ICANN74 | Policy Forum – Tech Day (1 of 3)
Monday, June 13, 2022 – 13:15 to 14:30 AMS

EBERHARD LISSE: Are you going to say some introductory remarks, Kim or Kathy?

KATHY SCHNITT: Hello and welcome to the Tech Day Part One of Three. Please note this session is being recorded and is governed by the Expected Standards of Behavior. During this session, questions or comments submitted in chat will be read aloud, if put in the proper form, as noted in chat. Taking part today via audio, if you are remote, please wait until you're called upon and unmute your microphone. For those of you that are in the main room or in the secondary room, please raise your hand in Zoom. When called upon, unmute your table mic and speak. Please remember to turn your microphone off.

For the benefit of other participants, please state your name for the record and speak at a reasonable pace. You may access all available features for this session in the Zoom toolbar. And with that, I'm happy to turn the floor over to Eberhard.

EBERHARD LISSE: Thank you very much. Welcome from lovely Swakopmund on the Coast of Namibia where I am self-isolating because almost my

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.

whole family got tested positive. I have very mild symptoms but haven't tested yet because I wanted to see what happens. I will test probably tomorrow. So that means if you see me sometimes put a mask on, it's not that I want to share your pain. It's just that one of my family members is going to walk by.

As usual, we have a mixed bag, if I can even fire this up. Hang on. My computer doesn't want to do what I want to do. Okay. So as usual, we have a nice little mixed bag with priority or the main presentations coming from Europe, given where the meeting is held. But we have also got something from outside. We'll start with Byrony Hill—I hope I say this correctly—who is from Nominet. And they have got a tool that looks for domain registrations that become malicious after a few while and see what we can do about it. Then Hugo Salgado is going to speak about ZONEMD. That's a new DNS resource record. They have got some experience in .CL so he can spend a few minutes and explain how that works.

Then we have the host presentation. Moritz Müller will talk about what SIDN is doing again this time, this year. As you know, we always have a host presentation where the local organizer or the ccTLD manager of the country we are in gives a little presentation of a topic of their choice. They discovered a bug in Google DNSSEC to my understanding. So we'll talk about that and a little bit else.

Then we have got Roy Arends, who many of you know. Please, everybody mute themselves who is not speaking. Can you please mute yourself? Can somebody please mute? Thank you. Roy Arends has done a Root Hints Survey. He will talk more about it. It's a bit complicated but it's interesting to see what tools he used and what research they did. And he always comes up with interesting stuff.

Then Pieter Robberechts is a PhD candidate at the University of Leuven and he's applying some machine learning versus DNS abuse. I read an article yesterday about somebody getting fired from one of the big companies for saying that one of his programs has become sentient. So machine learning is going to come into the future. And it might be interesting to hear what we can do about that.

Then we all know about Zonemaster. We have heard a few presentations. It has been updated recently. I saw a new version coming out yesterday or day before yesterday. Mats will give us a short update remotely. Hang on. I don't want to cough into the microphone.

And then Brett Carr will speak about now Nomient, who have about 50 gTLDs running, how they transition in and out gTLDs in [tuple], the programming language for old people. If you do it more than twice, you should provide a program. So it's probably a cool idea to see, if only to write down what you need to do so that you have got all the things written down. But what you can automate, you should automate.

And then Jordi Iparraguirre, who works at .EU, is going to talk a little bit about data sharing between .EU and .BE and .DK. Since these are all European countries, you can imagine that it's not only a technical issue. It's also a severe GDPR-influenced issue. So it's quite interesting to hear how to do this in a way that is GDPR compliant because we all might be conflicted or concerned with. Some of us like to share data. Some of us don't. But if we do, we must be—if there is a European resident involved, be compliant.

And then, as usual, one of the Tech Day members is doing a short wrap-up of 10 minutes. This time, it's Cristian Hesselman. In between each session, we have mandated breaks. I am no in favor of breaks, generally. But in particular, this time, because they apparently need to clean the room, we abide without any grumbling and carry on after the break is finished.

That said, I took two minutes more than I thought I would Bryony Hill, you have the floor. Please somebody share her presentation or you're on your own.

BRYONY HILL: I'll share my own. Thank you. Thank you very much for the introduction. I'm just going to share … This is new to me. Desktop two. I'm guessing you can't see my presentation yet. One second.

KATHY SCHNITT: If you want, Kim can share for you.

BRYONY HILL: Could you share for me? Apologies. I think Zoom's not quite working like I was expecting.

KATHY SCHNITT: No problem. Just say "next slide" when you're ready for her.

BRYONY HILL: Lovely. Thank you. So my name is Bryony Hill. I'm a data scientist at Nominet, so the .UK domain registry. And I'm talking today about DomainWatch, which is an initiative that we started about three years ago to look at blocking phishing domains at registration . Can I have the next slide, please? Thank you.

So yes. I'm a data scientist at Nominet. So I'm interested in machine learning models. And the sorts of things I've done in the past are around classifying the use of domains, so seeing what's at the end of the domain, the website, and trying to understand what industry might be behind it. And also, modeling retention of domains was another big project that I worked on.

I also want to just give a bit of a background around the kind of things that we have historically done for DNS abuse. We've got all the routes for suspension. In particular, we get a lot of suspensions through law enforcement agencies in the UK. So they will let us know if there are domains that they feel should be suspended.

And in recent years, we started putting landing pages on some suspended domains so that if a normal user reaches a domain that's been suspended, they can get more information as to why it's been suspended and a link through to that law enforcement agency to find out more about the kind of laws that have been broken. And then, around that, we've looked at the traffic that suspended domains get to try and understand the impact of suspending domains.

We've also historically done reporting on illegal terms in domain names. So these are very dubious terms that shouldn't really appear in domain names. And we report those and monitor those.

Another piece of work is one called Domain Health. So this is around looking at registrars and seeing how much domain abuse they have on their portfolios and ranking registrars against one another. We measure it in terms of what we can see on feeds and looking on what's on a registrar's portfolio.

Something that I'm currently looking at as well is benchmarking technical abuse—so trying to understand how much there is on .UK and looking at how we can reduce that. Could I have the next slide, please? Thank you.

So this presentation is around DomainWatch. The aim of DomainWatch is to make .UK a safer registry by suspending those domains that have been registered for abuse at the point of registration. So by registered for abuse, I'm talking about not compromised domains, maliciously registered domains. And in particular, phishing domains are the ones that we're looking at.

So I 've got a few examples there of the kind of domains that we've recently spotted. Verifypaypal-id, that's not the kind of domain that you'd ever register for a legitimate reason. You also get typo-squatting. So there's Facebook with a typo in there. Sometimes there's no brand but it's a series of words that clearly look suspicious. Account-support center suggests that might be part of a malicious registration, or sometimes you get a www attached to the beginning of a domain. So basically typo squatting again on wwwicloud.co.uk. That's the aim of DomainWatch. Could I have the next slide, please?

Thank you. So this is a potted history of what we've done over the last few years. In 2018, we began by using Netcraft, who are a phishing feed provider. And they had an API that we were using. So whenever a new domain was registered, we would query their API and they would give back a score. So they would take this domain name and give it a score out of 10 or something. And depending on what that score was, we would then have a look at it and suspend it. Now, this was reasonably expensive and the accuracy wasn't quite as high as we'd like.

So we then moved on to do some internal model development at Nominet with the aim of reducing costs and improving accuracy. Then, in 2019, we implemented our own API, which combines a couple of models. One uses neural networks and the other one uses regular expressions to identify which domains are likely to be suspicious.

Now, I will note here that we are focusing only, at the moment, on the domain name. So when someone registers a domain, there's a lot of information in that registration. But at the moment, the current focus is just on that domain name. So these APIs are just being asked to score a domain name. Could I have the next slide, please.

So this is the process for DomainWatch at the moment., just to give you an idea of who it works. A domain is registered, at the top left. Then the registry makes a call to an API and that returns a score. So this all happens very quickly, within 10 minutes of the domain being registered. We then have a check. Is that score above the threshold?

And if it is above the threshold, it gets manually checked by the customer resolution team at Nominet. So they have a look at that domain name, together with other information—so the registrant information, which registrar it comes through. And they have their own workflow that they work at in building up knowledge in terms of what looks good or bad. Now, they may say, "No. That's okay." And it goes through to the BAU lifecycle. So it just becomes a normal domain.

Or they may say, "No. That does look suspicious." So the domain then gets suspended. At that point, a notification is sent out to the registrant and they can supply an ID and justification for their registration to reassure us that it was a legitimate registration. This is then checked and either, "That's okay. Yep. That's a legitimate reason," or it gets escalated to a committee if it's not clear. And it could, then, go back into the BAU lifecycle.

So if you go on to the next slide—thank you—I've got an example. So that wwwicloud.co.uk, that domain was registered. And the score came back. The combined score from the two models was 9.9, which triggers our manual check process.

We also get a bit more information back. So the TensorFlow model is a bit black box. It doesn't tell you much. It just gives you a score. But the Regex model does give a bit more information. So when the team are manually checking this domain, they've got the information that "icloud" was a term that was spotted and the domain started with "www," which is known to be a bit suspicious.

So in that case, it was triggered. It was manually checked and they suspended it. And the registrant was notified. And as it happened, we heard nothing back so that domain is still suspended. Could we go on to the next slide, please? Thank you.

This is the bit that's of particular interest to me. This is the modeling side of things. And as I said, we've got a couple of models in there at the moment. So there is a regular expression model, which is effectively looking for particular substrings in the domains. So we had a team of analysts who manually built this part of the model. What it does is it checks for substrings in the domain, looking for a fixed list of terms and brands.

So there's a mixture of brands, like Amazon and Facebook, and terms, like "account" and "login" and things like that. Each term is given a score. And that score can change, depending where it is in the domain name. And there's also some kind of fuzzy matching to allow for spelling mistakes. So this one returns a score. And to add new terms to it, we do need to manually add those new scores for new terms.

The second model is the TensorFlow model. So this is a neural network machine learning model. And this is where the patterns are learnt from the data. So this is a neural network built on just the domain name. And the domain name is treated as a sequence of characters and it's looking for patterns within those characters.

It was built from security feed data. So we looked at security feeds to see what had been flagged as phishing and then made some filtering to pick out the ones that appeared to be maliciously registered, and therefore looked suspicious, as opposed to compromised domains, which are ones that might have been hacked or something. And then that will give us a bad set of data—a set of domains that are likely to be malicious.

And then we use the registry data to get a set of domains that appear to be good. For this, we took a set of domains that had been registered for quite a long time and then only picked the ones that had active content as well to try and get a list of domains that looked legitimate and used by normal, everyday businesses.

What else did we do with that one? Registry data and feed data. I can't think what else we did. But we did a bit of filtering just to pull out those two. Oh. That was it. We omitted brand protection registrars. So that would make sure that if, say, the bank, HSBC registered a domain, that wouldn't go into the good set because we're trying to pick out brands in particular.

And when we ran that model, built the model and applied it to test data, we got a 98% accuracy. But this was on the artificial data that we created. When we applied it to real data, so looking at the registrations coming through on a daily basis, the precision dropped to about 60%. So when you looked at domains that got above a certain score, about 60%/ of them looked to be legitimately malicious. Hence, we have this manual check as well. So the other thing that we did look at was bringing in more data types and found that certain things did appear to be quite important in predicting phishing.

There's various attributes around the e-mail address—for example, what the host is. So is it Gmail or Hotmail? And also whether there's similar strings between the e-mail address and the registrant information. Name server is another one. Whether or not the phone number is a mobile number or not. The registrar. And interestingly, the time of week or the time of day seemed to have an impact on whether a domain was likely to be phishing or not. Could I have the next slide, please? Thank you.

So this slide just shows some of the outcomes of what we've been doing over the last few years. This is for a typical month in 2022. We have about 200,000 .UK domains registered a month. And of those, the model identified 700 domains that are worth checking. The manual checkers then suspend about half of those, so about 400, typically. Then a following 40 domains of those are then unsuspended following appeal, which is reasonably low.

The other thing that we're getting out of it is we're starting to collect more information around the phishing domains and what's going on. We've seen legitimate use cases for suspicious-looking domains. One of those is penetration testers, so people who are pretending to be malicious to test the security of the company is a valid use case for a domain which may look very suspicious.

We're also seeing repeat offenders come up, registering lots of domains. And sometimes there's domainers who register multiple domains because that's just for resale. They're not actively using them. It's a slightly different case.

And also, we are starting to see more patterns in registrant data. So I mentioned e-mail providers. Also, the country that the registrant says they're from. Registrars. And we're seeing different terms in domain names pop up.

So if you go on to the next slide, this is a word cloud of some of the common terms that we've seen the last few months. These are ones that are either trending upwards, or just generally, we have a lot of these terms in suspended domains. I've highlighted the brand ones in blue. But what's interesting is there's quite a lot—online, service, alert, support. There are quite a lot of key terms that aren't brand terms. So even though you might be wanting to target brand terms, there's a lot more out there that's worth monitoring. And the next slide, please.

In terms of what we're doing now, what we're doing next, we're looking at dropping the threshold to check some more domains. So at the moment, we've got a hard and fast threshold. And it would be interesting to dip a bit lower and see the near misses that we have identified.

We're looking at a trial of taking action against existing registrations. So DomainWatch is very much focused on new registrations. But that doesn't mean that everything that's on the registry already is all good. So we're looking at some domains on feeds and seeing what we can do to take action against existing, potentially malicious, registrations.

Also, we're looking into collecting more data—so using the web crawler to collect more data for detecting abuse. And we've identified a few ways that this will help. And then, on the modeling side of things, I'm currently looking at reimplementing it on a new architecture so we can keep it up-to-date and update things easily. But there's a lot that we can do around improving the model accuracy. One of those things involves taking a lot more data in. So registrant data feels like it would have a lot of value.

Also, reassessing how the two models are combined and whether, actually, just one model would be sufficient but one that gives everything that both the models do give.

And a final thing is automating brand detection. So where we've got this list of terms that feed into the Regex model, this is static and manually-updated. And there's definitely a piece of work around automating how we detect new brands and new terms. A good example of this is coronavirus. As of 2019, it wasn't really a thing. But come 2020, we had a lot of malicious registrations with the words "COVID" and "coronavirus" in. And that's the kind of thing that we could have tried to pick up automatically. That's a bit of a big one—a bit of an obvious one. But there are other ones that it would be worth trying to identify as we go. And next slide.

EBERHARD LISSE: One minute left.

BRYONY HILL: Yeah. I think that's it. Next slide. Yes. Thank you for listening. Thank you very much.

EBERHARD LISSE: Thank you very much. There are a few questions in the chat which basically mirror what I was going to ask. One is, is the domain name already checked between registration? Is it already active between registration or being checked? Or do you run this before you activate it?

BRYONY HILL: It is active. So yes. We have a period of anywhere from a few hours to over a weekend—it may be a few days—when that registration is active.

EBERHARD LISSE: Okay. Joel Karubiu asked whether you already check registered domains. You answered that in your presentation afterwards.

And then my main question, which Mark Elkins phrased, is are you sharing? Is this open-source?

BRYONY HILL: No. It isn't open-source at the moment. I don't know if that's something we would look into doing but I'd have to talk to colleagues about that.

EBERHARD LISSE: And Calvin Browne from CO.ZA, or from DNS Africa, asks, "How many staff hours does this take and what does this cost you, Nominet?"

BRYONY HILL: That's a very good question. In terms of staff hours, I believe that we have a small team of about three people who probably take, between them, an hour a day to manually check the domains, I believe. So that's the staff hours cost. And I think that is the main cost behind this.

EBERHARD LISSE: Roy Arends asked whether you'll recheck registered domains when the machine learning gets richer and better. But you said you're not doing this just yet.

BRYONY HILL: Sorry? Do we reach out to existing domains?

EBERHARD LISSE: Do we recheck registered names when the machine learning gets richer and better?

BRYONY HILL: I think that's something that would be good to do, although obviously, we've got a very large backlog. So I have looked at applying the models to historic data. And there's a decent number. It's thousands that come back scoring above our threshold, which would then be a very large chunk of work for the team to have a look at. So it's not something that we're doing at the moment but it's something that's on the radar.

EBERHARD LISSE: There is two questions. I'm only going to ask one, which I find more interesting than the other. "How do the registrants react to suspension?"

BRYONY HILL: We get a mix. We get a lot of people who are very impressed with the fact that we're doing this. And we get some quite positive feedback. And they don't mind having to go through an extra hoop to justify their reasons for registering a domain. I think, on balance, that's the majority. But I think there are probably still a few people who aren't happy with it.

EBERHARD LISSE: I'm quite sure that malicious registrants just move on. They don't talk. It's all automated and they move on.

BRYONY HILL: Yeah. We get some funny responses with people using fake IDs to try and get around our systems. But they're quite obviously fake.

EBERHARD LISSE: Okay. The other question can be sent directly via e-mail. Addresses of all speakers are clickable in the agenda. The agenda is posted. Thank you very much. Very interesting presentation. I hope you make this source code available.

BRYONY HILL: Thank you. Thank you very much.

EBERHARD LISSE: Okay. Hugo Salgado from .CL, also presenting remotely. And the presentation is online. You have the floor. And you are still muted. No you are not but we can't hear you. Now, there you go. Now it's good.

HUGO SALGADO: Perfect. Thank you. So hello. Good morning for everyone. I'm Hugo Salgado from NIC Chile, .CL. And I wanted to tell you about our experience with the publication and use of the ZONEMD record in our root zone. As far as we know, we are the first TLD to use it. We know that it will soon come to the root, that we'll use it mainly for out-of-DNS root zone distribution. So next slide, please.

So to begin with, what is the ZONEMD record? It's a relatively new record from last year that basically allows you to have digest for root zones. It allows you to have checksum for the zone file, mainly. At least that's how we use it. This checksum gives us the characteristic of giving us an integrity test that the zone file is complete and has not been modified. But also, if we add DNSSEC to the use of ZONEMD—that is, if the zone is signed—it also provides authenticity. That is basically, when using DNSSEC, it's known that whoever should have signed it. These sentences, I have copied directly from the RFC that defines the standard. Next slide.

Here, I wanted to make a parallel between what ZONEMD means for zones with what the checksums mean that we are used to seeing in the software that is downloaded from the Internet. On the left side, we have how the checksums normally work for downloads. And on the right side, what is new from ZONEMD.

Normally, when you download a software or some image of an operating system, you download the .iso that is [available]. And you can also download from the same place, from the same web page or FTP directory, two extra files, where one is usually called the same but with an extension .sha256, for example, indicating the algorithm used. This is the one that gives us the integrity of the file.

When we download the .iso, we can calculate this .sha256 hash with a local tool and compare it to be the same as the one in the downloaded file. With that, we are happy that the .iso was well-downloaded. It comes complete and there were no failures in the transmission or in the copy.

But in addition to that, we can download a file that is generally .asc, where a PGB signature for the same checksum comes. So in this way, in addition to verifying that it's complete, we can know that it is authentic through the cryptographic validation that PGP gives us associated with the identity of the manufacturer of that software.

So the left side is how we normally act so far. On the right side, we have the equivalent for zones. Above will be our zone in text format. If it has a ZONEMD record, it means there is a checksum with the entirety of the zone file and also if it's signed with DNSSEC and has its corresponding RRSIGs, it's also authentic. Next slide, please.

Now an important difference regarding the software download is two things, the ZONEMD and signature are already included in the same zone file. So it's not necessary to obtain them separately but everything is integrated and contained in the same format. Next slide, please.

How does it look? Since it's a record that is within the same zone, it can be [consulted] externally, it can be [queried] like any record in a zone. Here we have a query to .cl where you can see the comment, and the response comes with the ZONEMD records. Within the fields that composite, we have the serial of the zone, then some numbers that indicate the type of algorithm, and finally, the checksum itself. Next slide, please.

So why use ZONEMD in .cl? Mainly because we already had something very similar. For many years, we have had an internal system that checks the zone files on all secondary .cl nodes under our control verifying that it is correct. This was borne out of an event we had many years ago when one of our nodes loaded a corrupted zone. Actually, it was a truncated zone due to DNS server implementation error that we were never able to track down or [nor did it] repeat itself.

It was our own solution implemented by us which has some difference with ZONEMD, mainly because our solution did not go inside the zone file, it was external so we had to dump the zone file to disk, calculate a checksum in each of the nodes and send it to a centralized system that made the comparisons. The system is where it compared and checked that the checksum was the same in all the nodes and alerted in case there was a different one.

For this very reason, when we learned in the IETF that ZONEMD was being planned, we supported it from the beginning. We gave some suggestions in the drafts and that is why we are one of the few TLDs to use it.

In addition, our [inaudible] created a tool that calculates and verifies ZONEMD which is available open source for anyone who wants to use it. It's called DNS Tools.

Also, moving to ZONEMD allows us to use other implementations, both authoritative DNS server and command line tools, and use better crypto than we had. So only advantages. Next slide, please.

How do we use it? After a zone is generated and signed with DNSSEC, the ZONEMD record is calculated and sent into the zone file. At the beginning, due to how we have the architecture for the signing process, at that stage, we did not have access to the DNSSEC private key. So when we launched at the beginning, the ZONEMD record was not signed. It did not have RRSIG nor did it appear on the [inaudible] NSEC3 type map.

Now, since this record is actually for internal consumption, that is, for our own calculations and monitoring, our zone file is private. And the idea is that this is never consumed by an external party. So in reality, we did not care that it was not signed. Next slide, please.

However, as of a couple months ago, it’s already signed. We've made some adjustments in the way of implementing it, and now the .cl ZONEMD has its RRSIG and it comes out correctly in the NSEC3 record.

To sign, we used a tool, not from ours but Verisign labs called ldns-zone-digest, and it takes about 33 seconds over an entire zone file of 1.3 million records. We used generic format of record for reasons that I will explain a little later. Next slide, please.

Okay, so that was the signature. Now for the verification. Each of our nodes of our three Anycast clouds that are maintained by us. The others are external services. In each of these nodes, we execute a process some minutes after loading a zone that dumps it to disk, it’s validated that it’s not too different from the one that signs, that’s our own DNS tools that takes a little less time for verification, 24 seconds, and it reports to the centralized system for the purpose of viewing the status for each one of them.

It's basically that we’re hoping that the support in the DNS servers, in the software, will be more complete with ZONEMD. Perhaps [in a few moments,] it'll no longer be necessary to write this and calculate it with an external tool, but rather, it will be the

DNS server itself that calculates it at the time of receiving and verify. Next slide, please.

Here, I wanted to show you a little how we're working. If we see on the bottom left, that's our registry, our database where we generate zone file which is signed with DNSSEC. That is then passed through another operator that computes the ZONEMD record and [inserts] and signs it.

This is then sent as a file to our central monitoring, and [inaudible] is already distributed to the normal DNS distribution platform using the normal transfer protocol of DNS. From there, it goes to all the nodes of the clouds. Next slide, please.

Then each of the nodes, what you see here on the right, at the time of receiving the zone via the transfer, in addition to starting to send it to clients on the Internet, a few minutes later, the zone is dumped to disk and the ZONEMD check is performed.

In addition to this local check, it's reported to the central monitoring system. One important thing here that I want to mention is that one of the advantages of using ZONEMD is that each node is totally autonomous. Since the checksum comes within the same zone, each node does not need more information than that to make the calculation and detect if it's correct or not. Each node can know if the zone it received is correct or not and even decide to stop responding if there is a problem.

This is unlike our previous system where simple checksum was separate. We needed centralized system to make the comparison and decide at that moment if everything was fine. So this is an advantage of the new scheme. Next slide, please.

I wanted to tell you some things about what we learned, a little experience. The first thing is that our DNS tool is written in Go language which requires a lot of RAM memory. We need at least 1GB of ram available for the check which caused certain problems in nodes that were a little tight on memory. It also requires quite up to date OpenSSL libraries due to Go dependency issue. So it’s an issue that must be taken into account before it’s deployed.

In addition, as I told you a little earlier, the format we used for the record in the zone is generic using this characteristic of using TYPE63, because the server did not know the ZONEMD record and gave errors due to unknown format. We will then move to the particular format once full support is in place.

And lastly, one thing that is very important to be careful of is that there's a certain software for example cannot—that when ZONEMD begins to have internal support, is capable of deciding not to load a zone if a record does not validate. That bothers a bit because we don’t want so much automate such a delicate part. We always prefer that a human intervenes. So we have to be careful.

Those two directives that are there are to prevent [inaudible] from generating ZONEMD record and also from deciding to suspend the load if the verification is unsuccessful. What we want is for the verification to generate an alert and a human operator to act. Next slide, please.

So that's all. As I was telling you, it was super natural for us to adopt it. it was a good experience and we're quite happy. Previously, we had about two or three failures a week, all false positives, mainly, because in the previous architecture, having to notify the checksums to a central monitor for comparison caused problems due to communication between the nodes and the monitor. They were not reachable, network problems, which gave false positives. So it wasn't really common for operators to receive these alerts and verify that everything was fine.

And in this year or so that we have been running, we have not had any false positives except for some alerts that have been, as I told you before, memory problems in one of the machines or delays in transfers. But other than that, nothing important. So it's much more stable than our previous solution.

So that's all. Thanks. I don't know if we have any comments or questions.

EBERHARD LISSE: Thank you very much. Did I understand you correctly in saying this would be open-source?

HUGO SALGADO: Well, we have our code that is pretty intertwined with our kernel system. So I don't think it's valuable at all. What we expect is that the DNS software itself gets proper support for ZONEMD records so in the future it will mean it's not necessary to have your own code.

EBERHARD LISSE: Cool. Thank you very much. Always nice to hear things that are cutting-edge. And therefore, let me just quickly look who is the next one. Next will be our host presentation, Moritz Müller from SIDN. You have the floor. And try to not get too much into the break in front of some questions if at all possible.

MORITZ MÜLLER: I will try my best. Thank you. So hi, everyone here in the room. And hi, everyone online. My name is Moritz Müller. I work for SIDN. We are the registry of the .NL ccTLD. And I work specifically for SIDN Labs, which is the research department of SIDN. And I'm also a part-time researcher at the Universiteit, also here in the Netherlands.

In this presentation, I would like to give you a quick overview of what we are doing at SIDN to keep .NL running but also to talk about what we're doing at SIDN Labs, introduce you to some of the research projects that we're doing, and finally give you a quick overview of a project I was part of at the beginning of this year where we found a bug—a vulnerability in Google's public DNS.

So just a quick overview about SIDN. We are a foundation. And our objective is to increase the society's confidence in the Internet. So we are interested in the confidence in the Internet in the Netherlands, in Europe, but also worldwide. We try to provide secure and fault-tolerant registry service for .NL. And we do that by running Anycast DNS services with DNSSEC support. And we'll talk about that a bit on the next slide. And we have additional services where we have, for example, registration and domain protection services.

Additionally, we try to increase the value of the Internet by enabling safe and novel uses of the Internet itself. So we have, for example, SIDN Fonds, which tries to support other small startups, for example, which share the same mission as SIDN. And we also operate IRMA, which is a project that has the goal to enable privacy-friendly authentication on the Internet. And what I am part of is we have SIDN Labs to increase the infrastructure security and trustworthiness. And I will talk a bit about the projects that we're doing later on.

As I've mentioned before, we are a not-for-profit organization with a public role. We are not located in Amsterdam. We are not located in Rotterdam, not even The Hague. But we're located in Arnhem, which is one and a half hours away from here by train. And if you have some free time, people here in the room, I welcome you to visit the small city of Arnhem.

Here's an overview of the number of domain names which are currently registered for .NL. We have more than six million since some time now. And what we are especially proud of is that more than 3 million of them are DNSSEC signed, which makes us one of the biggest zones with signed domain names.

This is a screenshot from our own statistics website, sidnlabs.nl. And there you can find all sorts of technical information about the .NL domain name, ranging from DNS and DNSSEC related information to also information about the [DAP] content in .NL, looking at TLS certificates. But also there, we look into things like RPKI deployment, in case you're interested.

Then I've mentioned that we try to maintain a stable .NL. And we do that by spreading .NL across four different name servers. All of them are Anycasted. Three of them are provided by third-party providers. And there's one, since a few months, run by ourselves, where our operations department set up their own Anycast servers with support of insights that we gained in research that we carried out at SIDN Labs.

Currently, it's consisting of virtual machines at different cloud providers. I think, at the moment, it's one cloud provider with different sites all over the world. So we try to achieve coverage not only in Europe but also on the other continents as well.

And one of the parts that we at SIDN Labs especially were involved was the optimization of the catchments of the different Anycast sites so there, we tried to provide some more insights into how the catchments of the different sites look like. There's a screenshot on the bottom left where you can see the catchment, I believe, of a site in South America—I think in Brazil. And you already see that the catchment is not ideal because it seems to attract some traffic from Italy as well. So we've tried to provide our operations department with these kinds of insights such that it can optimize the DNS services.

Now let's talk about the team I am part of, the SIDN Labs team. We are a technical diverse team—only technically, not gender-wise or racial-wise. But at least from the technical side, we are quite diverse. We have people from an academic background. I'm one of those people. But we also have people from an operational background, with a software development background. Since I started SIDN in 2014 as an intern, we grew quite a bit. Back then, I think we were four people, I believe, including our manager. And now we are 12 people.

Overall, we write open-source software. So we try to contribute back to the community. And usually, we also try to have some maths students in our team as well, to foster interaction with universities, and that allows us to work on small research projects.

Our goal is to increase the trustworthiness of our Internet infrastructure for the Netherlands in particular, but of course, also worldwide. We do that by applied research. So we usually don't do ground research, we don't try to bring things in production ourselves. But we try to lay the groundwork for services and production. We do that by measurements. And I will come with an example later on as well—by designing different new services, prototyping, and evaluating the services.

And our goal is always to make these results publicly-available if possible so we write academic papers to publish our results or we try to make our software open-source and contribute thereby back to the community. We work together with our own operations team, as in the example I've described before with the Anycast severs. But we also work together with universities in the Netherlands but also in other countries. And we work also together with other research teams outside of SIDN.

We have three research areas that we work on. The first one is probably the most obvious, which is focusing on DNAS, BGP, and NTP. We have NTP servers since a few years, I believe, as well, which even serves more queries per second than our DNS servers. We have a pillar which focuses on domain name and IoT security, that we look into projects similar to the one that Nominet was presenting—for example, detecting malicious registrations. But also, we have a software which has to go to secure the user's home network and especially detect malicious IoT devices and protect these devices as well.

And then we have a third pillar, which is the secure future Internet infrastructures, where we look into alternative Internet infrastructures, not to replace the Internet but to find if there are use cases where these alternatives future Internet infrastructures make sense.

To make this a bit more concrete, here are a bunch of examples. On the top left, we have one of these measurements studies I was part of, this measurement study where we've looked into DNSSEC algorithm agility, and measured the deployment of certain signing algorithms. The picture here in the top center is our NTP antenna, you could say, on our office, getting a time signal from Germany and also from GPS satellites.

The screenshot here shows our SPIN software to detect malicious activity in home networks, which is open source as well. And you can install it also on your own hardware. This one is an example of a project where we try to identify malicious activity in .nl domain names. Here, this is a project where we tried to identify logos and websites. This was in cooperation with the Dutch government and had the goal to give the Dutch government information where its logo is being used, and thereby helping them to detect malicious use of the logos.

And these two here in the bottom are related to the future Internet project where we, for example, look into the future Internet protocol SCION and implementing that on P4 to actually run it on hardware and see how it performs on hardware actually, whether it's possible to implemented in hardware, but also helping to bring it a bit into production. And this is all part of [inaudible] project where we do this together with other partners here also in the Netherlands. If you have any questions about these recent projects, I can try to answer them later on. Or I can just connect you to one of my colleagues who are involved with them.

And, as I've mentioned, we do experimental research. So we don't do fundamental research. We don't do operations, but we are somewhere in the middle. All these projects we don't do on our own. We have partners from academia, with different universities here in the Netherlands, with universities outside of Europe as well. We work together with the government, we also work together with ICANN. So we really want to connect to other people and to broaden our scope as well.

And with that, finally, to the research project I was involved in, though you couldn’t call it really a research project, but more a side project in which we by accident found a vulnerability in Google's public DNS service. This vulnerability had the potential impacts that you could spoof resource records in Google public DNS of domain names of a domain name, despite of the domain name being signed with DNSSEC.

We found this vulnerability in early January 2022. And it was fixed by Google a few weeks later. So it was pretty fast. So how did we actually find this vulnerability? My colleague, Marco, he's operating servfial.nl and this is a domain name which is on purpose not correctly signed. So it's a bogus domain name, basically. And this domain is being used by us internally for monitoring, for testing, for measurements. But it's also being used by other parties as well to test their DNSSEC implementations.

And of course, as we all know, you can break DNSSEC in many different ways. And until the end of last year, servfail.nl was broken in one way, but Marco wants to break it in another way. Because for him, it made more sense. And he decided to make it bogus by signing the records in serfail.nl with a nonexisting key. This was the basic idea.

So this is then how this domain name looks like. This is a screenshot of DNSViz showing the servfail.nl zone. Here on the bottom, we have the four resource records. Here on the right, we have the ZSK, we have the KSK, and the KSK is signing the ZSK or the whole key record set. And also the DS is a hash of the KSK. So everything was fine on this side. But however, if you look here on the left, then you will notice that these records here should be signed with this key 45918. But this key, as indicated by DNSViz, does not exist. And this makes these records bogus. So they should not be valid, these domain names, and this is also indicated by DNSViz with this red line around the records. And if you [inaudible] for example quad A with a validating resolver, in this case, quad 9 coincidentally, then quad 9 would give you the error code, servfail, and the actual resource record is not in the answer so this is how it should be.

However, when we checked this with Google public DNS beginning of January, then we actually received the status, no error, which is already a problem. And we also saw that it's contained the actual resource record, the quad A record. And if you would look at the flags here, then you'd also notice that the [ID] flag is not set. So apparently, Google public DNS did not validate the record.

So this is definitely something that doesn't look right. So how could you misuse the situation? In theory, you select a domain name of your choice, which is signed with DNSSEC. Say for example, sidn.nl as an example. And you want to make sure that you get redirected to your malicious web server where you serve the users with a malicious version of sidn.nl to infect the user's computer or to serve some kind of information. So you create fake resource records of the targeted domain name. And then also create fake signatures of the resource record with non-existing key.

This is relatively easy. The harder part is probably performing a cache poisoning attack against Google public DNS, performing a regular cache poisoning attack, as you would do it also with non-signed domain names, using this spoofed [inaudible] record and using the fake signature. And if you succeed with this cache poisoning attack, then you would be able to put these malicious records into the cache of Google. And therefore, the public DNS server would then respond to the clients with the malicious records if they would ask for sidn.nl, with all the consequences I've described before.

If you look at the actual impact, then this is a bit harder to estimate. From our perspective, Google public DNS was likely the only affected resolver that had this bug. We ran measurements on the Internet using web Atlas and scanned for more than 10,000 other resolvers out there. And Google public DNS was the only one that had this behavior at this point in time.

Also, Google themselves don't believe that they have been misused. And this is because of their quite complicated caching infrastructure. So cache poisoning attacks against them should be quite hard. And they fixed this problem within one and a half months. But we're not sure how long this problem actually existed in the code.

If you want to know more about this whole problem, then you can find more information in our blog post where we also sketch some more information on how we actually found this vulnerability, how we checked whether other resolvers were affected, and also a detailed timeline.

Are there some takeaways from this? Can we learn something from this? I guess one of the main things that we can learn here is that DNSSEC is still hard and does have many corner cases. And in this case, I think the RFC was quite clear. This is a record that should be bogus. But this is just one of the examples that there are many different ways to break DNSSEC, and this can cause problems.

Another issue here was by PowerDNS. They had some internal testing which was not a security issue at all but with their internal testing, they also noticed a corner case which they didn't think of. So if you want to do DNSSEC validation, then you probably should rely on existing libraries and resolver software to do this. Of course, also there, you never can be sure that all the corner cases are covered, but at least more people have looked at that.

If you want to implement DNSSEC validation yourself, think of as many corner cases as possible, break things, and you will probably notice that your validation software will have some problems at some point. And with that, that's all there is. I think I've managed in time. So if there's some questions.

EBERHARD LISSE: Thank you very much. Actually, you've used much less time than we thought you would. Not a problem, that gives us more time for discussion. And there was one. Did you get Google bounty money for the bug bounty?

MORITZ MÜLLER: Yes, we did get Google bounty money, and we donated the money to an open source project.

EBERHARD LISSE: Thank you. I thought something like this would be the answer, being a nonprofit and so on. Then what else? Any hands that are going up to ask questions directly? Ivan Minic on the chat asks, why do you think that over 60% of.nl domain names have active DNSSEC?

MORITZ MÜLLER: Because we measured and we see it in our registrations. So we not only see the DS records, but also make sure that these domain names are validly signed. So [only uploading] DS is not enough to be counted, but we actually check whether this whole chain is actually valid or not.

EBERHARD LISSE: Oh, I don't think that was the question. The question was, why are so many signings? Nobody's [inaudible] the number, the question is, why? Are you giving them a discount?

MORITZ MÜLLER: Understood. Yes. This helps, I think. I think we also do a lot of outreach to people to educate people about DNSSEC. But we also give discounts.

EBERHARD LISSE: I think that’s an important thing. If I'm not mistaken, .cz does the same or similar thing. And I think that's still one of the ways to get—money talks. That's one of the ways to get things signed because most end users, most end companies, even most registrars couldn't care less. Only if they are forced or if it's commercially interesting for them, then they become interested. That's at least my view. We are not able to convince anybody locally. They think HTTPS is as good as they need. Are there any questions? There is Terence Eden from.gov.uk. You seem to be sitting locally, so please unmute your desk microphone.

TERENCE EDEN: Hi there. Really interesting presentation. Do you do anything specifically looking at government websites in the Netherlands or beyond?

MORITZ MÜLLER: I know that government websites as far as I know need to be signed with DNSSEC. So we have a list of standards that government websites have to comply with. And I believe DNSSEC is on there as well with for example, also IPv6. We specifically at SIDN do not look at those domain names, maybe occasionally for research or so, but we don't have any structural tool that looks at these domain names.

TERENCE EDEN: Thank you.

EBERHARD LISSE: Okay, I don't see any more hands. Let me look in the chat if there is anything. I don't see anything there. Good, then we are a little bit early. So I'm calling the break now. We must be back together in half an hour. So that means 36 minutes, 13:00 UTC. Thank you very much, and we'll see each other after the break.

[END OF TRANSCRIPTION]