixed now?

The API for the backend is JSON over HTTP. It's a well known type of API. And of course you don't need to use the frontend to use the backend. You can write your own application if you wish.

The frontend is meant for normal users with a web GUI. It's also a service of course, and you can run the frontend on a different server from the backend so you can put the backend more hidden if you want, or more secured and then have the frontend more exposed and just have that communication open.

But of course the backend must of course be connected to the Internet so it can perform all the DNS queries. If you want to test this, you can look at zonemaster.net where we have the GUI running. Again, you're welcome to download the GUI or the whole set and then modify the GUI for your own look and feel.

Zonemaster is configurable. You can restrict what tests to run. You can also change what result an error or warning. You might have a looser or stricter regulation and you can do that. And we're now adding a filter feature on RIPE's request to have a more fine grain control of policy per name server. They want to exclude some name servers from the testing because of the way they use it.

In the future, today CIRA has a set up for .ca. They have taken the Zonemaster and they have modified the web GUI for their own look and feel so it looks different.

IIS, we are working on an installation for .se and .nu or for our customers. We will do a change. We have more [inaudible] the web interface for our look and feel. RIPE NCC is working on an installation and we know that that there are more installations to come.

IIS and AFNIC are dedicated to support Zonemaster. They will be needed for our customers, so for us they are very important. But other users and contributors are welcome. The more users we have, the better we find any issues or bugs that might exist in the code. Also, as I said earlier, if translations, if someone want to have it translated some other language we will of course help with the format of such translations and that include the new translations into the code so it's there with new releases.

Here are the references. So if you get my presentation or our presentation, you can find the link to zonemaster.net. Everything is published on GitHub. The Internet draft that you can read the requirements is there. Well, right now it's -00, but it might be a later draft when you look. So remove -00. And specifically the test specifications, they are in the same GitHub repository but there's a direct link to that if you want to read the test specifications.

EBERHARD LISSE: Any questions?

ROBERT [FREY]: Yes. It looks very interesting. Can you explain a little bit about how you incorporate this into your own processes inside your organization? Because I think many here might not see where this would fit in. Some might, some might not. And I think that was not in the slides.

MATS DUFBERG: Currently we run DNSCheck in the same way as we will run Zonemaster very soon. And firstly, we link to that so that our customers – the .se and .nu registrants – can check their own domain. And secondly, we run every night a check on all domains that have changed delegation on. If there are errors on any one, we send an e-mail to the registrar and send the information to them with a link back to the DNSCheck. We will do the same with Zonemaster which go back to a performed test in the database.

This is the actual test that was performed yesterday. This is not how it is today. This is the errors that you saw when we performed the test. So that's the purpose of the database, that you can point back. These are the errors that we saw because they could be errors that are only there at night or something. So we see this as a service to our customers and a way to keep the quality of the .se and .nu domains.

UNIDENTIFIED MALE: Let me abuse the prerogative of the Chair a little bit. We in .na require from our registrants that they maintain working name service. It's relatively difficult to standardize this or to sort of objectivize this. Then we have an object measuring tool, if we then put in our contract you must comply with Zonemaster less than notes. We will run this once in a while and then you send, "Look, we found this and this and this. Please fix this." We find that real name server errors, all our registrars are very happy to be informed that they can fix it, especially when they've been told what to do. Our experiences are not as good with the WHOIS accuracy and all related issues, but if you have a system where you can tell them in an objective manner, "This is not working. This is not working. This is not working. Your agreement requires that you must do this." There I'm going to fix this.

UNIDENTIFIED MALE: [inaudible] .tk. I think you could actually [get to] my point [Abrahad]. I'm just going to ask what your reaction was from your registrars. When we proposed to start doing that, our registrars would "Go away. Don't fiddle. What a mess. I decided how I want my servers. You don't tell me how to do it." They sound like they're not interested and even if we want to send an e-mail they'll just throw it away. So what's been the reaction for you?

MATS DUFBERG: IIS does not require the domain names to be up. So you're allowed to have broken domains. But the registrars welcome the information and I guess that some registrars just see them and throw them away. But there is no negative reaction or that, on the contrary.

And I also wanted to comment, if you're running – some of you might be running DNSCheck today, and that software is deprecated. So you shouldn't count on any updates in the future of DNSCheck.

EBERHARD LISSE: We have terminated one registrar for cause over the 10 years that we have formal [inaudible] that we have. It was persistent problems, persistent failing name servers, persistent incorrect WHOIS data, persistent lack of response, and persistent incompetence. Eventually we sent them a notice asking them why we shouldn't terminate the agreement. They couldn't give us other than some insults which I usually appreciate very much. And then we gave them notice and they started to squeal, rave, and rant, and nothing happened. And in the end, the domains moved over to other registrars. And even though we haven't got it formally [inaudible], if you tell them, "Look, you have an issue," they always fix it. Name servers, they always fix it because if you have an issue on your name servers their own clients will be affected.

WHOIS is a different thing. That's just a cost center, that's not a profit center. So the interest in fixing the WHOIS is purely related to the amount of pressure you can apply. But operational things that you can, "Look, you have an issue. This is what's going to…" They say, "Can you run this tool again?" We have seen this. We've used the French tool in the past. They once in a while said, "Can you run the tool again?" I tell them, "Well you can run it yourself," but they're so happy if I run it for them.

So generally speaking, you will never get the really lazy or criminal registrants who just don't want to do it. But many registrars want to provide a service to their clients, especially if they're not too big. And they appreciate if you can do that and they start something incorporating these things especially since there is an API that they incorporate into their old tool that they [won't] register things at registration and break the rules. It's just a problem what happens afterwards because it's a DNS is file and forget. You put your thing on you never do anything about it. That you run name servers from 20 years of age because it's working. Why fix it?

Anybody else? Thank you very much. I am busy to try to install it from CPAN on my Mac laptop, but it runs into a missing program which is a known issue non-related to Zonemaster. I'm tracing the bug.

Okay, the next individual is Patrick Wallström equally from IIS. He's going to talk about the change in their policy about zone transfers or zone files.

Use the clicker.

PATRICK WALLSTRÖM: Hello. I'm going to talk about how we handle our .se and .nu zone files . Oh, the clicker.

First of all, the big news that we are making the zone content for the .se and .nu domains completely open and available and [that] they have been open for a month or so. And you can fetch them through our website or through AXFR.

And some back ground for this. In Sweden we have what we call a DNS Reference Group which is meetings twice a year arranged by IIS where we go there all day, DNS expertise in Sweden. So a year ago we had the discussion on if we should change the .se zone from using NSEC to NSEC3 because our management tends to talk about that a lot.

The problem is that we don't really know why we consider the zone file to be secret in the first place. It's also kind of hard to make the transition from NSEC to NSEC3. We had to upgrade the DNSSEC algorithms and such. So our conclusion from this reference group meeting was, rather than make a decision on moving from NSEC to NSEC3, is that we should consider the possibility of releasing the zone file publicly. And we took the DNS Reference Group recommendation quite seriously, so I took upon me to write a document.

This all turned out to be a very long document. First, some background why there are limitations on fetching the zone file right now. And a decision made like 10 or 15 years ago said that we should not release the zone file because of issues with spam and phishing and such.

Then also a very long description of what a zone file actually is, the content of it, a complete overview of what is an IP address for example, what is the different record types in the zone, what is a TTL, what does all these values mean – because management are not DNS tech people – and also an example how zone files are being used. You have to explain what the [resolver] is and how [to] use it to fetch parts of the zone file content.

There is also high interest in the zone file content. At registry we receive queries for fetching the content a couple of times per month or so for different reasons. Most of it are research.

We're also very behind the open data issues. We want to share as a non-profit registry, we'd like to share the stuff that we're doing. I also sent out a Centr survey, asked how other registries in the ccTLD community in Europe sees their zone files and why they are not publishing the content. Most of them handle the zone file as secret and also when they are releasing the zone file to individuals that they have a quite strict agreement on what they can use the zone file for.

And also the obvious thing here, the zone file is available anyway. People running large resolvers are extremely large portions of your zone file. You can do NSEC and NSEC3 [walking]. And the problem in if you're deploying NSEC3, instead of NSEC [walking] you would have dictionary text which increase your load on your name servers.

What could possibly be secret in a zone file? For our registrars this means that when they are registering domain names used by trademarks that have not yet been registered or campaigns that are not being launched yet, they should recommend to their registrants that they are not delegating domain names so that it's published in the zone file.

There are also benefits of freely distributing the zone file content. We don't have to care about protecting it in any way at all. We also did a full risk analysis of what could possibly happen when you release the zone file. It was also a description of how other people distributing the zone files. All the gTLDs have their zone files available in some way or another and all the new gTLDs are making their zone files available through the ICANN centralized zone distribution system where you can log in and request copies of the new gTLD zone files and then the backend provider or someone has to approve that my reason is good enough. But I think I have 800 new gTLD zone files on my drive. And also how the zone files should be distributed to limit the problem for our other systems. I mean the zone file is quite big. And also recommend the decision. In this case, the recommendation was to release the zone file.

So what the management decided then was that we should probably have more eyes on this. So we then decided to perform an external legal analysis of the problem. We [got] a lawyer agency to do this for us. And also they assessed/performed an external technical analysis and the NLNet Labs did that for us as well.

Wistrand did a legal analysis. We have two laws in Sweden. We have a law Top Level Domains and we also have the Personal Data Act that mentions things like IP addresses can be personal identifiers in some cases. And this report said that there is nothing in the legal text that prohibits the distribution of the zone file. And if it were, we would have other legal problems with doing our core business anyway. So that was good news.

And then NLNet Labs did an inventory of the technical consequences. I mean we shouldn't open [exit] for all our name servers, for example. They recommended also that we use [XPS] and [XF4] for distributing zone file. And since [we all did] do some really good protection of our personal contacts in the WHOIS database, this was not a problem either. You cannot use a zone file and have that as a bridge into our WHOIS systems in any way because the web WHOIS is protected by Captcha and our WHOIS on Protocol 43 they cannot see any personal information at all.

After updating all my documents again, I sent it to management, they said, "Okay, let's do this but we have to ask the Board to verify the decision." So the document went up to the Board as well. So the decision from the IIS Board was made in March. The zone is to be released without restrictions under a Creative Commons Licens. And this was for the .se zone. And then we also asked the IUSN Foundation if it was okay to also release the .nu zone file. So that was okay.

So after these decision, we had to do a lot of external communications. We have the PTS which is the regulator overseeing the gTLD law in Sweden and just informed our plans to release the zone file. We also had a lot of discussions with our registrars and also informed them of the problems with delegating zones in the zone file.

We also informed the Swedish Data Protection authority. I was again informed the DNS Reference Group they had early access to the web page where they can fetch the zone file and just could see that it worked. We informed the public on our website. We also did a press release. I did a presentation last month at the Center Jamboree Meeting. Now I'm here at the ICANN ccNSO meeting as well.

The problem with stating copyright on the zone file is that there is no place in the zone file where it can have this information available so I wanted to add this to the .se zone file. It is not yet done, but it would be nice if there was some sort of recommendation somehow to do this. So I wrote a draft on how you can put license and copyright information directly in the zone file.

On my personal domain here you can see an example on how I published this information there. And this also the ITF draft that I wrote on it. I'm not sure that we're going to do this but it's an idea on how you can actually publish this information in the zone file. Because when you access the zone file over [X4] you never get to see any license or copyright information at all other than looking at the zone file itself.

End users, we informed them that you should not delegate your domains if you don't want to make them public because not publishing them in DNS is the only way people can notice that they are available. So our registrars are aware of this and people [who] have been working with trademarks and campaigns for a very long time are also aware of the issues.

And this is the website you see. You can go there and fetch the zone file just by clicking on the link for .se and .nu. And you get the [GC] compressed zone file. And it's also available with [X4]. The only thing here with this website is that we ask people not to abuse it, otherwise we can just filter out their web IP addresses if they abuse the website.

A common question we get is that, "Can I configure my name servers with .se zone file?" Of course you can, but we will never send out notifiers from this system to your name server. So you have to make it up to date. We publish the zone file every two hours here, so you can fetch the zone file every two hours to have a fresh copy.

And the reactions on the release is that we have no negative comments at all. Only positive ones, such as, "Now I can skip my

NSEC-walker." And, "Well done. You're doing the right thing." And, "Can I configure my name server?" That's all.

And of course this one from – there's a Tumblr called DNS Reactions – so this is an automated gift for you. Someone pushed a button all the time.

So some recommendations if you wanted to do it. Have a well protected WHOIS service because that's the thing that you really want to protect anyway, the personal details of all the domain name holders. We have contact information by the Captcha as I said. The zone data machine that we have is completely separate piece of infrastructure that we have in order to if people are going to fetch the zone file, like hundreds per second or so, they will not destroy our other infrastructure.

So my question to you, what do you think? Will you release your zone file content?

EBERHARD LISSE: Thank you very much. The zone file is clearly intellectual property. What you do with it, whether you release it under whatever license you do or not, is your decision in the end. But the point is you then allow, as you noticed, them to walk your WHOIS, systematically query for every domain name who is owning it, which is one of the issues. I don't want to have to have somebody access my intellectual property like this on principle and for some other reasons, but if you want to do it and if you have got mechanisms in place to corrupt the abuse, fine with me. Patrick had a question.

PATRICK JONES: We have a remote participant question. They are asking, “How do you simulate the increase of spam abuse that others expected?” And you said you don’t.

PATRICK WALLSTRÖM: As I said earlier, the zone file is available anyway, so people that are doing abuse such as spam, they already have very large collections of mail addresses and of course they [access] anyway because they know how to [inaudible] walking. I mean you saw the presentation these people who were here this morning, [Mathias] was it? He collected the zone files from all over the place doing NSEC and NSEC3 walking. The information is out there if you want to abuse it or not. It’s not up to us.

EBERHARD LISSE: I actually disagree. If you don’t want somebody to abuse it you have to do whatever you can to not just give in. Dimitri?

DIMITRI LORENKO: Okay. It's Dimitri [inaudible] domain. We never ever released the zone file. I guess now it's another argument. I repeatedly get requests to provide zone file to various companies. Sometimes they say we are research, sometime they say we marketers, sometimes they say nothing. But I just have a very quick question. You talk about zone file intellectual property. You mean zone file basically list your names, list your name servers, list your [inaudible] records. That's it. No other personal data is that except for the name. We may argue whether domain name is personal data, whether name server is a personal data, but do you plan to release any information for example list of, I don't know, registrants?

PATRICK WALLSTRÖM: No.

DIMITRI LORENKO: Not at all. And it's not at all. Because in case of Ukrainian registration for example, same thing. I can probably do what you did, but I would definitely never do the list of people to domain names or the company name to the domain name.

PATRICK WALLSTRÖM: We never publish any contact information. Only for companies.

DIMITRI LORENKO: And you don’t include any of the NSEC records I suppose? So basically you can say you publish basically [unsigned] zone.

PATRICK WALLSTRÖM: No, we published a fully [signed] zone.

DIMITRI LORENKO: Oh, you publish the NSEC records [inaudible]? Okay, great.

PATRICK WALLSTRÖM: Yes, everything that is [signed] zone.

DIMITRI LORENKO: So as much as I can get [inaudible].

PATRICK WALLSTRÖM: You can verify the integrity of the zone file using DNS validations.

DIMITRI LORENKO: Thanks. I think it’s a good effort. I would probably encourage now other people to consider doing that, if only maybe for the limited amount of [source].

EBERHARD LISSE: Dimitri, I get the same request for zone file access as you get because they send it to all TLD managers who don't have filters on it. I usually ask them how much are you going to pay and if they answer that with any figure, I say, "Okay, per domain per month." And then they never answer.

DIMITRI LORENKO: Thanks. I'll use that technique as well.

EBERHARD LISSE: Next.

SIMON MCCALLA: Hi there. Simon McCalla, Nominets. We recently released our zone file a couple of months back under license and one of the things we've noticed is we've seen a lot more attempt at enumerating via the WHOIS and through our various tools, and we've also seen quite a lot of unusual DNS traffic spikes as a result since. Now we haven't been able to tie it back directly to the zone file, but I just wondered if you'd seen any other patterns like that since you've released yours?

PATRICK WALLSTRÖM: No, we haven't actually.

ADAMS: [inaudible] Adams, ICANN. Simon, here's a tip. You can actually embed this funny string in your zone file that you release and see if people query for that.

So the root zone is available, all the gTLD zones are available. For those the ccTLDs for which I, if I wanted to, do not have a zone file. What I do is I'll take the com zone, strip off com, at, let's say .na and enumerate them if I wanted to do that.

So I understand your personal reasons or business reasons to keep that private. The fact is it's not private. You may want it as private as you like, but folks have ways around this and they have had that for many years. So I'm not saying you should release it, absolutely not. But I'm saying there are ways for other people to get that zone file. Plus, you might actually reduce some of the loads on your systems. But just what Simon said, the load actually increases so there goes my argument.

EBERHARD LISSE: I do not agree that if somebody commits a crime, you should lay back and enjoy the ride. You should use every means at your disposal to work against that. It is a crime to abuse or to use intellectual property that doesn't belong to you for purposes that the owner doesn't want you to do. It's clear. Under every data directive country in the European Union, even currently still in the UK.

ADAMS: Well, my law degree says – I'm sure you're not a lawyer, but I don't think you [inaudible].

EBERHARD LISSE: I'm a gynecologist but what's the difference?

UNIDENTIFIED MALE: You're both… Oh never mind, I'm not going to go there.

UNIDENTIFIED MALE: [Inaudible] .dk. I'm just going to give a Danish perspective and I do agree, just what Roy said. The number I'm seeing, if you take .com zone for .dk, you get about 80% of the zone. That's [more tricks]. You can do more, so it's not really that secret.

On the other hand, we are under the obligation to publish contact information unless there's other regulation that forbids it. And we actually have to validate all Danish uses and companies against the registries of the government. So we actually know if the right person, including name and address unless they have name and address protection. So I'm not sure what we want to do because –

UNIDENTIFIED MALE: [Inaudible] is available anyways.

UNIDENTIFIED MALE: 80%. And the other comment is we do know some of our registrars use it to [walk] WHOIS database every week. So we do see that kind of information gathering.

NIGEL ROBERTS: I shan't prolong the agony too much. I've noticed that lots of the questions and comments are –

EBERHARD LISSE: Nigel Roberts, can you please identify yourself for the record?

NIGEL ROBERTS: My name is Nigel Roberts, .dj. I shan't prolong the agony. I notice that most of the comments that you heard here are not really technical in nature. I mean, it's fairly simple to transfer zone file with AXFR. Interesting comments about NSEC 3 of course. But as I was getting up here – this is not my laptop; it was thrust in my hand by one of my colleagues of a long period of time who reminded me – he probably knew what I was going to say – that in 2002, ICANN attempted to mandate the transfer of the ccTLDs, as [inaudible] puts it, intellectual property to them.

And I think you have to remember and understand the DNS wars more often to the zone file wars. ICANN said, “We own your TLD; give me all your zone so we can give it to somebody else if we want to.” There is a serious mistrust that still persists to this day, residual mistrust perhaps, but there’s a serious mistrust to this day that if your zone file is available to people that you haven’t given it to, they will use it to take your ccTLD away from you if you fall out with them. And some ccTLDs have fallen out with ICANN over the years.

Now the ICANN that we’ve got today is a completely different animal to the out of control blackmailing. That’s a literal word. ICANN refused to do IANA updates unless data that would allow them to do a redelegation was provided. So we’re not in that position anymore. But one day again, we could be. As we found in the UK last week, never say never.

EBERHARD LISSE: Okay, Shane Kerr and then I think I want to close the list after Patricio.

SHANE KERR: Shane Kerr, Beijing Internet Institute. Have you had anyone attempt to download the zone and do loopback version of SC or anything like that that you know of?

PATRICK WALLSTRÖM: We don’t know that.

SHANE KERR: Okay. Interesting. There are ways. Okay.

UNIDENTIFIED MALE: Okay. Sorry, I’d just like to add a comment. I don’t think anybody owning a copy of the file that any ccTLD has would be able to hijack the domain from you because it’s all built on mutual trust. The discussions on registrars, when they actually threatened us to [inaudible] our domain without – that was back over ten years ago – I don’t think it would work because you have to delegate the domains. So basically, it’s just like let’s say you’re in China and you’re working the Internet. I know it’s a hypothesis, right? At some point, you can still have a copy of the zone, or like talk about the project that [Shane] was participating in, all the tests of the copy of the root zone file which is just like the original but different. So I really don’t think that’s a threat. But I also don’t think there is any utility in it. I think it’s only for search.

UNIDENTIFIED MALE: You are most welcome to release your zone file.

UNIDENTIFIED MALE: Oh, sure. I just don't think I wouldn't call it crime and I wouldn't talk it a threat to the owner. And yes, it's their choice. And I readily agree with you that – sorry, you, I guess too – that ICANN or any other body should never force on any operator truly. That I would really side with because, hey, if as a manager, you are responsible [inaudible] is, you know. It's like your task too.

EBERHARD LISSE: I don't want this to [inaudible] in a debate about intellectual property. The point is if I ask once to do this under the open common license, it's perfect in order to do so. If I don't want to, it's equally in order to do so. Patricio.

PATRICIO POBLETE: Patricio Poblete from NIC Chile. We don't release our zone file and that dates back to the days of the zone file wars as Nigel was remembering. From time to time, we've been getting requests recently, mainly from mass marketers and they found out that they could use the Chilean transparency laws.

As NIC Chile is operated by the University of Chile, a public university, and it was recently decided that those laws did apply to our university. So we got one such request and that law also says that in case there are people involved, we should notify them that their data is being requested. So we did that. We fired off something like 240,000 e-mails and we got back in a couple of days over 50,000 e-mails of people that were opposing that their research and data being released. And that was about the end of that. So that's what happened.

EBERHARD LISSE: Okay. Thank you much. I think this was an interesting topic because it also generates a little bit of discussion and it generates a little bit of thought in everybody in our present TLD managers or ccTLD managers whether or not to do it, how it can affect them and what intended or unintended consequences it has. But one thing which I didn't realize is you can only do this if you have really good WHOIS security because they [work] my zone every day. I see requests every day. So since I have taken to diligently e-mail my provider to block those posts from which it is coming. Now that [inaudible] is blocked, it is much less.

All right, next is Jay Daley.

JAY DALEY: Hello, everyone again. I am the Chief Executive of the .nz registry. And I'm talking today about our registrar portal. I've talked a bit about this before, but we're in the midst of some new development about it and so I thought it would be just useful to give you an update about it.

This is a standalone web application – so many of you people are used to these – launched in January of 2015. It replaced a section of our main website with private registrar transaction data that was just tables of data for people to use. We've started phase two development and I forgot to collect any usage statistics so there's no data to show you about that.

This is a very different registrar portal, I think, from many others. This has the main objective of influencing registrars. We set out with some objectives. The first one of these was to reduce support costs. So on the financial side where we need to chase late payments, where we need to resend invoices to people, where we need to go through an invoice with somebody, we wanted to eliminate some of those problems. And on the technical side, where people have registration system errors and they want to understand those more, we wanted to reduce our costs there.

We also wanted to influence our .nz sales by helping registrars in a number of ways, helping them to find new ways of selling. So that's new markets, new customers and new sales to existing customers. We wanted them to improve their sales process as well so that they understand how the sales they've made one month can be better than the sales they've made another month even if their numbers are the same, to understand their customers better, to retain customers by timely intervention as well. That, of the sales bit, we've done some of that, not all of that, and I'll talk more through that.

We wanted to improve the quality of our top-level domain, again by influencing registrars, to improve the quality of the WHOIS data, to manage the DNS better, and to begin to build processes for acting quickly on phishing or compromised sites.

We also decided there are some things that it must not do. We did not want to create a class of registrars who could rely on our portal and not use EPP. We also have our own proprietary system called SRS which is XML-based over HTTPS. It's not similar to PP, but again, it's an automated system. It is our view that there should be a very clear minimum technical barrier for a registrar that they can properly use EPP and talk to an EPP server, so they can write automated systems.

We also wanted to avoid allowing one part of registrar to work in an uncoordinated way with another part. So we regularly find this when we visit registrars, that one team has access to their EPP interface – it can do things – and they act as a gatekeeper of company processes. And you see another team, and the other team asks you, "Can you give them back door access to their own data in the registry?" to do things that the other team won't let them do. And this is just something we just want to avoid entirely, you know. We don't want that.

We also didn't want this system to lock things away too much. We didn't want to end up putting things on here privately that we really should make public to everybody. We didn't want to provide a new way to attack a domain, which if there was any form of registration in there, that would be possible to do. And we didn't want to substantially increase our support costs. This is meant to reduce them. So there is no functionality in our portal at all that duplicates EPP.

Okay. Kim, will you? Do I need to in Adobe Connect? Hit the share button. All right.

Hold on. Sorry. Okay, I'm just going to let the browser. I'll let it share my screen and various things and then we're done. Sorry. Slight technical delay.

EBERHARD LISSE: Technical delays only happen in technical working groups.

JAY DALEY: I'm going to be showing you our demonstration portal which is not intended that well to be used on the other side of the world from where it's hosted so it's a little bit slow.

Okay. It's coming back and I will be set up and then showing you something. But we're already ahead, so that's fine.

Kim, can you bring my slides back. Thanks.

So I'm going to just talk while we [fill] about that, about some things that we're not doing yet. Well, actually, that the developers are doing that they are building currently for us. So I can't show you this screen yet.

We do a scan of all of our zones once a week. Currently, we use a customized version of a Zonecheck and we're in the process of switching over to using Zonemaster. From that, we have, I think it's approximately 80 errors or something – well, errors, warnings, and notices, the standard zone check ones plus some more – that we then collect for every domain name.

We're just in the process of developing a screen for our registrars that lets them see every domain and then lets them see which of those have any of these errors, any of these warnings or any of these notices. So effectively, there will be a line of 80 dots that could be colored in or not colored in there to be shown against each domain name. And they will be able to see what portion of the registry total they have of those and be able to look at those.

This is part of our attempt to drive people for better quality in the zone by showing them the errors that they're getting right down to the very domain name level for them to look at. There are no obligations on them. We're not forcing them to respond in any way or do anything. But we are expecting them to just look at it and we will track that and begin to learn and understand the impact that zone errors have on them there. Okay?

EBERHARD LISSE: But it looks promising.

JAY DALEY: Replace that. Yeah? All right. Okay, sorry. I was previously logged in but the screen's crashed and things so this, oh no, it's good.

Okay. So this is the initial admin interface where I choose a registrar. This is all fake data, entirely fake registrars. So this is the invoice screen which we show to registrars and there is the small psychological nudge of the little snail with the dollar icon on its back to show when they paid late. We paid to have that icon made and it's now creative common so you can use that icon as much as you wish. All of our things should generally be creative common.

So when a registrar ends up with three snails on one screen, they will then receive a fine from us and we will then just send a screen copy, a screenshot of the three snails along with the invoice for having to just deal with their processing problems. And they know it's coming so that just makes life a bit better.

There is the pink up here which is showing you our estimation of the month end billing that they will have. Now this is just our billing that we bill them. What I would love to know is for every domain name, what the retail price was that they sold it for because I could then show them value on this, not cost, and it would be great to get to understanding value about things.

As with every screen here, on throughout, we have the comparison of them with previous periods or them with the registry overall and they can download the data as necessary and they can move back month to month as well.

Okay, so the transactions. This will just take a little bit to load. We used to just publish a table of transactions, every transaction on the numbers for each registrar that only they could see. We have now broken this up into something that is more targeted with a message for them. So this is a [gamified] page. It is won and lost where won is your create and your transfers in and lost is your cancellations and your transfers out to another registrar. It's just attempting to present the same data to them in a way that drives their behavior really a bit better.

Because this is fake data, you can't see. There aren't trends in this. This is relatively random data. But on a real registrar, when we look at that, you can see quite a trend of the way they're going up or down or how they're attempting to recover from a bad position.

Growth is pretty straightforward. We show that against the registry total. There's nothing unusual there.

Then we have the renewal rate. And the renewal rate doesn't normally look like this. The lines are closer together in real data. We're going to make some effort to get some better real data here. Our renewal rate is generally about 83% in our registry which I think is very high. And registrars can vary, obviously, either side of that but don't generally vary very much either side of that. There's a very low spread there of registrars.

And for those that are below that registry renewal rate, they can see that immediately and begin to understand that they're not doing as well as other registrars and they should be doing better with that.

Then we provide them significant data on transfers in and out. We show them their transfers against every other registrar, net transfers and just the straight in and out here. In the live system, we have 80 registrars. Each person generally shows something like 40 or 50 of those registrars will have appeared because that's the level of term that takes place. So that's the very basic transaction data. That's nothing special. We're doing maybe a few more things on that, but not a lot.

Okay, so the interesting bit now comes in with quality. Every night, we run a series of fairly simple pattern matching scripts or routines against our WHOIS data, our registrant data, and we look for obvious mistakes or obvious problems. So this one here is the city as the name; this is quite common. We must have, I think, imagine, out of 660,000 domains, we probably have 1,500 of these. So this is where people deliberately give the same name both as their name and as the city name and possibly something else. It's an indicator that they are trying to create privacy through the back door.

Of our 660,000 domains, we have slightly less than 5,000 domains in total trigger these data quality errors. So it's a very small number and we have some false positives in there as well.

So we look for obfuscated private, it should be called here. There's one obfuscated NA that we've now merged together. But it is looking for any way that the registrant has attempted to create privacy which is not allowed in .nz. So they have put NAN or they have just put private or they have put XXX or something. A few weeks ago, we found somebody who had put in the address of 1 Fictitious Place against all of their domain names, so they were easy to find.

Then some people, the next one is salutation as first name. Some people just put “Mr.” or “Ms.” as their only name in the registrar so we look for those.

There are a number who have, apparently people have entirely numeric names. That may be the case but not in New Zealand. And entirely numeric cities and I’m not aware of any entirely numeric cities in the world. If anybody knows of any, please let me know. So we look for entirely numerics.

There are some people who only provide cities. I can click on one of these and it will show those. And so just the city, the same thing on both of the addresses there.

We have some marvelous wrong country/city combinations. Now we only know about cities in New Zealand. We only look for cities in New Zealand and we look for cities in New Zealand where the country is not New Zealand and in every case, the country is Afghanistan because it is the first one on the drop-down list and the people are too lazy. And this just annoys me. Okay, there is nothing really wrong with the data because we don’t need the country thing. It just really annoys me so we are trying to eliminate that one.

Then the final one is a known privacy service which we don’t have any of. Nobody attempts it because we know and we find out about it.

Then the next one on the quality section is our system errors. These are the mixture of EPP plus our internal errors. The data doesn't normally look like this. It, for a registrant, normally shows that a registrar has a problem. A big one is where they are attempting something with an invalid authorization code and you'll see that five times on one day, ten times the next day, five times. You'll see a pattern about their misuse. And that's very easy for people to see already. So that's useful. This is not one we look at. This is one where they can just work themselves and do it.

There are two more bits we are in the process of developing up here in the quality section. One I showed you earlier is the Zonecheck data or the zone data. And the other one is threat intelligence. This is where we consume three separate feeds of phishing sites or compromised sites and we then break that down, those domain names, by registrar and then put those up for the registrar to see so they can see those domain names, they can see the confidence level, and they can see where, which feed we got that from. We can publish that to them. And we've only started off with three feeds and that's still in development. It should be ready in the next couple of weeks.

Okay. And I'll talk more about those feeds and things straight after the demo.

Opportunities. Opportunities is where we provide lists to the registrars of domain names and registrants that they may want to sell something to. So currently, we have, well, for many years, you could only registrar a domain in .nz at the third level under .code.nz or .org.nz. You can now register directly under .nz, so these are, the first one here, second level is about people who have a .code.nz but don't own the corresponding .nz and so you can upsell to them to buy that.

We're looking to extend these opportunities in future to, for example, using traffic analysis, find domain names that are in the top ten of traffic and then put those as an opportunity to sell a ten year registration to because we offer registrations from one month up to ten years. So it won't make us more money – well, it will with interest – but it's just a better cash flow position, better if they can upsell those as well. And it's better for the registrar if they're in the top 10% of traffic that they recognize their domain name is important and do that.

The other thing, of course, is our drop list as well which is just the domains that are coming free and when they're coming free which is useful for the secondary market.

Then we just have a few other bits. The portfolio bit, we have a complicated conflict system for opening up the second level and that shows people things there. And the resources are marked in resources, communication packs and things that we are probably going to make public anyway.

Right. So can I switch back to the slides, Kim? Thank you.

So recap then. The transaction data, this is split into sections each with a specific aim and there is some gentle gamification. The opportunities are lists that they can use to upsell, cross-sell, retain people. There's no tracking. We don't push people to use these. It's up to them. But we do see these used regularly. And then the quality which is currently is bad WHOIS data and EPP or SRS errors, but we're also adding on zone errors and compromised and phishing sites onto those quite soon.

So finally, some bits on the technical info. This is Ruby on Rails web app. The charting library is mCharts which is marvelous, very nice library. It's a standalone post-[inaudible] database. So this has no connection to our production database. That's very deliberate. If this is compromised in any way, then it stops there.

The threat intelligence feeds that we use, that we're building in, are the Shadowserver compromise sites for .nz which are free if you walk to Shadowserver, OpenPhish Premium which they gave us for free as well because they like those things, and the APWG block list which comes through our paid membership which is expensive.

Okay. So we have a few more little plans to go. These are things we've agreed but we haven't fully implemented. We're going to do anonymous ranking for registrars. Again, nudging people. So a registrar will see you are the number one worst for quality and we hope that that will nudge them to do something more about it before we start having to publish that.

We are going to support the role splits within registrars with access controlled roles and we're going to add true fact authentication. I mean, everybody does these things.

We already have an active program using machine learning to provide an industry categorization for every domain name. It's our intention to give our registrars a breakdown of their portfolio of names by industry category and then compare that to the overall registry category. This goes back to the objective of helping them understand their customers better so they can see that data, pull that data in.

And there are some things we are probably going to do. We're still thinking through it. We are building a probabilistic model of cancellations so that we know from talking to registrars, they have a limited number of contact points which they can get away with, with a registrant. So we want to maximize the value of that by helping them understand just when to intervene with a domain name to prevent it being canceled. There are some good indicators of this, again, from traffic analysis. If the number of lookups of the MX record drops significantly, then that is a very good indicator that domain name will be canceled.

And so these are the things we look for to build that model to help then on the opportunities list that I mentioned earlier to provide a list of domains that are likely to cancel and you should intervene about.

We want to extend that model to domain name value. We're actively working on these, so these are more than just perhaps wishes. I want to show the historical lifetime value of a domain name. Now currently, I can show the historical lifetime cost of a domain name. I don't know the retail price so I can't show the value, but if I use my cost as a proxy, I can still show data for people.

And then I want to have a model for when a domain name is registered that can give me – again, it would be a probabilistic value of how long that domain name is likely to sit on the register form because a registrar may register 1,000 domain names in one week and 1,000 domain names in the next week and believe they've done exactly the same business. I would like to be able to look at that and say, "No, the quality of this week was much better than the quality of last week so your sales process must have improved over that time and helped drive the sales process" by measurement and value of those.

And then the next thing is we have an algorithm for the detecting portfolios. We want to refine that and start then putting that up there so that – this is customers with lots of domain names or more than one domain name – so that a registrar can understand which of their customers like buying domain names or need them because they're investors or because they're a large company that sells lots of different brands or something so they can understand who they can sell more to in that type of way.

So that's it. And there's some text that is reversed and upside down. I don't know how that happened.

EBERHARD LISSE: Are you sure that's not a little snake with a dollar sign?

JAY DALEY: It might be.

EBERHARD LISSE: Thank you very much. Any questions?

UNIDENTIFIED MALE: Which part of these tools that you have a sort of available to be taken out, outsourced so that somebody else or a different TLD can use them on their own thing?

JAY DALEY: All of our research tools, like the machine learning to categorize by industry code and traffic analysis tools and all those things, those all open entirely. The three principles of that team are open knowledge, open source and open data. And so those, the source code is on GitHub and the data is on our Internet data portal or in the process of being put onto it which is IDP.nz.

The code for active systems are, that generally does become open but it takes us more time to be absolutely certain that we haven't included a password or an API key or some script that gives things away or those type of things when we put that and make that open. So we are generally quite behind in doing those.

EBERHARD LISSE: Simon?

SIMON MCCALLA: Firstly, Jay, I think this is fantastic, absolutely love it. It's really, really good. We've had some similar thoughts and done some similar stuff ourselves and I really like the way you brought it together.

It's quite interesting your registrar scoring stuff. We've done a similar thing and [inaudible] debating whether we should publish to our registrars not only what their own score is but actually where their peers are get them to use, essentially, gamify even further. I wonder if you, we haven't done so yet, but we're contemplating it. I wonder if you'd had the same thoughts.

JAY DALEY: We are going to have that conversation in our registrars, yes, to see. I'm not that convinced of the value of it. Pride and shame internally are better rather than people's then reaction to other people's views about things, you know. This is a psychologically designed system more than anything else and I don't know how much that publishing necessarily works. We find we already have lots of people making claims that they're the biggest at this or the best at that or those sorts of things and that doesn't, if you're not careful, that ends up misdirecting the registrants. So yeah, it's not a decision for me. That's a broader decision we need to make and we're going to start that process now.

SIMON MCCALLA: Yeah, the second add-on question to that was we wondered whether publishing some of this information to both Internet users and other registrants is useful, whether that be how good is your registrar or whether that just be general information about the other domains you have on the registrar.

JAY DALEY: I think we certainly could see a point in our industry not too far away where there are registrar comparison sites, where you can go in and the same way you can do that for buying electricity or something and you can see somebody has got a set of data and can tell you who they think is the best choice for you.

SIMON MCCALLA: It looks fantastic anyway.

JAY DALEY: Thank you.

EBERHARD LISSE: Probably, you will not get your registrars to agree to that, but you may get, if you tell them you're going to do it anyway, you may get to agree then what data they want to be published about themselves.

JAY DALEY: I don't think you can guess what they think. They may well agree. I don't know. It's a conversation for us and the regulator, have with them so there's a lot to go in it. But many registrars are quite, either not worried about competition because they don't compete; they have a niche that they deal with or they are very happy to compete because they do that, they have that mentality already and so they don't mind about sharing that. And some of them want to be able to make those claims and if the data is public, then they can point to that data and say, "This is how we make that claim."

EBERHARD LISSE: Okay. Thank you very much. Now Roy Arends is going to tell us a little bit of yet another DDoS Attack.

ROY ARENDS: Oh perfect, thank you. Hi, everyone. I work at ICANN. I'm going to do this a little bit fast because I've got a lot more to talk about than just abuse in [inaudible].

EBERHARD LISSE: You don't have to do fast. We have got time.

ROY ARENDS: Okay. So I was born… no. At ICANN, the DNS engineering team runs the name server names NS.ICANN.org. And at ICANN, we wanted to test [inaudible] so we have an evaluation license at the time that I used to look at traffic from NS.ICANN.org. And we found a few interesting things.

And let me first explain what NS.ICANN.org looks like in terms of architecture. Those are two name servers in two different locations, so four in total. IAD, those are basically the three character [ISA] codes. It's not actually at the airport. It's some data center close by. In IAD, that's Dallas and LAX which is in L.A. It is a few weeks of captured traffic. We uploaded that and the results are going to be seen in a minute.

So where are, authority for a few top-level domains. If you do this trick, you basically get all the names, all the zones that NS.ICANN.org is responsible for which is INT, museum and UG which is Uganda. It runs a whole bunch more but I'm going to concentrate on two of the three top-level domains.

Chapter one, the telephone company. TPC.ntifu, if you were at the DNS quiz this afternoon, one of the questions was "What does TPC stand for?" It stands for TPC.int. It comes from RFC 15, 8, 28, 29, 30, etc. Legacy stuff, don't look it up. It's really old. It's basically remote printing. Yes, you can send an e-mail through a fax. This is literally how it works.

In order to do that, you need a phone number, reverse it and look it up the DNS. But why should calls the phone company? It's because of this movie which is the precedence analysts which has this guy, James Coburn. The way it works is you basically take a phone number, you reverse it – yes, just like [inaudible] but a lot earlier – and you look up the delegation. All right? Sorry, you look up that phone number.

UNIDENTIFIED MALE: Can you stand a little bit away from the microphone because you're breathing into it?

ROY ARENDS: Oh, sorry.

So let's first look at the TPC.int delegation. Sorry I didn't get that so I'm just going to continue.

UNIDENTIFIED MALE: You said something about our nationality.

ROY ARENDS: Okay. Let's look at the delegation for TPC.int which looks something like this. If you do DickSN@ICANN.org for TPC.int type NS – some of you a little bit more technical probably have done this before in their lifetime – you get something like this. You see five name servers for TPC.int.

Now if you translate those host names to IP addresses, you get four IP addresses. The reason you get four, not five, is because the middle one doesn't actually exist.

UNIDENTIFIED MALE: [Inaudible].

ROY ARENDS: Yeah, go for it.

So there are two addresses in there that don't actually route so you're left with basically two routable IP addresses. The first one is the primary. The second one is the secondary, if you will. The primary, actually, is timed out. It doesn't run a name share for that address. And that's probably the reason why the secondary gives a [inaudible].

So TPC.int, yes, it's delegated. The delegation exists but it can't be reached. Keep this in mind. It can't be reached. It doesn't work.

The zone file that I could put my hands on was actually from 2013. This is the latest of the latest-latest. But currently, you can't get it. So it has stopped working years ago, but still, we see a lot of traffic at NS.ICANN.org for TPC.int.

Now the first three are basically normal stuff that you kind of would expect even though it still doesn't work. The first one there that has a telephone number in it, is IDDD which is International Direct Distance Dialing. That number is actually a telephone number that works. There's a fax listing on the other side. Plus 852 is Hong Kong. I obscured it a little bit in order to satisfy ICANN Legal. Plus, 1-212 is New York. Yes, people are still using it and it doesn't work anymore.

Anyway, let's move on to IPC.int. I see Mr. Jeff [Houston] in there. He had something to do with that. He actually killed it years ago. This is what he wrote. He wrote in 2005, I think, that people, please stop using IPC.int. That's over a decade ago. More than ten years ago, Jeff wrote, "Please stop using this."

UNIDENTIFIED MALE: Nobody listens.

ROY ARENDS: Nobody listens. You can never get rid of this stuff.

UNIDENTIFIED MALE: [Inaudible].

ROY ARENDS: Yes, that's a good thing actually.

So what is the idea of IPC.int? Well, currently you have an [inaudible].arpa and IP6.arpa which is the reverse address space for addresses in the DNS. Folks started using this in August 2001. Jeff asked people to stop using it in September 2005 and so, of course, there's absolutely no traffic anymore for this.

Yeah, this is what happens when you roll things in the DNS. Right? You roll to .arpa.

This is a graph of all the traffic to NS.ICANN.org. It's a little bit unclear because I didn't adjust the scale so I'm going to quickly skip through that. This one is interesting. Most of this stuff, this is all the traffic for IP6.int. Of course, NXDOMAIN because IP6 doesn't actually exist anymore.

If you compare the traffic, if you look at the traffic, all the traffic that we see on average, right – this is per day – to NS.ICANN.org is 131 million DNS requests. That's not a whole lot. It's a large number, but that's not a whole lot. I've heard of systems that can do a lot more, that see a lot more.

All of the traffic that goes to INT is actually 105 million. Keep in mind the same machine runs ICANN.org, a whole lot of domains, Uganda top-level domain, .museum, but 105 million of those 131 go to .int. Now of IP6.int is 77 million. This is a huge number. I'm sorry. I don't know what happened there, but this is 59% and I can magically inflate that number by doing some dubious statistics. This is 70% of all INT traffic is for IP6.int. And keep in mind, this should have stopped more than ten years ago. The domain doesn't exist. People are happily querying for it. It's probably automated systems that are happily querying for it. No one knows this. No one sees this. No one is aware of it until you basically have a look at the traffic.

So a small sidestep, NS.ICANN.org used to be responsible for .om which is Oman. March 21st in 2012, again more than four years ago, NS.ICANN.org is delisted from the OM zone apex so it's not anymore in zone file. It's not anymore on NS.ICANN.org and it's also not surfed from NS.ICANN.org anymore. So the second line, April 4, it basically means it's not listed in the root zone file.

So OM exists, just not on NS.ICANN.org anymore. But we still see traffic. We still see traffic even though it was delegated away four years ago. This happens to be – and this is a snapshot – it happens to be China. I know it's from a few more [ASN] numbers. This just happened to be all China.

Now how is this possible? I mean, what do you need to do? Well, the way to do this is to configure a root zone in your resolver in the hard way, the one that doesn't update. Right? And if you did that four years ago, yes, you will still be asking us, NS.ICANN.org, for queries for Oman. Plus you will lose a whole lot of visibility to a lot of other names, to a lot of other top-level domains, but yet, it still happens.

So this is actually where the title of the presentation was about the denial of service in Uganda. Uganda, the pearl of African's crown, and this is their flag. Uganda is here. It's one of the fastest moving countries in the world. That is because it's on the equator. This is a picture on the equator facing west as you can see. Big fan of the equator. I have a special relationship with Ecuador where my wife is from which is also a fast-moving country.

Uganda is interesting because it's one of the very few, maybe the only one, land-locked country that has a Navy. And the reason for that is, of course, Lake Victoria. A few statistics about Uganda. This is some data I'm going to show you to show how people use the Internet in Uganda.

This is the number of mobile subscribers, about 18, 19, 20 million and growing. This is the latest I could get but it's from a couple of years ago. This is the number of mobile subscribers. Of those, right, you see here the graph that shows an increase in mobile Internet subscriptions and that's about 3 to 4 million. Now Uganda, as you could see, is of course the size of Canada in terms of population and it's the size of the islands of Great Britain which is in terms of land mass. I think it has about, if I'm not mistaken, 27, 28 million people. Yeah, like I said, 3 to 4 million mobile Internet subscriptions. And if you look at the fixed line Internet subscriptions, that's a miniscule amount. And this is the amount of estimated Internet users. June 14, about 8.5 million.

So from these few statistics, you can basically say that about 8.5 million Internet users are using the Internet over a mobile subscription, over mobile phones. So even though this is a growth market – not a growth market, sorry – the Internet use is growing there in a fairly novel way, mostly used on mobile subscriptions, it's actually pretty fragile because the landlines going into Uganda are not that wide, are not that available if that makes any sense.

So if you look at the zone file, there are about 5,500 registered domains and they see about 3.2 million queries a day. Now for a zone about that size, that's not that strange. That's the right ballpark. We see that for similar countries, similar zone size, similar Internet statistics as mobile subscribers and landline users. That's a similar amount.

However, about 2 million results in NXDOMAIN which is a very high amount. This is about 62%. This is very, very high. A graph to show it is very, very high and this is, you see [inaudible] here again with a top end function. This is everything that returns NXDOMAIN. So the first one is ICANN. Yes, with denial of service. No, we don't denial of service. We send a lot of traffic and this is to make sure that name servers are up. This comes from a wide variety of probes around the world and we check every top-level domain to see if they're up or not. This is not our group. This is a slightly different group so I can't actually answer questions about that. I need to defer to Francisco Arias for that.

But yeah, this is the first query, dot, dot, dot, dot, zz, dash, dash ICANN [inaudible] monitoring. This is chosen specifically so it will return in the next domain.

The next two, broadbandcompany.ug is a well-known domain. No, you can't go and register it now. It's actually registered but it's taken offline. So those are basically the domains that you expect.

Now the rest of this lot is mumbo jumbo. It's unpronounceable, basically. There's a whole bunch of funny domains. And I thought this might be a domain generating algorithm as part of a botnet. Now a domain generating algorithm is basically a trick that botnet deployers use to find a rendezvous spot. Right? They generate, each of these spots generate the same amount of domains every day, the same amount for all these different ones but different every day, in order to find their commanding control server.

Now if you take this bunch, and I did that – I could just grab this – and you place it here, and this is domain generating, the DGA Archive of [inaudible] Institute in Germany. This is free to use. Yes, you can use it as well. If you have a funny-looking domain that you've never seen before, you want to check if it's part of a botnet, they have a whole library of domain generating algorithms that they've reverse engineered and basically, it will take various different dates from way back until today and will generate all possible domains, put them in a library and you can look them up. So I did the 97 that were on my screen, put that here, and all of the, no exception, all of them are part of this botnet, a single botnet called [nickers] DGA and then some code identifying string. You can do this at home, [inaudible] Institute DGA Archive. The slides are available.

But this is actually pretty bad. This is a large part of bandwidth basically wasted on domain generating algorithm. So if you write a botnet, can you please stop doing this? Seriously though, the botnet writers, they actually don't care. They throw in a few top-level domains because the white hats, in order to defend against this, they need to register each domain and each top-level domain in order to defend it. But the black hats can actually just register one domain five minute before midnight in order to abuse this.

They happen to choose .uj, .ug, which is one of the smaller top-level domains which actually could do without this traffic. We just happened to see it because NS.ICANN.org happens to be a secondary for his domain.

Anyway, so in short, 5,500 registered domains, 62% NXDOMAIN. That's really high. If you see a number like that, something wrong is going on. Yes, Roy, a single botnet, we know that.

So in short, stuff never goes away. Once deployed, it will never go away. You think you're doing a test. You put code out there, people will use it. It will never go away. The Internet never forgets. A single botnet can easily overwhelm a smaller TLD and for me, the most interesting part is analyzing DNS traffic is fun.

I'm going to do a series of this. I'm not sure what I'm going to call them, but I just love this stuff and I find things almost every week and I promise right here to put them in a presentation for next time or maybe in two ICANN meetings. So hopefully I will see you again. Anyway, thank you. Any questions?

EBERHARD LISSE: Did Uganda ccTLD manager notice he was being attacked?

ROY ARENDS: I've sent the information to the UG manager. I forgot his name. I'm sorry. But I sent him information. He was very interested in this.

EBERHARD LISSE: Because I remember, half a year, a year ago we saw the same patterns also coming through one of our name servers that we were running, an NS that we were running a probe on. I speak to my main name server people like PCH, [inaudible], "Yeah, we've seen this. It doesn't bother us. We didn't notice. It makes no impact. Don't bother. Don't bother." It didn't really make an impact on the probe we were running in Namibia but there is stuff going on.

UNIDENTIFIED MALE: Yeah.

ROY ARENDS: Hold on. Can I answer that? There's nothing that the Ugandese registry can do about this traffic. This is something put in a botnet by a third party, a malicious party, put out there. The machines that are querying are just resolvers. They're not open resolvers, just infected stuff. Should a registry be made aware of this? I think so. I think it’s important because it’s their traffic even though it’s also elsewhere.

EBERHARD LISSE: Basically, what it did to me is it taught me how to learn bit more Perl and a little bit more [inaudible]. In particular, the graphics more [inaudible] to plot it nicely. But it did not really affect our local servers and the guys from PCH and then they said, “It doesn’t even bother us.”

ROBERT MARTIN-LEGENE: We do have a note of, we do secondary for Uganda as well. We answer these queries anywhere in the world including inside Uganda as well. So, I mean, the answer would be Anycast from inside the country and outside the country. But we are only one of the names in the name server, so, and the name server is it.

UNIDENTIFIED MALE: I’m kind of interested about what’s happening .int. About a year ago, [inaudible] from NLNet Labs, about a year ago, actually we had a couple of days [inaudible] .int but because being curious and I can confirm the TPC’s status [inaudible]. But it’s actually for about half of the domains in .int. Most of them are completely dead. And I wonder if there is any plans and you’re probably not the person to answer but maybe you know whether there are any plans to do something about that.

ROY ARENDS: One thing I noticed that .int is not signed and I'm a big fan of top-level domains getting signed, .int including, independent of any political discussion between various large acronym organizations, ICANN being one of them, ITU being one of them. However, yes, this is way above my paycheck to answer that so I won't.

JEFF [HOUSTON]: I've been looking at a related issue and I'm kind of curious about your presentation. I've been seating unique domain names and they have a time field. And my authoritative name server, I'm now able to sort of point out that at least one-third of all the queries I get are bullshit because they don't relate to any particular action from an end user. It's just the name is sort of caught in the DNS and constantly re-queries me as the authoritative name server.

Now here are you a little bit further up the tree because you were a name server for the top-level domain. Your presentation seems to sort of imply that the queries are real and the real question is could you tell if the queries were completely bogus, that the names that you were seeing and are seeing are simply caught up in the DNS and you're seeing recursive resolvers simply mindlessly re-querying just to keep the cache alive because God forbid that you should ever forget a name. My suspicion is that the latter is starting to drown us all, that at least 40% of the DNS right now is actually rubbish, and obviously [I'll] just pull up more capacity. But in essence what you're doing is just encouraging the DNS to get more and more trash in it. 40% of all the queries you're seeing are rubbish.

No one really wants to fix it. I'm wondering if what you're seeing further up the tree is the same rubbish, that these queries just are senseless.

ROY ARENDS: Yes. There's a whole lot of crust. What I've seen is, for instance – you know when you have Chrome and you happen to type in something that looks like a domain name, or even worse – there's this gamer site called Twitch. It has this automated scrolling check box going by. If you type in a domain name there or something – if you forget a space after a dot in a sentence – so you say basically, "Hello.Hi," it will try to resolve that. Chrome will think it's a domain name. It will try to resolve it. So at the root, you'll see that traffic immediately because "Hello.Hi" is isn't cached. That's one thing.

You see –

JEFF [HOUSTON]: But then, will you see that same query echo again and again and again without any trigger action?

ROY ARENDS: Yes. I've seen scripts that are purposely designed to take – excuse my French – a dump from BIND's cache and reiterate that just for the cache. I know the script was popular a couple of years ago by a few Internet Service Providers. They'd run it at midnight. They would stop BIND, restart BIND, and repopulate the cache by using the dump that was just done ten minutes ago before BIND stopped. Those are very popular scripts, and they go from generation to generation to older folks, who push it to the new folks. The new folks don't want to touch it because the old guy over there actually used it, so it must be good. Yeah, once deployed, it is very, very hard to get rid of.

We also know that some organizations that look at interesting traffic – it's a little bit of a strange area. If they don't understand this, they want to know what the query resolves to. So they've observed the traffic. Now they don't understand the traffic, so they want to see what traffic is doing. So you will see a repeat.

Yeah, I used to see all that kind of traffic. Hopefully soon at a root I will see the traffic as well.

Sorry, go ahead.

JEFF [HOUSTON]: No, that's okay. That was my point.

UNIDENTIFIED MALE: Okay.

EBERHARD LISSE: Okay, thank you very much. Now Norm Ritchie will speak about the tool called Luminous from the Secure Domain Foundation.

NORM RITCHIE: [inaudible].

EBERHARD LISSE: No, they can be loaded for you.

NORM RITCHIE: Oh, okay. So you have to [inaudible]

EBERHARD LISSE: Talk.

NORM RITCHIE: Hello. My name is Norm Ritchie. I'm the Chair and the Founder of the Secure Domain Foundation. Sitting beside me is the Director of Operations, who I work with very closely every day, Drew Bagley.

I guess many of you probably know me. I've been around for a while in the ccTLD community and in cybersecurity more recently, and also I had to do a mandatory stint at the ISC.

Drew is a privacy lawyer, a very good one, and also comes from the FBI. So he's a great addition. So we're going to do this together.

I'll let you talk.

DREW BAGLEY: All right. Yes, I'm Norm's slightly less famous sidekick. Slightly younger. Today we're going to talk about our new DNS intelligence platform, but before Norm goes into that, I wanted to give you guys a little bit of a background on who we are and what our current platform has been.

EBERHARD LISSE: Use this one.

DREW BAGLEY: Oh, that's right. I'm using the wrong control. All right. Here we go. Many of you know us. Some of you may not. We are a non-profit incorporated in Canada, founded about two-and-a-half years ago. Our mission is to give Internet infrastructure providers the tools they need to fight back against DNS abuse and ultimately cybercrime.

The way we do this is we serve as a clearinghouse for malicious domain name intelligence. We get our members to share information with us on suspended domain names, IP address, e-mail address, and so on, and then we try to add value to that. We have the past couple years with our current API by mapping relationships and producing a reputation score for those indicators.

Another role we serve is as a trust group so that we can be a trusted third party through which registrars, registries, and hosting companies can share information, ask questions behind the scenes when they're dealing with DNS abuse, and also really come together and form best practices.

What we've done so far and what we're hoping to do a bit more of later this summer especially is to engage more with researchers and to produce more of our own research. As we've already heard from today from many of the speakers, there's lots of interesting trends if you're looking at which registrars could potentially be the problem registrars. That can actually solve problems if you can have some data to figure out what can be done to stop there from being some many abusive domain names with certain registrars, or if we're looking at pricing data or whatnot. So those are other things that we do.

With our current API, what we've been able to do is help serve use cases for registries, registrars, cybersecurity analysts, and researchers just with queries themselves. For registrars, we've helped to answers questions about whether or not account holders might be associated with previous known malicious behavior. For registries, they can look at how clean their TLD is, even on a daily basis, and keep track of how many malicious registrations are popping up.

For cybersecurity analysts who might have one little piece of information about a known bad actor, they can come to the SDF and find out more information and see how much more prevalent that malicious behavior might be and how that might be spread across TLDs. For researchers, as I mentioned, a lot of our data is useful for statistical analysis because we benefit from having data from a wide variety of members.

With our current API, being a reputational API that serves those who query the API and provides a score as to whether or not a domain name is associated with previous known badness, one of the best use cases we've seen for proactive anti-abuse, which is what we try to promote is when our members will actually go ahead and integrate our API into their registrations process, for registrars in particular.

What his allows us to do is, in the same way that credit cards are verified, we can have a registrar verify that a would-be customer is not someone else's bad customer who just got suspended the day before from somewhere else and is now trying to do the same thing again.

What we've seen as part of this is, when registrars are able to integrate our current API in this way, they actually save cost on the back end, because then they are dealing with less abuse complaints, which requires fewer man hours, and therefore which requires fewer resources spent.

So this has been a model that we've been hoping to promote going forward, but as I'll explain, it's been a bit hard to scale with our current API because we've been a victim of our own success.

So what we can really find out with these pieces of information, though, from the data that we have, and with a query that a registrar, registry, or hosting company will do, is we can actually find out a lot. We can map out relationships going back historically because we have a wealth of historical WHOIS data that we're able to pivot these suspended domain names, IP address, and e-mail addresses off of.

As you can see, you start off with one indicator for a query. This query could take place proactively. Or even if it's taken place after the fact as part of an investigation, it then can link a single domain name that's already known to be bad or that's being questioned to a bunch of other indicators. Or you could start off with an e-mail address at registration and find out that this e-mail address has been used before with a known phishing domain which was linked to the IP address of a known botnet command-and-control domain name, and so on.

Our current API has been able to just give a real basic yes/no reputation score. We took this and thought about what more we could do with mapping out these relationships and decided that we needed to get to work on developing a new API and a new platform.

But as you can see, with already our current infrastructure, we've been able to tell a lot more than just a simple yes/no to a block list. So that's how we've seen ourselves. As I was mentioning, we're an information clearinghouse, and this is our value-added. Now we're trying to take this to the next level because the bad guys have been able to scale for the past several years, and this is our response, attempting to scale back.

As I've alluded to, the problem with having lots of data, using that data, and mapping it in complex ways is that it's a lot of data to deal with, and therefore is very resource intensive. The data changes constantly, as you all know. Data can change on a daily basis, a weekly basis. But even within hours, you can have registration data that's changing. And the more we've seen our members use our API, the more queries we've ended up getting as members have wanted to rely on it more and more.

Additionally, searches need to be able to incorporate wild cards and not just be exact matches to really be useful in tracking what's going on with malicious actors, as they may use DGAs or as they may just modify their own e-mail addresses that they're using to command their empires. On top of all of that, just maintaining our current system has been operationally intensive.

All of this has led us, for the past year, to work on a much more scalable, more powerful system that's also much more useful for our members.

Norm is going to go ahead and talk about that.

NORM RITCHIE: Okay. Today we're introducing our new interface that we call Luminous. As Drew mentioned, we've been in this for quite a while. One thing you got to realize is, as the small non-profit organization, we're very resource constrained.

We actually started this two years ago. When we first launched the SFD, it was quite successful right out of the gate. It got a lot of usage, but that usage killed us. We couldn't keep up with the demand for it. Operationally, it was a lot of long nights trying to process all the data at daytime.

Knowing that that wasn't going to work too well for us, we simply started designing a new system, like I said, two years ago. It's based on Cassandra, so it's much, much faster. Knowing that we couldn't predict what would be required in the future, we keep all data parsed all the time. It's very scalable because [inaudible] WS. So we can add servers as the demand increases. Also with Cassandra we can have redundancy on the servers as well.

So there shouldn't be any downtime. It should scale with demand. It has much lower operational needs. The old system required us basically block lists every night and then pump out the results back into a database. The current one just loads it in. it does everything on the fly. It's very cool, actually.

It's also highly flexible. But more flexibility in the API also adds a bit of complexity from the user's point of view. What we were trying to do before is basically predict what everyone wanted and make a nice, easy query for that. However, whenever you need a change, we had to go back and reprogram things. So the new interface has a lower level command that then you can just pick and choose and incorporate them as you see fit.

Different levels of query interfaces. Right now there is a command level, which is not available to users, unless there's a special set-up to get access to it. API is the normal way, of course. There's going to be a web interface as well: sometimes you just want to single commands rather than automated API use. And the output is going to be an XML, JSON and text as well. But why you'd use text after having XML and JSON, I'm not sure.

So, what's there? I think I'll show you what commands are there in a second, but I just want to give you an idea of the data that's there. We've been collecting data for two years for gLTD data. So we have a great inventory of gTLD data. There's about 80 million records that we're processing. We have another set of somewhere around 180 million records. Plus, it's growing every day, about 120-150k per day. So we have all that WHOIS data. As we mentioned before, that's what we use to discover new malicious indicators.

We also have quite a few black lists and block lists that we import. Now, these came from a bunch of cybersecurity experts. We asked them which ones we should get, and they told us. But for the most part, if you want to look, look at Critical Stack. They list about 105 block lists. So for the most part, we have a lot of those. We also have some unique and some customized lists as well.

We have about 70 million unique indicators currently. To be honest, we have about 75 million flagged pieces of data that are in some way malicious. So the amount of flagged information is actually quite high. And we're adding constantly 10-100k per day, because there's just that much of it.

The classifications we use for the data. It's one thing to say, "This is bad," whether this is an e-mail or a domain or an IP address or whatever. It can be bad, but the question is, why is it bad?

This shows the breakdown that we use for classifying malicious domain. If you've ever looked at ThreatExchange – that's the Facebook-sponsored exchange – this will look similar because the [inaudible] are the same thing. My favorite in there is the botnets. I just have a thing about combating botnets and command-and-controls.

We can add to this list, or if something is not being used, it could be taken away, I guess, as well. But we can also add in new classifiers quite simply.

The other thing we do is add in white lists, which is an interesting one. Let’s say Google hosts a lot of sites, and we know Google itself gets a free ride. So we can’t say the Google domain is malicious because it’s hosting a lot of content that is malicious on their site. So that’s why we white list certain domains. You’d probably recognize them as probably the top 100,000 websites in the world.

So there’s a sample breakdown of that number before that I showed, which was about 7 million unique indicators. It shows the breakdown, so you get an idea of what’s there. Looking at the bottom of that list, you see phishing malware. This classification called suspicious is like an indirect indicator. So it’s not directly attributed as being malicious, but it’s associated with one that is.

The botnets are actually a small number, but obviously higher in priority to deal with them.

Okay. Some example commands that we have in here. Some of these are exposed on the API. Some are not. To be honest, I did this the lazy way and took the command-level commands.

We are able to expose each command on the API for each user. So we can actually change that, depending on who the user might be. Some of these are useful. Some are probably not. WHOIS actually just returns a WHOIS record, but it’s parsed. So it's much better, I believe, to house parsed WHOIS data because then you could do something with it. I'll show you what it looks like in a second.

Things like the WHOIS server returns or a WHOIS server for a domain is not very useful for anybody here. However, we use that internally, and that's why I put it on there: because I say I got lazy.

Something like a WHOIS reference actually will take a domain or an e-mail address and then pivot on that. It'd say, "Okay. Take this e-mail address. Tell me all the other places that it occurs, wherever you can find it."

Bearing in mind that we have a lot of WHOIS data, nobody, including us, has a complete set of WHOIS data. So anybody saying they have all WHOIS history? They don't. It's not possible. The reason for that is that only the registrars know when a change is made to WHOIS data. You can have cases where you can change a record, do something malicious with it, and change it back within 24 hours, and most people never pick up that change, just because there's just so much WHOIS data to go through that you can't do that. You can't catch them all.

So we will report on what we have, and it keeps getting better because we're constantly processing WHOIS data. Currently in our production system, there are just shy of 42 million records right now processed in there. We add about two million per day, and we'll keep doing that. We also keep adding malicious indicators every day. So over time the system will get smarter, will get better.

The other thing we will do is run our own agents in the background. That will just go through and run out scenarios, starting with a malicious domain name, a search [tracing] through itself to just make new discoveries.

The flags are what's set, but there's another feature here that's Export, which is probably the most useful for a registry. With one command, you can export all the malicious domains, e-mails, or whatever for a TLD. If you want to break it down to sub-TLDs, we can do that as well. So you can look at those at the domains.

The other ones work on DNS, etc., MX. The last one I want to mention here is one called Report. You can actually build a report with an XML template that will do the various queries for you. So you submit that template and generate a report out of it. It's actually pretty cool. It's hard to explain, but it is cool.

I was going to show you some live, and I feel like I'd probably fumble around if I did that. So luckily I actually have a slide here, because I can't switch easily.

Here’s an example of WHOIS query. It actually is for securedomain.org. In this case, this domain was not in the databases from a test database. So it ran out, got the record, parsed it, and then presents the information back with some geolocation, if anybody is into doing maps and stuff like that.

You can see how it’s parsed the phone number. It broke that down into its components. It has parsed the address fully – it was able to work that out – and e-mail addresses that I already mentioned, and the dates. So now you can actually make this a lot more useful if you’re doing any type of analysis to get the things tagged. That record is truncated because it just wouldn’t fit on the slide.

Up next, we’ve been doing betas now, and we have some early adopters. A big thanks out to CoCCA for bearing with us as we were developing our system and they were trying to integrate it at the same time. It’s very exciting for us. I hope it’s also exciting for them, but I’ll be honest: it was a lot of long nights and a lot of pain.

Near-term, we’re going to keep evolving Luminous. We need that in submissions and vetting of those submissions. So when we’re bringing a source from somebody, we have to have some way of vetting that that actually is malicious.

We need to be able to remove content as well. If you have a domain or URL or whatever, it can be malicious now. That doesn't mean that two years from now it's still malicious. So we have to be able to take those back out.

We're adding a watch list. There's some cases where you just want to keep your eye on something. I'm quite keen on proactive anti-abuse. Lots of times when you see a domain name being registered, you know it's going to be used maliciously. You just know it. But it's not doing anything, so you put it on a watch list. It'll just notify you when the name servers come alive or whatever, or anything changes on it.

Batch processing. Of course, we adding that very shortly, in the next couple of ways. Ongoing, we'll keep crunching more and more WHOIS data and indicator data.

DREW BAGLEY: What anyone who's interested can do is join SDF. Joining SDF is free, and use of the API is free. Everything is free because we're a non-profit. If you want to take anti-abuse as seriously as we do, then I implore you to share suspended domain data with us, because when you become an SDF member and you're able to use this great system, the more data you contribute to the system, the better the system is for everyone because it's that many more relationships being mapped out. So it's truly better for everyone. The more you give, the more you really get out of this system. We are looking for any suspended domain lists, e-mail addresses, IP addresses, and so on.

Also, at ICANN54 in Dublin, we held a session – and this message is more geared towards registrars, although registries might be interested – on abuse reporting standards because that's a common problem: there are inconsistent abuse reporting standards. So even when you have abuse complaints, it's not always easy for people to follow up on them.

We're working with the i2 Coalition to try to get the community to come together and agree upon some standards. I encourage you to visit our website and contribute to that discussion, too, because we're hoping that, when we go to Hyderabad, we can further that discussion and get the abuse reporting criteria more consistent, which will help routing abuse, investigating abuse, and then, in our role with you using our system, to take care of abuse.

The signup process for SDF is simple. Just go to our website or e-mail us at register@securedomain.org. We have a data sharing agreement that we've developed. Get that signed, and then you get an API key, and then you can build out your own system from there, or use the user interface and whatnot, and, like I said, share data.

NORM RITCHIE: Do we have question time, or are you going right to Garth?

EBERHARD LISSE: Yeah, we're having –

DREW BAGLEY: Maybe because mine's on.

EBERHARD LISSE: Turn yours off.

NORM RITCHIE: Oh.

EBERHARD LISSE: Oh, that's much better. Any questions? In the meantime, I have a question. Have you got an API for Perl?

NORM RITCHIE: For what?

EBERHARD LISSE: For the Practical Extraction and Report Language, Perl.

NORM RITCHIE: Our developer – his name is Q, honestly, that's actually his name; it's very cool – actually thinks we should be building a library for this, so I'm surprised that was the first question, actually, because I thought, "Are you crazy?"

EBERHARD LISSE: You know me very well.

NORM RITCHIE: Yeah. At the end of the day, what we want is for this to be used, because that's what really makes it work. So if that's what's required, then we'll put that on the list of things to get done, or hopefully get someone to contribute Perl interface.

EBERHARD LISSE: Yeah. My point is that we run CoCCA, and when it's working, CoCCA will upgrade the version that does it. But that only does it when you register them. I'm currently working on something including Zonemaster and some other thing into a Perl script that I could run. It would be nice, if, while I'm checking that, I can check that as well.

NORM RITCHIE: Yeah, you could just do [cURL]. It would work.

EBERHARD LISSE: Dimitri?

DIMITRI LORENKO: Well, I would probably second the Perl support. Actually I would give you another thing to consider. I love the XML – actually, I don't. Do you think you call also do JSON output with REST API entry point because [inaudible] –

NORM RITCHIE: Yes, yes. Sorry. JSON is there as well.

DIMITRI LORENKO: Okay, great. That's good. Then I'll pass it over to [inaudible].

NORM RITCHIE: Yeah. And text.

JOHN LEVINE: I'm John Levine. I got here late, so I don't know whether you addressed this. There are some widely used domain black lists, like the ones from Spamhaus and SURBL. They don't accept third-party contributions because the data is so cruddy. Partly it's because people are malicious, partly because there are people who believe that any domain associated with pornography is evil.

So what's your plan for maintaining data quality?

NORM RITCHIE: The submissions will only come from our members, so we wouldn't be taking submissions from the general public. It'd only be from our members, our registries and registrars, and hopefully some hosting companies soon.

JOHN LEVINE: If I may be maliciously cynical, how hard is to join?

NORM RITCHIE: You have to get past us.

JOHN LEVINE: Okay.

NORM RITCHIE: It's a manual process for joining.

JOHN LEVINE: Okay.

DREW BAGLEY: You see how strong we look, too. Intimidating.

EBERHARD LISSE: No more questions? Thank you very much. Garth Miller from CoCCA is up next. He will talk about that implementation of the latest version of CoCCA tools that Norm just mentioned.

GARTH MILLER: [inaudible].

EBERHARD LISSE: It’s there. You use the clicker.

GARTH MILLER: Thank you very much. I’m wearing two hats today. I’m the administrative contact for Christmas Island, which is a very small ccTLD external territory of Australia. We have about 40,ooo registrations. I’m also the Project Manager for the CoCCA software project.

CoCCA is a free registry system currently used by I think almost 60 TLD, gTLDs and ccTLDs. One of the advantages of being a small ccTLD, like Christmas Island, and also working with CoCCA, is that you can just do things quickly. When Norm comes up with his interface and this tool, we can take an executive decision and just do it. We don't have to consult widely. We just do things, which is an advantage.

We've been able to work with Norm over the last few weeks and over the last year or so, but mostly on this new API to develop and integrate it into CoCCA. If somebody is using the CoCCA software, they don't have to do any additional development work. They can just connect to the SDF right away as soon as they get credentials, and they're often running.

This is really our third attempt at essentially – as a small ccTLD, maybe I'll preface this discussion. In the gTLD environment, there's an assumption that a lot of the burden on abuse mitigation and everything happens at the registrar level. In small ccTLDs, the registrars are often small, localized [Ps], people that register manually. They don't have a lot of resources to prevent or mitigate abuse.

Under the CoCCA model, a lot of the abuse mitigation attempts to control that actually occur at the registry level, which is a little bit different than a large European or gTLD, where there's an assumption that the registrars have a lot of resources.

If we look at the NZ model, we assume that the registrations are by EPP. There's a certain level of competence. Under the CoCCA registry system, we assume that the registrars don't have a lot of resources, don't have a lot of technical competence, and we've shifted the burden of actually trying to mitigate abuse, or at least provided those tools, to the registry level.

Historically, it was always pushed out to the registrars, but increasingly, a lot of registries are actually feeling heat from regulators and whoever else to actually mitigate abuse at the registry level. That seems like the logical place to try to do it.

EBERHARD LISSE: [inaudible].

GARTH MILLER: All right. This is our third attempt at trying to do something at the registry level. Our first attempt was in 2012. What we did was we built into the CoCCA platform a tool where, every time someone registered a domain, the registry would send the registrant and the admin contact an e-mail with a link in it. They would then have to return to the registry, accept the policy, and confirm the contact details before the domain was delegated.

So from the registrar's point of view, the registration happened completely normally. We just didn't actually delegate the domain until we'd actually had the registrant come back and agree to the policy and verify that.

There were two problems with this. One is it didn't actually look at any abuse. It was just checking to see whether the e-mail was valid. The other problem is it created a lot of workload for registrars because, with names that were registered, people weren't clicking on the link because they didn't recognize who the sender was or they just didn't get it because of spam. So we had a lot of complaints from registrars saying that it created additional burden for them to actually sell our domains.

So a lot of the users of the CoCCA platform were reluctant to use this tool because it was reducing their sales because it was increasing a burden on the registrars.

So that was our initial attempt. It sounded good at the time, but even for Christmas Island, we abandoned that fairly quickly.

What was our additional refinement? In 2014 we added an additional layer, which is that we escrowed data in ICANN format for Christmas Island, which is also part of the CoCCA software; it's all built in, if you want. Some of the escrow agents basically started to offer services as part of their ICANN compliance things, these domain-assured sort of things, where they would basically extract the escrow data on a daily basis, compare it to various databases, and then send us an e-mail saying this domain appeared on it.

That was a problem because it was completely a manual process. Someone in our staff had to trawl through the e-mails and look at things. It wasn't integrated with the registry, and the actual action, if any was taken, was taken very delayed. It might be two to three delays by the time we registered. It's escrowed. We get the report back. A lot of abuse happens very quickly within the first 24-48 hours of malicious domains.

So that didn't really work too well, and it was also a paid service, so a lot of the users of the CoCCA software were reluctant to pay an escrow agent for the escrow and for the additional abuse mitigation.

Out latest effort with Norm and his new API is a hybrid of the initial two efforts. What we've done is we've integrated the API to the software, and what we do is we try to evaluate at the time of registration whether the domain is high risk or no risk. So essentially we've taken our first attempt, which is to force the registrant or admin to activate the domain manually and combine that with SDF, where we basically take a registration. After the registration, we then go to the SDF, extract whatever information we can about name servers, IP addresses, e-mails, MX servers – whatever we can look at – and, if we consider it to be a low-risk registration, then we automatically activate it.

If it's a high-risk registration, then we go through our legacy activation thing, where we send an e-mail, and the registrant has to come back and manually activate the domain. So it doesn't block registrations. It just puts a delay in there and creates an additional burden to do the activation.

[inaudible] down at the bottom there.

Currently, there are a lot of tools. I'd say we're an early adopter, so as of a few hours ago, we're still trying to trawl through Norm's effort and see what we can extract, and that effort will continue. But currently we look at existing e-mails, and then we look at the domain associated with that e-mail, and also the actual name because sometimes people reuse the same name to see if that's been associated with abuse. Then we do a pivot and we look at the MX server to see whether or not – because they might change the user name but they're still using the same domain and MX server. Then we look at name servers; whether or not that name server has actually been associated; either the host name or the IP.

So we're not blocking the registration. We're just saying that, if it has a reasonably high risk of abuse, then we force it to go through a manual process. The manual process is still done either by the registry, the registrar, or the registrant, depending on your policy matrix. But if it's a bogus registration, there is some delay involved, and whoever the malicious party is will have to actually come back to the registry to actually have the domain delegated.

If the IP that they come from is actually known to be malicious, then we automatically force an administrative delegation. So even if it's a manual registration, if it's a known IP that's malicious, then we'll block the registration, and the delegation will only occur after manual intervention.

So the idea is really to decrease the amount of registry/registrant contacts [that] the registrars don't like, so the registry is only contacting the registrants directly in the case where it's basically a high-risk registration.

Manual registration. We basically look at the IP address. If someone does return to the registry and tries to register and activate it, we see if that's on Norm's black list.

In addition to the tools that we're using at the time of registration, we're also, in the background, continuously walking through the database and comparing it to SDF. So we have at the time of registration, but we're also just continuously looking through the inventory of contacts and domains, and comparing that, just continuously walking through.

Any information that we get from the SDF is actually stored in the registry database. So basically, as I say, it might be good today but bad tomorrow or bad tomorrow and good today, so we're constantly keeping that historical record and associating it with the domain history.

There are policy considerations, obviously; whether or not the registry and the registrars agree to the registry contacting registrants directly. That has to be considered. The other issue is, in any registry/registrar interaction, one of the issues we had in the past was that they didn't recognize the registry, nor did they recognize the registrar because it was sold through a reseller. So when we would send an e-mail to the registrant saying, "You've registered this domain through X registrar," they still didn't know who that was.

In the latest versions of CoCCA, we do have an EPP extension which allows registrars to associate resellers with the domains so that, when we send an e-mail from the registry and we tell the registrant who we registered through, we actually give the name of the reseller, as opposed to the registrar. That hopefully will cut down on the amount of confusion to the registrants.

Inside of CoCCA, it's very simple to configure. You basically just set up the SDF. It's all done through the administrative GUI. It takes three steps and just a couple of minutes. You basically just get the credentials from Norm and add those. You go to the zone and say, “Require Activation,” and then you say, “Activate old [inaudible].” The new feature is high-risk domains, so you would just highlight that, and off you go, pretty much.

This is just manual activation. If you did happen to be a high-risk domain, you would be set back to the registry. You need to confirm that the contact details are correct. Also, you need to agree to the policy. Often when people register domains the registry requires that the registrant be told of the policy, but once you go through a registrar and a reseller level, hardly [inaudible] should be made aware of the policy or agreed to the policy.

So by forcing high-risk ones through this way, we’re basically forcing them the registrant to read and agree to the AUP and the other policies. We track all this information and store it in the database – what time and date they agreed, from what IP address – so that if something is high-risk and it turns out to be a source of complaints or suspension, we actually have evidence that they read and agreed to the policy, of which the reality is, through most registrar reseller arrangements, no one has ever actually looked at the policy.

That’s pretty much it.

EBERHARD LISSE: Thank you very much. Any questions? Excellent. We're not running that version yet. We'll wait for the beta testing to be completed. When it's in production, we want to have a look at it for .na. But also, I didn't know that it would go an ongoing assessment, which is quite helpful because then I won't have to do it in Perl.

Thank you very much.

Okay, now we have Gary Gale from What3words. We had to have a slight change in the program. We had to move him up one slot because he's traveling a little bit earlier than I thought. So he will give us the next presentation.

GARY GALE: [inaudible].

EBERHARD LISSE: Yeah. Or we can do it manually on the machine [inaudible]. We're just waiting for the presentation to upload for a second, as you can see.

Unfortunately, it's not a bandwidth program. 78 megs. Okay.

It probably gets uploaded from here to the U.S. and then played from there.

GARY GALE: All right. This is going to be interesting. Hello, everyone. My name is Gary. I don't know how to work this piece of software. This is really bizarre. All right. We'll give it a go.

You're not my normal audience. My normal audience deals with maps. We deal with geolocation. We deal with location technologies. About the only time I typically come into your world is either managing the zone records for the company I work for, or registering a new domain.

However, this is another touch point, and that's the reason I think I've been generously invited her to talk to you. Very, very quick introduction.

Does this work? Yes it does. Sort of.

If you're on Twitter, that's me. In addition to working for What3words, I also run a map blog, which has been going for quite a few years.

Before I started at What3words, I worked at the Ordinance Survey, which is the United Kingdom's national mapping agency. Before that, I worked for small Finnish company which made mobile phones, called Nokia. Before that, I worked at Yahoo, back when they were still cool and were still doing stuff.

To try to set some context for this, I'm going to start with a map because all of my talks start with maps. This is a very, very old map of the world. It’s called the Len0x-Hunt globe, and it’s quite famous in my circles because it’s one of only two places that are known to have the infamous phrase “Here Be Dragons,” or, in Latin, “Hic dracones” put there on a map to indicate that, basically, the cartographers of the day hadn’t got a clue what was there, and it was better than a blank space.

Fast forward from when this was first produced to today’s modern age, where we all carry around our smartphones, which have nice little maps with that reassuring blue dot, which glows to show where you are. Basically, it’d be easy to think that I’m out of business. We’ve mapped the world. We know exactly what’s going on. We can all just head to the bar and have a beer.

But that most certainly isn’t the case for a surprisingly large amount of the world. This is a picture of a place in Dar es Salaam in Tanzania. It’s actually Dar’s central market. This is where the city gets all of its fresh food from. It looks a little bit untidy, and the reason it looks a little bit untidy because this is in the middle of a slum, or as the government puts it, an unplanned settlement. The market is actually engulfed by the ever-growing metropolis. I’ve been there. It’s full of life. It’s full of vibrancy.

Yet, when you look on today’s maps, there’s not much there. This is here, Nokia’s old mapping service, and it’s got some roads, which is good. If I look at Google, it’s a little bit better.

Some of the roads have names. Some of the points of interest on the map are listed.

It's only when you look at the satellite imagery that you realize just quite crowded and how populated that area is. The last time an official census was taken there, I think somewhere in the region of 3,000 people lived in this tiny, tiny little enclave. It's taken community mapping projects, like OpenStreetMap, to actually fill out the map to show what's really there.

Now, the reason I give you this example is because, when it comes to addresses, those nice little strings of characters with commas, names, numbers, and postcodes in them, we tend to take that for granted. But the world itself is as poorly addressed as it is poorly mapped.

Most people don't realize this and don't come across it because they use one of these things, a geocoder. This is the magic piece of computer software which translates an address into a latitude and longitude, and vice versa.

Geocoders are to maps almost what DNS to the Internet. When it works, nobody cares. Magic happens. You put a domain name in, you get an IP address out. Likewise, with a geocoder, you put an address in, you get coordinates out, and your map works. Nobody know or cares, really, about what goes on, outside of these little industry niches, until something goes wrong or it doesn't work. Then people complain and bitch and moan like anything.

The reason is because people tend to make assumptions. They make assumptions like this. They think an address will always start with or at least include the number of a building, except when it doesn't.

Okay, so we're going to have numbers. We accept that. But none of those numbers will be zero, except when it is. But they won't be negative. The numbers of the houses on the street will definitely be positive, except when they're negative. The building number will only be used once and the street, except when it is.

This is a real road in Ireland, separated roughly by 20 miles, but there'll only be one number in the address. We can work that out. Ah. But there'll be a road name. We will definitely have a road name. Ah.

Actually, this is a remote sensing station somewhere in the middle of Oklahoma, run by the U.S. Geological survey. They actually get posts delivered. Amazon actually delivers to them.

Now, all of this is a talk it in itself, and there's a wonderful blog post which has been written called "Falsehoods Programmers Believe About Addresses," which is nice and funny and convenient. It's easy to laugh, just in the same way as, when the Internet first started out, people in Britain came across web forms which asked for telephone numbers and zip codes rather than postal codes, and our crazy postal system in the UK didn't validate. But, actually, 75% of the people living in the world either have no address, an inadequate address, or a poor address.

That 75%, to put it into context, roughly means four billion people don't have any form of address.

EBERHARD LISSE: Are those people [inaudible]?

GARY GALE: I think those people would love to be on the Internet, which means they fall back on something like this, which, in my industry, is recognizable as geographic coordinates, latitude and longitude. But unless you're incredibly clever, and some of the people I work with are, you can't really tell much about that.

So, forced to use latitude and longitude, people come up with alternative suggestions. In Ghana, for example, there's two addressing systems which are currently being run as a pilot. There's UMLIS, and there's the ASI ZIP Codes. I find it quite touching that both of these programs have so much faith in their new addressing system but they don't actually use it in any of their signage, which means that you get street numbers like this. And this. And this.

Now, this is deeply confusing. I haven't got a clue what this means, and a lot of people who live there haven't got a clue what it means. They just put these numbers which they've been given up onto the walls of their house in the hope that somebody will be able to work out some form of meaning.

That means that, really, they tend to fall back on things like this. Opposite the [inaudible] oil petrol pump. In my industry, we see a lot of things like this. We see behind two pond trees, and, one of my favorites, where the old building used to be, which is a well-known geographic landmark.

Again, it's easy to find this funny, yet people like the Universal Postal Union think this is very, very important. People think that a 1% improvement in addressing data can bring around a 25 billion euros reduction in cost. Now, whether that's before or after Brexit, I'm not quite sure.

There's only 50 or 60 countries which actually have a postcode which is even kept reasonably current. Even the founder of the biggest crowdsource mapping project in the world thinks that actually addressing rather than mapping is rather more interesting.

It's expensive. It's frustrating in the developed world, but let's be honest. Nobody is going to die or be uneducated if they can't get their Amazon Prime delivery. But in the developing world, it costs lives and it limits growth.

So we thought as a company that the world would be a much better place if everybody could easily talk about every single part of it. We called it What3words.

What we did was we took the world metaphorically, not literally, and we divided it up into a series of three meter by three meter squares – round about 57 trillion of those, to be precise. We worked out the geographic position of each one of those squares. Then we assigned three words – three unique, unambiguous words – to refer to each one of them.

For example, this clever little slide, which one of my creatives put together, shows the address of the What3words office in London. It's called Index Home Raft. It works everywhere, land of sea. The reason we did this is because there's a reason why phone numbers are the number of digits that they are. People have a very, very difficult time remembering long strings of digits. People cannot remember or communicate accurately latitude and longitude to the degree that you need to accurately position something. An alphanumeric string? That's round about 10% recall half the time. A latitude and longitude? Forget it.

But words, what we use in our everyday language, are very, very easily remembered. Because they're words, you can use them pretty much anywhere. You can talk about them. You can write them down. You can send them in e-mails. You can put them on webpages.

One of the things that we've done is, in the ordering of the words, you can spot errors. For example, we make sure that a plural form of one word is geographically the other side of the world, if such a things exists, which means that you can plug into our software, into our algorithm, a three-word address, and, provided you tell it where you are, it will say, "Well, are you sure you mean that? Because that's in Australia. However, there's an alternate form which is a lot closer to you."

It also means that the shorter words, the words which we tend to use most in our languages, are located on the land in areas where that language is natively spoken. The longer words are in the seas. Of course, we're not quite as arrogant as we used to be, and we recognize that not everybody speaks English. Therefore, we support multiple languages. This works in pretty every industry sector that we've spoken to.

Now, I want to touch back on geocoding, because this is where my world and your world intersect: those pesky little sections of a domain registration from which ask for your address. There's a buzzword which is going around in computing terms, and that's big data. Geocoders are typically big data-ers. They work online, on the Internet. You may see a single API endpoint, but in reality, that is hundreds of servers on a backend trying to work out what it is you're typing.

What three words fits in roughly 10 meg of storage? That's data and algorithm. For those of you who are old enough to remember floppy disks, this will fit on round about eight of them.

EBERHARD LISSE: Those are stiffy disks.

GARY GALE: I'm British, so there's massive innuendo to be had in that phrase, so I'm just going to move on. That means that we can put this in a mobile SKD which fits on a tablet or on one of your mobile phones, and it just works. So for those pesky times where you don't have any data coverage, or, alternatively, when your country chooses to secede from the European Union, which means that roaming rates won't be capped, this still works.

Of course, for those of you who like to live in the online world, we have a nice little API. It spits out JSON because humans can read JSON. It spits out GeoJSON, which is a geospatial form of

JSON. For those of you who like to live in XML, it will do that as well.

It also means that we can stick this into desktop GIS systems, like Esri, which is the largest geographic information provider in the world. It also means that there's a whole community which has sprung up around this of people wrapping our API and our SDK in natural language bindings to make this being used in as many places as is possible.

To illustrate why this works both on the macro- and on the micro-level, a couple of weeks ago, I needed to go to the Apple store in the Westfield shopping mall in Shepherd's Bush. I get an address, which I plug into the system, and helpfully showed me that the Apple store is slap-bang in the middle of this fast [boarding] shopping mall, which kind of gets me there, or at least only gets me there to do the door because, actually, if I'm driving, the address of where I want to go is somewhere to the north of the mall, where the car park entrance is.

If I'm going by public transport, then the address is somewhere to the left-hand corner of the mall if I'm going to one particular station. But if I'm coming from the opposite direction, the entrance is another place. Addresses in a lot of cases aren't nearly granular enough. The address for where I'm sitting doesn't really have that much relationship to one of the many entrance points to the massive building that we're in now. But I'm definitely here.

I'm not out in the parking lot. I'm not out at one of the pedestrian entrances. The entrance where I might want to deliver something to the shopping mall is different again.

But these are first-world problems. Going to the Apple store is very, very nice and handy when your trusty map book breaks, but in the real world we find that our solution is being used to help give people an address where there's no power, so battery-powered lighting to enable people who are totally and utterly disenfranchised and not even viewed as being officially there by the government. They can have light.

It's being used to locate mosquito traps in rural Africa because, for some strange reason, these mosquito traps are colored green, which they then place in the foliage near the water where mosquitoes like to breed. So they're really, really difficult to find.

It's being used in the shantytowns in Brazil to deliver postal services and packages to people who have been neatly carved out of legislation. So the Brazilian Postal Service officially doesn't have to deliver to them because the government doesn't recognize that they live there.

It’s being used in the townships in South Africa to deliver medicines to people who really need healthcare and yet would normally be faced with up to a three-hour walk and then a six-hour wait at the hospital to get their prescriptions.

It’s being used by the United Nations to report disasters. We tend to laugh that geography is relatively constant, apart from continental drift and natural disasters, but in the case of natural disasters, where you have a significantly displaced population, they’re definitely there, yet they’re not officially there.

And for more fun things, it’s being used right now for [Glastonbury] Festival so that people can know where to meet when their tent gets washed away in a torrent of mud. It’s being used for bars, hotels, and restaurants, especially in very, very old cities, where the streets are very, very narrow and it’s very difficult to drive to.

It’s even being used by the British Museum to document over a million archaeological finds, because the cities where these archeological finds were originally produced often are now just fields. History has moved on.

We’re embedding this in as many places as we possibly can, and people actually quite like this. We’ve been given quite a few awards, which is always gratifying.

Really what a three-word address means is that a population which quite often can't get access to or find the resources they need for life – fresh water and sanitation – they places now have an address. These places will never have posts delivered to them, so they will probably never have a formal address.

It means that if you live in area where you have to have an address to register your children for education, you can now do this. It means that, for example, in Mongolia, where they have no postal service and no addressing system, there is one out of the box. It means that you can register for healthcare and have your friends, your family, and your loved ones looked after. Everyone everywhere has a simple three-word address.

Because we're start up, we have to have a mission statement, and our mission statement is: we want to make the world better, more efficient, less frustrating, and safer. Maybe, just maybe, this could also help in the problem of people who don't have addresses being confronted with a domain registration form which demands a real address.

It works online. It works on the web. It works on mobile. Download it if you'd like. It's free to use. Have a play.

Thank you very much.

EBERHARD LISSE: It’s a really nice idea. I like it. It’s just that you can confuse it with domain names, which is a bad idea.

GARY GALE: Yes, you possibly could. However, the words have to be of at least four characters.

UNIDENTIFIED MALE: [inaudible].

GARY GALE: Yes, that’s true, because you now have these crazy top-level domains.

EBERHARD LISSE: The problem is that you don’t know how many names will still be registered, but the existing ones, are they excluded or can they be excluded from the right-hand side?

GARY GALE: I’m not quite sure what you mean there.

EBERHARD LISSE: .shop. Can you exclude the word “shop” from being the third word in your [inaudible]?

GARY GALE: For new languages, yes, that is definitely a possibility. For existing languages, one of the promises that we've made is that, once a language is released, it's not going to change. If it changed, that would mean the offline facility, which is one of the key selling points, will cease to work.

ROBERT MARTIN-LEGENE: Robert from PCH. Well, we all saw the [dots] in the address, which we also see in other parts of software development and stuff like that. My question is, are you going to stick it in DNS?

EBERHARD LISSE: My question to this is, why?

ROBERT MARTIN-LEGENE: [inaudible] DNS.

EBERHARD LISSE: What I understand from it, and I will just defer to him, is it's offline. It's so fast it doesn't need a fast evolution like DNS. Finding my location on my old 5C with four gigabytes of data is instantly – loading that application takes about ten seconds. But finding is quick.

[PATRICIO POBLETE]: How does it help? [Patricio Poblete] from the Chile. Still. How does the process of assigning these three words to these three by three squares work? Have all those squares already been assigned, or is it an ongoing process?

GARY GALE: For each language, we effectively start from scratch. In the first attempts at internationalization that we made, we realized that the critical mass of around 27,000 words that you need in order to roll out a new language was soon going to be exhausted. The more languages you do, the more you realize that. Plus I –

[PATRICIO POBLETE]: Why do you speak about languages rather than regions or countries or whatever?

GARY GALE: There is absolutely no administrative hierarchy imposed on this.

[PATRICIO POBLETE]: When you say “languages,” does that mean that this place has three words in English and also three words in Spanish?

GARY GALE: Yes. And three words in Italian and German.

[PATRICIO POBLETE]: Okay. I understand now. Now, people, being how they are, it's quite likely that somebody will find offensive some of the words that were assigned to his place. Say you assign something that contains the word "steak" to somebody who's a vegetarian or something. Well, people complain.

GARY GALE: We do do careful editorialization to make sure that offensive, profane sexual connotations are removed before we release a language. We also work not with language speakers but with people who are native to the countries that we internationalize for because the cultural nuances are everything. A phrase which could be absolutely innocuous to an English speaker on a translation basis could be the most deeply offensive insult about someone's mother-in-law that we could ever come across. So we try to weed those out.

To be honest, though, there is a saying that you can please some people some of the time. People don't like their street names – let's put it that way – where they exist. So you have to strike a balance.

[PATRICIO POBLETE]: Well, I wish it would, like with the filtering. Even with just restricting yourself to one language – say, Spanish, there are many examples where a world that's perfectly safe in one country that speaks Spanish is a no-no in the country next door.

GARY GALE: That is very true.

EBERHARD LISSE: I want to have Patrick with a remote question first, but to answer – and, Patricia, you can always move three meters to the left.

GARY GALE: And people do.

PATRICK JONES: Patrick Jones with a remote question from Carlos Martinez in Uruguay, who's asking why does the mapping not preserve locality? If you move one square, all the words change.

GARY GALE: That's because each three-meter square has its own address. I don't mean to sound flippant, but it's a little bit like saying, "I'm looking at my latitude and longitude, and when I move, the numbers change. The location has changed." Okay, maybe –

UNIDENITIFIED MALE: [inaudible].

GARY GALE: When we assign the three-word addresses to a square, the process that is done is we start with the centers of population for the countries that that language is either a native language or is a formally-recognized language. Then we move out from that.

We tried originally looking at inferring some form of road, neighborhood, district, city hierarchy. It made the offline footprint far, far, far too big. It also meant that, because these are algorithmically assigned, it ruined our built-in error connection.

Does that answer the question now?

PATRICK JONES: No. What he's asking is: why is an index home raft next to index home toilet? Something like that.

EBERHARD LISSE: So that you don't mix it up too easily.

GARY GALE: So that you don’t mix it up.

PATRICK JONES: So it’s deliberately so you don’t mix it up?

GARY GALE: Yes.

PATRICK JONES: Okay [inaudible].

GARY GALE: Thank you for translating, by the way.

PATRICK JONES: [inaudible].

EBERHARD LISSE: For me as a doctor, this would be interesting in Namibia, which is the second most-sparsely populated country after Mongolia. We have an app where you can push a button and the ambulance comes if you have data. If you live next to my mother-in-law’s place, I can give you the string that leads to the gate of her little place, which I cannot send over data because there is no exception, but which I can sort of send a chart with a telephone into reception and it can even phone the string through and then the ambulance will know exactly where.

So what I like is that obviously is that you have to number 57 trillion squares, and then you take addresses of words in a list to the power of three. What I like so much about it is that this is something that you don't need sophisticated infrastructure or communication servers to communicate over the wire. 911 in America has featured very strongly John Oliver's show because they find it's sometimes difficult to dispatch ambulances in the richest country in the world. That would help.

UNIDENTIFIED MALE: Hi. Well, as somebody who goes to other ICANN meetings – we went to favela in Brazil, and I actually know what they're talking about. By the way, my home place in Kiev called harmless reserved bets, which is not offensive to me, but it can be offensive to others.

I actually have two questions. I see that you already used it Mongolia. That's how I learned about you guys. I wonder what the criteria is used to add the language. For example, would you add Russian, which is not my native language but may be popular? Because I know in Russia there's a saying that there's no roads, just directions.

GARY GALE: We have Russian in beta.

UNIDENTIFIED MALE: Oh, great. Okay. So the criteria for languages – and I guess our TLD can probably think about adding that field as a one way of expressing addresses, although I would not bet on that. And I guess that's why you're here.

My other thing that – and again, to give you an example. I've been to Costa Rican address that was specified as like Dos Palmeras, [inaudible], doscientos norte, [trescientos sueste], stuff like that, basically direction-wise. You know, some places are not flat, so the thing about streets all addresses is that sometimes they specify the height. So how do you address the cubic, the height of the place?

GARY GALE: Z axis is a notoriously difficult problem to solve.

UNIDENTIFIED MALE: Yeah, I know it's difficult, but [inaudible]?

GARY GALE: I think the thing to understand here is that we're not saying that we are the new addressing system for the entire world. Tremble ye mortal postcodes!

UNIDENTIFIED MALE: So only for the [inaudible].

GARY GALE: No, but we are there as an enabler. For places that don't have an address, it works out of the box. For things that will never had an address, it works out the box. For places that already have an existing addressing infrastructure, we can provide additional insight. There's nothing to stop you – and indeed, people in Africa are doing this. They're saying, "Second floor. Building Name. Three-word address."

UNIDENTIFIED MALE: Yeah. Sure. Like I said, you're good, and it may work. I just think that maybe the benefit of your system [inaudible]. So that's just something to think about.

GARY GALE: I can definitely say that elevation on it is something which is on our backlog, which some of the more clever people are looking at.

UNIDENTIFIED MALE: Okay, thanks.

JAY DALEY: Hi. Jay Dailey from .nz. I've already seen one great advantage of this. Living in a relatively not densely populated country, my block of land is two-and-a-half thousand square meters. So if I use one of these words and I get banned on the system, I've got plenty more I can keep on using.

I have two questions for you. First of all, is the three by three square always horizontal or does it shift on the elevation model? Basically, does it match the contours of the landscape?

GARY GALE: In effect, you're talking about the problems that mapmakers have, which is trying to unwrap the oblate spheroid, which is the world, and transform it into a 2D representation.

We draw our squares on the surface of the earth according to think WGS84, which is the world geodetic data rationalized in 1984. So those squares do not reflect the terrain.

JAY DALEY: Okay. So they're always horizontal.

GARY GALE: And they are equal all the way up to the poles, at which point in time you do get a slight degree of distortion.

JAY DALEY: Okay. The other thing is, do you have list of words that are used in the third position that you could make available to some of us so that we could then see the conflicts that exist with top-level domains? Or have you done that work yourself?

GARY GALE: No, we haven't done that work. For an organization such as yourself, I think that would be a really interesting thing to do. My e-mail address is there. Somebody please get in touch and tell me who I should talk to.

JAY DALEY: Okay. Great. Thanks.

EBERHARD LISSE: Robert?

ROBERT MARTIN-LEGENE: Robert from PCH again. Well, I think, taking the registration data, you're asking registries to start accepting these as some alternative for some locations [of what]?

GARY GALE: I'm not asking. The reason I'm here is because I was asked because I think that there was a thought that maybe this could be useful to the problems of incomplete addresses that you have in the registration records. I would say to you as an industry as the same as we say to any other industry sector that we come into contact with. I'm sure there's some way that we can work together, but by no means are we mandating this or wielding a club saying, "You shall do this!" History, and especially the history of the Internet, has proven that, when we all work together for a common goal, things tend to get done a lot quicker. That's literally the reason I'm here.

ROBERT MARTIN-LEGENE: I think most registries wouldn't mind which format an address comes in. What do they want to do with it? They want to be able to send a letter to someone or tell the police where to go. They don't care if it's called the blue house or if it's called whatever.

About not finding an address offensive, my phone right now has an address that does not really mean anything about me personally, but it's "restore bedroom skunks." I thought it was pretty funny. I don't smell, by the way.

EBERHARD LISSE: Three meters to the right.

NIGEL ROBERTS: Thank you. My name is Nigel Roberts from the Guernsey registry. I had the advantage of looking at this a couple days ago. I got a hint that this was being presented today. I like it a lot.

GARY GALE: Thank you.

NIGEL ROBERTS: But a couple of comments and observations and questions, really. First of all, we're getting quite hung up about the fact that the rightmost word could be in the DNS or something like that. But it strikes me that the full stops between the three words is pure convention, nothing else.

GARY GALE: It's an opaque string. Absolutely.

NIGEL ROBERTS: So there's no reason you couldn't put a slash or a hyphen or an @ sign or whatever anybody likes. It's the three words that matter.

GARY GALE: Yes and no. Originally, our search facility allowed pretty much any punctuation character in between the three words, and it meant that people were finding it very, very hard to be able to parse out from free-format text fields whether a three-word address was there. Given that the full stop exists in almost all character sets and is used as a delimiter is almost all scripts, that's we standardized on that.

NIGEL ROBERTS: So you could say that's a standard, but you could be tolerant of other expressions of it, just the way that phone numbers are sometimes [mangled].

GARY GALE: Yeah. Absolutely.

NIGEL ROBERTS: Thank you.

EBERHARD LISSE: I am getting more concerned about him making his plan than he is because it's such an interesting debate, but I have to –

GARY GALE: No, I do have to run.

EBERHARD LISSE: I'm cutting the list short now because he has to get the plan at 7:00. Thank you very much for coming. This was one of the more interesting presentations I've heard in a while.

We have had presentations that were totally out of our normal sphere before, so I think he deserves a good hand of applause.

GARY GALE: [inaudible] absolutely fascinating.

EBERHARD LISSE: Okay. Then we have our last presentation, which is more traditional: Analysis of a Botnet Campaign. As I said, we are not strapped for time on this presentation. We were just constricted by him having to reach a plane.

UNIDENTIFIED MALE: [inaudible].

EBERHARD LISSE: Where is he going?

UNIDENTIFIED MALE: You scared him off [inaudible].

EBERHARD LISSE: No, he's looking for the microphone, I think. He wants to probably stand.

UNIDENTIFIED MALE: [inaudible].

ANDY SETTLE: One, two. Hello? Can you hear me?

UNIDENTIFIED MALE: Yes.

ANDY SETTLE: I'm very fidgety when I talk, so I'm going to be moving around. I don't want to sit down. I've been sat down all afternoon.

Is this okay for the sound, guys?

UNIDENTIFIED MALE: Yeah.

ANDY SETTLE: Awesome. Brilliant. I’ve been told not to wander around too much because there’s filming going on.

Hi, my name is Andy Settle. I’m from Forcepoint. I promised a number of people under the threat of pain that I’m not going to do any marketing or sales, so go look at what Forcepoint is. We used to be called Websense.

Okay. Failure on the first… So talk amongst yourselves.

UNIDENTIFIED MALE: [inaudible] myself?

ANDY SETTLE: Yeah, smashing. I’m going to give a brief introduction about myself – I’m not going to read that to you – but then mainly health warnings because there’s going to be quite a few opinions and conjecture in here. They are all my own. They are not necessarily the opinions or statements from my employers because there’s a couple of employers in there.

My connection goes back to ’98 with the GG and J ccTLDs, where I worked with my colleague over here, Nigel, back in ‘98, at the same time when ICANN was formed.

I'll move quickly on. I'm going to talk to you about a botnet that we started looking at about six months ago. This was part of my team, the special investigation team. We were looking at different aspects of thing that were just a little bit too weird, a little too complicated, a little bit strange for the rest of the security labs in which we work.

It all ended up in a whitepaper, which you can download. It also ended up – I love the artwork that our design guys have done, because they've translated at least the Korean and the Japanese infographics for those areas, and you'll see why in a moment why it's been translated into Korean and Japanese.

About November 2014, Kaspersky brought a report out called Darkhotel. The Darkhotel is about a campaign against people that frequent hotels that are predominantly used by business people. It was a technique by which the Wi-Fi networks were infiltrated and then started attacking and infiltrating the machines that were connected onto it. So, a great report. Very interesting. Got to commend them on it.

Just like every security researcher in the world, we go and read everybody else papers. The first thing we do is go and look for whether it matches what we already know. Does it aid our understanding of what we're already looking at, and are there any gaps in it?

This might not work. This is a PowerPoint presentation. This next slide – sorry. If you could just bear with me. There's things missing off there which are actually quite important. Can we switch the PDF, please? Sorry about this.

So we took the Darkhotel report and we took other reports that were actually produced at the time, and we started looking at the telemetry that we see when we start looking at botnets, another malicious activity –

Are we good? Yeah. Okay. Let me just flip forward. Are you okay with that in the back? Apologies for the small text. The things here are for illustrative reasons anyway.

So we heard the term "pivoting" before. It's been mentioned today. There are IP addresses in that report. We take those. We go and look forward in the DNS, and then we go and reverse the [banded] DNS techniques. Of course, then we start used the Passive DNS providers and Passive DNS information, which is the historical information that's available on those entries.

The example we give here is what did [pic3].moo, .com resolve to in the past? When we get those IPs, what did they resolve to in the past as well? We build a bigger picture up.

From those, we go and check in some cases. In this case, those web servers we checked against to see whether they were running web servers on there. On those web servers, we then go and check for known bad areas where botnets or other malicious code – we heard stuff about the other campaigns earlier – we've got a huge database of these command-and-control systems.

So with the Passive DNS information and with all this information here, we then go and drill down into the web servers and go and look for what we could call evil.

In the case here – I'll point this out if you can't see it in the back – if you go and look for a file called near.jpg, it just sounds like a standard image, which it normally is. If you Google for it, you get one of these pictures up here. Nothing really exciting about that so far. But actually, when you go and look on [pic3].moo that we saw earlier, you'll find a near.jpg. That's all looking okay so far.

Now, the story I'm telling you sounds very linear. For everything I say, there are probably 100 rabbit holes we went down. This was probably Rabbit Hole #57, but we struck gold because we noticed that that file was 451 meg in size. That's a pretty large JPEG file, especially to be in an image directory, which usually has the back and forward icons and so on and so forth.

So we went to have a look at it because that's a strange image. It wasn't a JPEG file at all. When we go and use our favorite LINUX command, we're told it's a SQLite database. From that SQLite database, we can run a PySQLite. It's the lightest SQL database system in the world, arguably. And we can see in there that there is a database in there. There's a database schema, and there's tables called Child and there's tables called History. There are other ones, but I'm just going to talk about those for the purpose of this talk.

Of those two, the History ones is pretty self-explanatory when you hear what I say, but the Child one was really fascinating. They got unique IDs in there. They don't want to track people by their IP address, which they do store the IP address. That's PIP – Public IP. They're stored by unique identifiers.

There's a version [number] there. There's something pretty serious about this. Someone's doing version number tracking and wants to make sure that they know which malware is on what machine. There's an info table, and I'll tell you that in the next slide. Then there's when the information was updated. There's a mine of information in there that we got ahold of.

The unique ID clearly means that they know that they're tracking people whose IP addresses may change. That kind of chimed back in with the Darkhotel pattern because that was clearly some set of people that were being moved around because they were using public Wi-Fi.

Well, let's have a look at one of those columns from the database. Don't worry about not being able to read it. That here is a dump of the current processes. They're interested in the processes that are being run on the machine. They're also interested in file names. Down here we have on the right-hand side the file names that we saw within the directory listings that they did.

The top – it's a good job. It's blurred actually, and probably nobody can read the Japanese in there anyway. They're pornography. Equally, there's copyright infringement in here. You can see Hotel Transylvania as one of the file names, and then a couple of others.

Equally interesting is personal information. There are just some of the examples that we took out. There were images of people's passports in there because people were scanning their passports in. People do that. When the malware gets on there, these guys get to see that. This is all fairly standard, run-of-the-mill malware so far. We're not surprised. We unfortunately can't in my team have to investigate this every day because there's so much of it. We just write rules and then we move on.

But then we spotted this. If you can't see that at the back, I've highlighted some of the words. There are four letters – D, P, R, K. there was a set of victims in there that all had file names that references those four letters – D, P, R, K (Democratic People's Republic of Korea).

Now, we know that there is a challenge with connectivity in North Korea, and we know that there is a challenge with some of the threat sources that we see that match onto this in relation to that. I'm not doing any attribution here at all. What I'm saying that there was a set – and I do mean a set; I'll tell you what I mean by "sets" in a second – that were all referencing either Pyongyang or DPRK as a term.

Let's just look at what that was doing. Now, some of you that have seen this kind of bad behavior before may recognize that set of commands, not necessarily that identical set. That set is very common in the reconnaissance phase of someone that's more than interested in student credit card details. It's done as reconnaissance. It's done to exfiltrate configuration information, and that first passive data that you can get off of it to machine.

I collect these, and I've got a canonical list of all of these because it works for me where I can actually see the fingerprint of the attack.

Now, in this case, you can see there's a highlighted command in there. Those victims that I said were all relating to DPRK and Pyongyang did not have that command in there. But every one of the others – the thousands, if not tens of thousands – had that command in there. So there has been a slight to that TTP, the Tactics, Techniques, and Procedures, that are used by them. But they are almost identical. Looking through our fingerprints of those, we couldn't find that.

Let's just jump ahead now. We didn't just find one of these near.jpgs; we found a set of them. Again, it gets very unusual when you see the ones that we've had to redact for some security reasons. When you see the number of victims in there, there were only 17, at most 10.

So we're now looking at a botnet campaign that's got ten command-and-control servers hitting approximately 19,000 different victims. That's a conservative estimate. We think it's far, far higher than that, but we got to be careful with our figures. And then in one case, one command-and-control server was just looking at ten people.

So we tabulated that. This is from the report, so you can see this data as well. We like drawing things on maps. Those are where the command-and-control servers were. So we reached out to the certs. We reached out to those authority over there, and we start sharing that information with them, saying, "Hey, we found these things." Some of this falls on deaf ears, and sometimes they reach back to us.

The colors are the colors of each one of those ten. You can see that there's a lot in Thailand, one or two in Malaysia. The Singapore – and you can see it's all centered around that area.

Well, actually, that was the first thing that we start looking over: [to] APAC region.

I want to show you the graph now of the victim timeline. These are number of victims per each one of those sets. When I mentioned the 17, that very small number, that's that red line there. We refer to that as Red Raccoon. I will talk about, again, Red Raccoon in a moment.

It got quite complicated. You can see things coming and going. You can see victims disappearing and reappearing. At one point, what you can't see, unfortunately because of this graph, is that the Black [Safarists] command-and-control server gets attacked one Friday night. It gets attacked by a carding scam team that wanted to do an attack themselves on Spanish banking online sites.

Monday morning comes – Monday morning in our terms, so I'm not attributing when what time zone this all happens in – and bingo! The go and find out their command-and-control server had been exploited. They kicked them out and they started all over again.

In one case, they didn't manage to do that, and that's why some disappear, and that's why some go offline for a short period of time. So you can see there's a little bit of a tit for tat going on with the threat actors.

Targeting itself. When you start looking, because we can see some of the system info because we have access to that near.jpg file that's got that information about the victims, we can start doing some really interesting analysis on there.

These are meant to be three pie charts, but for the sake of being able to put a slide up, you have to trust me that the third pie chart looks exactly the same. It's victims per country, victims per language, and victims per time zone. So it fits a demographic of targeting, the time zone of Korea and Japan, the languages of Korean and Japanese, and equally the time zone as well. These victims sat over in Japan and Korea, and I'll just illustrate that with these maps very quickly.

You can see that quite a bit on the west coast. You can see it's been on the east coast. You see a scattering around Europe. There's a side story which I just want to bring up very quickly. If I can point it out for you, it's the data center in the middle of Algeria. There isn't a data center in the middle of Algeria.

If anyone is familiar with the MaxMind dataset, there was a news item that was in BBC News a few months ago, which was about a farmstead in Kansas that keeps getting debt collectors going, “We know that you’ve been running a scam from this location. We need our money back.” The problem is, it’s – my American geography is terrible, wherever it is, Kansas place or whatever – when MaxMind did not know where the IP address when you geolocate the IP, they put it in the middle of where they do know it is. So if they know if it’s Kansas, they put it in the middle of Kansas. If it’s Algeria, they put in the middle of Algeria. So we’ve got to be careful. What we can say that is it’s Algeria, but not necessarily that location in Algeria.

Equally, you get the same from Mongolia. There’s [a] temple, if I remember right. We went to look at that on Google Earth. There’s not enough buildings for that number of IP addresses there.

But with that in mind, just to come back to the mainline story, you can see that most of the victims are hit out of Korea and Japan.

How’s it done? Poison bit torrents. There is a nice build to this one. Unfortunately you see Mario gone. You see at the top that we got files up here that are sitting on bit torrent sites to be clicked through on. Someone goes and downloads them. In this case – Mario, [I’m going to] mention this because this is the one that was passed to us by Microsoft. They found it. We didn’t find this piece of malware. It's not a bad game, by the way. We ran it in sandbox. We didn't run it on our bare-metal machine. It's Angry Birds with a Mario theme.

When you run it, you get compromised by the Stage 1 malware. In there, we've got a PS2 emulator for PlayStation games that you can run on Windows machines. You've got some codec for, if you want to watch this movie that I've just gone and pirated, you have to download this codec to get the audio to work correctly. And you've got the WinRARs of the world, which are the unpackers for, when you're downloading this cracked software, you want it to unpack it, well, here, download this as well. That had the malware in there as well.

I can't count from here. One, two, three, four, five, six, seven, eight, nine, ten, eleven. In the interest of honesty, there are more pieces of malware out there for this botnet than we ever found because we know that from the version numbers. The version numbers relate to what was seeded with the malware in the first place.

A part of frustration for us is that we didn't find all the pieces of malware, but we found out how it was actually getting out there in the first place and getting onto these machines.

Well, how did it, because everybody uses legitimate versions of Windows and everybody uses antivirus software. Well, actually, no, not at all, because over 50% of these victims were actually running counterfeit software, specifically Windows in this case, but other counterfeit software as well.

The good news – and it is a good news story – is that the number of corporate victims in there was tiny. You can see there that we calculated this. We got to be careful with this one. There are probably more, but we could spot 153 unique corporates in there.

Truth be known, there were two customers of ours. They're really interesting stories in their own right. The first one are two road warriors for a U.S. industrial company. They were always on the road.

They have great security, obviously, but unfortunately they weren't configured to use that security when they connected to their cable connections at home. If they didn't phone-dial into their VPN, they didn't get afforded the protection that would stop this from happening. So they both got infected from home, and then when they went into the corporate environment, that's how we found them out, because they went and beaconed home again and then the alert went up on our sensors and we could tell the company what was happening.

The second, which was also interesting, was a mergers and acquisition. It's one of my pet subjects. When you take companies and you can join them, you don't know what you're conjoining those two companies with. In this case, it was a large civil engineering company that went and bought a company from the Far East. The day after they connected their two networks together, we spot a beacon come up on one of our sensors that was a piece of [inaudible] malware in their infrastructure.

They were both fairly thankful for us telling them that we'd found that.

I think I've summarize all those already. The [C2] database, in the first instance, was an absolute godsend. Being able to see inside their command-and-control system allowed us the level of insight that we had never seen before.

But it gets more interesting when you get down to second-stage. So you got a piece of malware on a machine. What are you going to do next? Well, we spotted that the malware was downloading a PNG, and we thought, "It's a PNG. It's going to be another SQLite database." Well, in this case, if you look at the XY coordinates of that PNG, that's the height and the width of that image. That's not, again, your average image for any known system that we've come across. That is not an image at all. It looks and it feels and it sniffs like an image, so the content checkers think it's an image. But actually, it's a malwareembedded file with a PNG header, and it's steganography. The malware is sitting interweaved inside this image. If you display it, it probably blows your machine up because it's a huge image in the first place.

I'm going to skip those in out of order. It's got compression in there. It does AV detection. Again, for the sake of honesty, it checks for one AV engine and one AV engine alone, BitDefender. If it finds it, it deletes itself and disappears. We don't know. BitDefender [don't]. We had a quick chat to BitDefender and they went, "I dunno."

Why malware would run away when they saw it? There's a few ideas. One of the thing that is my role in life is to look for answers from people like you guys. So if there's anything that I go "I don't know" to, it means that I'm looking for help because we don't have all the answers. The more we fill in this picture and the more we do the dots together and then color it all in, the better understanding we can have.

But let's just go back to bad RC4 encryption. The first rule of writing your own crypto is don't write your own crypto. That's the second, the third, the fourth, and carry on to N rule. But if you're going to, copy it, and if you're going to copy it, don't modify it and then break it because the reason why we can do all the rest – and I can talk about all the rest – is they took RC4, which isn't brilliant but is workable, and they modified it so that we were able to reverse engineer all the way down to the opcodes and then back again to C. So we can bring back the source code in effect through our reverse engineering. We take pride in our reverse engineering. We're a little bit obsessive about it.

That was the weakest link. You will hear me say that time and again: rule number one is don't write your own crypto.

With that in mind, we started looking at how that second part behaved. I will spare you some of those details anyway, but this was an interesting story because I was proofreading our final whitepaper the night before we published it and – top tip number two: if you've got a command-and-control server, don't call it command-and-control. That's pretty obvious as well. Now, these guys are not stupid. We laugh or we laughed and we still laugh to this day and it makes me smile, but it is a giveaway.

Then I went and saw that because I was giggling about that and then I read that and I thought, "Has anybody actually gone and Googled that from my team to make sure that it's not something unusual or we've missed something?" So I did that.

Now, there should be a very sexy build on this, and it doesn't work, unfortunately. So I put that into Google, and guess what? I get four results. You get a bit excited when that happens when you're doing malware analysis.

The top left-hand box shows you a header file. It's a C++ header file. Now, if it's a C++ header file that's got a class in there, you know it's going to have publics. If you take one of those publics and then you go plug that back into a Google search to find any more code that uses it, that's the bottom left-hand corner.

From there, we've got a library of code, and you can see there's already Korean text in there. Then you get to this blog post written in 2011. It's not the most commonly used blogging website in Korea – South Korea of course, and it was written in 2011. But it had all the code that, when compiled, generated a component within the malware. We're not saying that we do know whose site this is and we've redacted for him. We know whose site it is. We're not saying he wrote the malware, but his code was definitely in the malware.

And, crucially, all the source code was written in, as I said, in C++ with lots of Korean comments. Not English comments. So if you're going to read someone else's code, unless you are very, very, very bizarre, and you want to set a false flag – I would defy anyone – you're going to be native Korean speaker, because if you're not, you will have done it in your own language, or you'll have gone to what everyone else would have done: gone to

MSDN because this is a standard design pattern that MSDN documents four or five times in lots of different ways and you can download the examples.

But, no, this person used a piece of code from 2011 that was heavily commented in Korean. We can, with a reasonable level of certainty, say that whoever wrote that malware was a Korean speaker.

Now, this is where I got into a bit of trouble with the Tokyo Metropolitan Police Department because they said, “Why did you call it Red Raccoon?” “Well, because we take random names and we take a set of colors.” We used to use adjective and name for our project names. It’s completely randomized. But in this case, we used colors so that we could plot them on graphs. It just happened that Red Raccoon came up.

Now, truth be known, the Raccoon was fairly random, but we should have moved off it. When we conjoined Red Raccoon, we didn’t realize that it was this concept. It’s not a real thing because he’s not red and he’s not a raccoon. He’s a bear. In Japan and in China there is this idea of a red raccoon. It’s not. As I said, it’s a bear.

Of course, the Japanese then go, “Well, this is a reference to communism.” “No, no, no, no. It’s just a random color, and we regretted that." So that is it. It is a random color and a random name.

However, this refers to all the people relate to Pyongyang. As I said, there's 17 of them; scientists, academics, engineering company employees, government employees from other countries with interests in whatever manner possible with North Korea and NGOs.

There was a connection between two of those. Everyone we saw had a connection with the big red circle in the middle. But one of the victims had file names – well, they were actually linked in this case. They were linked to YouTube videos. There was a video file on the directory listing of the same name, which was a movie.

We went and looked on YouTube and we saw what that was, and it's a guy cycling around North Korea on a bicycle and with a GoPro on his handlebars. There's quite a few of these videos. They're gone now.

At one point in time, we were looking through the video and then we saw him ride back into his compound where he lived in North Korea in Pyongyang. Guess what? We saw a white vehicle. I think it might have been a Toyota Hilux, but it's that kind of vehicle, with the name of the NGO written on the side of it.

Now, we rewind again back to Darkhotel and we go "Bingo!" These guys are being targeted. We can't say with any certainty that these are being targeted in the same way that these two were, but, with a reasonable level of certainty again, it was the Wi-Fi network or whatever Internet connection they were being provided that had been compromised in whatever way. That's as much attribution as you're going to get from me.

Here's another strange story. John Underhill. Never met John Underhill before. John Underhill is a really nice chap, and what we did is we found a bit of code – I'll spare you how we did this; this was a lot of work to find – Some of you will be familiar with secure sanitization and disposal, forensically sound delete algorithms that file lots of zeros at a disk and then file lots of all ones and then fire random numbers and then do it a number of times.

There was a unique way in which this malware was actually doing it. It was annoying. I nearly swore then. Yeah, it was jolly annoying that we couldn't recognize the algorithm until we spotted that it looked very similar to a U.S. Air Force-proposed delete algorithm.

Well, we looked a bit further and we came up [inaudible] because we spotted John Underhill's code, called Secure File Shredder. Why am I laboring this point? It's because there were bugs in his code, which you can see from this line here that they actually fixed that big. But what they didn't do is they didn't fix that one. There was a coding error in John's code.

Now, that's not as strange as the next bit. The strange thing was that I wrote an e-mail to John – didn't know him – and said, "Hey. We're doing a malware investigation. We've seen some secure delete code, and we think it's your code. Could you send us the source code for your C version of Secure File Shredder? Because the one on the Internet is written in C#." He wrote back and said, "I never wrote this in C."

We had that Scooby Doo moment when you think, "Hmm. That doesn't compute." Now, the reason that doesn't compute is that someone must have sat down with John's code and line by line translated it into C so that it could go into their code. Then it ended up in this piece of malware, and the copied his programming errors.

John was really nice about it. He was actually quite upset that the algorithm for his code was sitting in some malware somewhere. But there's a level of reuse that's just getting a little bit weird.

If you think there's a big punchline to this, I don't want you to be disappointed. It's a story which we know the picture of, but we don't know what the big picture is. But we do know that someone is concertedly trying to, in this case, delete off a file system a piece of malware.

But it gets better because you saw that [tank] that I Googled for. We found the C++ code. In this case, we've got a command-and-control system which uses DNS. This is one of the conversations that I had with Nigel about the number of DNS command-and-control channels that we see.

In this case, we've got C names that are using MAC address and victim names that are beaconing out, going, "Hello. I'm here. Hello, I'm here," every three minutes as a C name. We've seen other ones.

There was, by the way, mention of DGAs, Domain Generation Algorithms. We're pretty into that as well. We do a lot of Doman Generation Algorithm reverse engineering, and that's what we thought this was. We thought this was a DGA. No, it's not. It's an encoding mechanism for the victim to say, "Hello. I'm here."

This is the one that fries my brain because I have never yet – fingers crossed I'm going to meet someone in a few seconds that's seen this before. There is a library called UDT, UPD-based Data Transfer. You can see where the stars are on the bottom left-hand corner that the community that uses it are the supercomputing people. These routines are there to actually get large amounts of data reliably and efficiently across between machines. That's in this piece of code as well.

You only want that if you're going to be moving a lot of data around a lot of nodes. One hypothesis is, is it to get through content checkers to go: it's a recognized piece of code? No? Because none of us have ever seen that code before. Is it a malicious pattern that is trying to [inaudible]? Well, possibly. But the level of effort these guys have gone to write this is phenomenal.

The good news is we've made lots of new friends. So I'm speaking to UK, Dutch, Japanese, Korean, Canadian, and a couple of other certs that we've just not put on this slide. Law enforcement. The Japanese have been really, really good. Their high-tech guys have gotten into a real discussion with us. I've got to actually say that you to the Korean cert guys as well because they proofread out report.

Microsoft, again, and there's a couple of other vendors which I'm allowed to mention because the boss goes, "You can't mention competitors on this." But at our level, at the researcher level, we talk to each other all the time. Microsoft is really good because they found a piece of malware and they named it JAKU, so they're acknowledging that this was our discovery. Then they go and give us some of their research they did where they went, “Just use it.” So I’ve got to say thank you to those guys.

Those headlines are resilient command-and-control channels, three forms of command-and-control channel. We’ve seen a [failback] but we’ve never seen three, and then one using a protocol that’s primarily aimed at supercomputing.

We did the maths, by the way. I’ll just skip to the next slide. We did the maths on this. The number of computers in this botnet or not on the supercomputing level. The supercomputing is just exponentially higher than the amount of power on here.

But arguably, if you’re looking for a piece of infrastructure or you’re strapped for infrastructure and you want a lot processing power, not necessarily to do bad things with but to actually do whatever you want, then this botnet was probably ideal. Lots of unwitting victims just offering their processing cycles to whoever was controlling it. That’s the report on the left-hand side.

There’s an, for the hard of reading, infographic in English and in Korean and in Japanese. There is an Easter egg in there. We write reports that we would want to read ourselves. So far there’s only three people that found those Easter eggs, and one of them was this morning. When I received an e-mail this morning, it made me smile because it’s got a code red in there.

It's very geeky to report, but we've really enjoyed reading it, and we've got some really massively good feedback on our research on this one.

I'll just reiterate. If this does raise any questions, if you are interested in going, "Actually, I could answer that in that way," then please, please, please, we're here to share that information.

Thank you very much for your time.

EBERHARD LISSE: Any questions? Well, the last presentation of the day is over. I'm trying to entertain. Thank you very much. I liked the presentation very much. Jacques will now give us –

UNIDENTIFIED MALE: [inaudible].

EBERHARD LISSE: And he took the tie off actually. Yes. Thank you very much again. Give him another hand.

As usual, somebody will close the procedure, and this time it's Jacques Latour.

JACQUES LATOUR: Hi. I'm Jacques. I'm with .ca, and I'm part of the Program Committee. I hope you enjoyed this half-day technical workshop.

I'll cover quickly what happened in the presentation today, and then some of my overview. With .fi [inaudible] went over their – I'm interested to see that they're changing their vision to adapt to the new world. They're going to deliver new services and try to compete with the changing world. So that was interesting.

Having .sc, .nu zone file publically available is something interesting. There's a lot of us that perhaps would want to do that. But I'd be curious to see if there's a direct correlation to abuse. If not, then I think more of us would be willing to make our zone file available. So that's something that maybe in the next session we can have more presentation around. That'd be interesting.

Jay did a nice presentation on his portal. There was talk about having make this offer available to all. That'd be something really interesting. We'd have a common way of interacting with registrars.

We had a nice presentation on Zonemaster. That's something we use CIRA. We had some issues branding it with .ca, but over time I think it's going to be useful for people to use that.

Roy did a nice presentation on the, as usual, history of the Internet. You went into the traffic analysis. That was interesting.

SDF, the Secure Domain Foundation. I think this is the second version of that. I'm seeing some value of having this in line with the registry to validate information.

CoCCA is implementing this. I think this is going to be important over time to register over time to measure the quality of the registration.

My address is "coder toolbar vouche," so that was interesting; what the three words were. I'm not sure how or what's going to happen with that in the future, but I think it's a nice way of doing it. The issue is that one house can have 20, 30, 50 addresses different, so we need to address the multiplicity of addresses for one address, and also the collision with the domain name.

Then we had this last presentation about JAKU malware analysis. That was interesting.

So that was the day. If you have stuff you like, stuff you don't like, let us know, the Program Committee, so that at the next Tech Day we have always a better session. I hope to see you in Hyderabad.

Thank you.

EBERHARD LISSE: And without wishing to offend anyone, I call it Puerto Abad.

[END OF TRANSCRIPTION]