JORDYN BUCHANAN: Thanks, and welcome, everyone to this Plenary call of the CCT Review Team. I'm Jordyn Buchanan, subbing in for Jonathan who is unfortunately absent today, as is Laureen.

Today we're going to spend most of our time, again, talking about DNS abuse, as folks hopefully saw – Drew sent around a revised version of the DNS abuse paper. But we're going to also have a presentation from the authors of the DNS Abuse report, assuming that Martin is able to join, but I see Maciej is here, at least so hopefully, we'll be able to get some time with the authors.

And then our only other agenda item is to take a look at Jonathan's latest revision of the parking paper.

Before we get to that, let me just ask if anyone has any revisions to their Statement of Interest. All right, it looks like no. Oh, Calvin's typing. Maybe he does. No, okay.

All right, no updates to Statements of Interest. I will go ahead and jump then to Agenda Item 4, which is a review of the updated version of the parking paper. I'll note that there's still some numbers that we're taking a look at, or specifically, I'm supposed to be taking a look at, that may, each week, some of the exact figures in this chapter.

I actually think even if we don't get those revised, it's not necessary for this iteration, which is really mostly supposed to be based on the new nTLDStats data which is now incorporated into this chapter.

Is someone typing loudly?

In any case, Jonathan sent around a few other edits. This is being presented right now. Is this scrollable? Can anyone scroll? I can't scroll. Aha, hopefully this means we can all scroll.

So, I just wanted to open this up and see if anyone has further comments on the parking paper, or questions.

Waudo's hand is up. Go ahead, Waudo.

WAUDO SIGANGA: Yes.

JORDYN BUCHANAN: Yes, go ahead.

WAUDO SIGANGA: Okay, Jordyn, about the title, I think I agree that the title should remain still parking, not this changed title. Is that correct?

JORDYN BUCHANAN: Good question. I do think we agreed that.

WAUDO SIGANGA: Sorry. Sorry, Jordyn. One other thing about that particular issue that we did discuss last time is that I kind of went through the main report again and there are many [cases] where it's talking about parking. So if we decide to change the title here, it's going to affect a lot of areas within the main report as well as in this particular paper.

In this paper, [inaudible] confident the word “parking” was appearing 65 times, so a change from the word “parking” is going to cause a lot of editing work.

JORDYN BUCHANAN: Yeah, thanks, Waudo. I do remember that –

MAARTEN WULLINK: Hello, this is Maarten. Sorry for the delay.

JORDYN BUCHANAN: Hello, Maarten. Thanks for joining. We’re just wrapping a conversation on another topic that we started, so we’ll be back to you in just a minute.

MAARTEN WULLINK: Okay, great. Thank you.

JORDYN BUCHANAN: Okay, Waudo. So yes, I do remember we discussed this on the last call and I think you’re right that we had agreed to just continue to use parking. That was the sense of the folks on the call. So let’s relay that feedback back to Jonathan who’s been holding the pen on this, but unfortunately, isn’t on the call today. Any other feedback on the latest revisions to the parking paper?

And we're getting to the point in this process that we're going to be finalizing these sessions or at least finalizing for the interim for the supplemental draft report that we'll be sending out in just a few weeks. So if you have any input, I think this is basically the last chance to provide it. Other than Waudo's comment, we'll consider this section pretty baked.

All right, great. So why don't we go ahead then, and we'll call that a wrap on the parking discussion and jump back to the Agenda Item #2, which is the presentation on the DNS Abuse study from Maarten and Maciej. So we can switch the other deck and turn the mic over to them.

BRIAN AITCHISON: So actually, Jordyn, I'll just give a quick introduction. Yeah, and you've actually already seen this. Could I get scrolling or presenter rights? There we go.

Yeah, I'll just walk you through the first few slides. You can see these are actually the slides we are going to be using for the webinar as well, so it's kind of a preview of what we're going to look at.

Just to give you a little background as you've already seen, this is basically coming out of a 2009 paper where we asked these security communities these four questions that you see on the screen, and then we get these safeguard recommendations that you're all quite familiar with by this time, and we've seen in Drew and Calvin's chapter.

You know that we have this 2016 paper that sort of crafted this research model that we're now seeing the sort of deliverable on, I suppose.

Then enter the Review Team and you have a mandate to look at all these issues, so this is all the timeline. This is what we're going to go over on the webinars tomorrow, just to give you kind of a preview of that.

So we won't spend too much time on that. We'll kind of turn it over to the stars of the show, Maarten and Maciej. Maciej, I see you're still on mute. That's a common problem on Adobe Connect, so you may want to take yourself off mute before we jump right in. I'm not sure who's going first, but Maarten, Maciej, I will turn it over to you.

MAARTEN WULLINK: Okay, thank you, Brian. I'll do the first section discussing a little bit about the data we used, then Maciej will take over and discuss domain reputation, and I'll finish off with registrar geographical location and privacy or proxy services.

Can I switch the next? Yeah, okay.

So basically, the goal of the study is I guess already known to everybody here. It's come up with a comparison between the DNS abuse and [the] legacy gTLDs. We have been looking at spam, phishing and malware for that, and also, wanted to have a fiscal analysis of potential abuse drivers, maybe things like DNSSEC, and of course, the motivation for all this is the ICANN New generic Top-Level Domain Program.

So for this study, we used what we think is six of the best blacklists out there for phishing, mainly Anti-Phishing Working Group, which is a very well-known blacklist, contains phishing URLs and stop [inaudible] where

it contains malware URLs, SURBL, which actually contains four separate blacklists, contains three types of content, such as phishing, spam, and malware domains.

And the last three are Spamhaus which is a huge depository of spam domains, [inaudible] which again, has three different types of feeds, which contains phishing, malware, and something they call defaced URLs, basically hack websites. And we also added the security domain foundation, which among other things has phishing and malware URLs.

So those are the blacklist data we use in order to be able to map domains that are getting reported through these blacklists. We need to have WizDta as well to be able to map the abuse to a registrar's geographic location.

So for that, we used the Wiz XML API which contains data for all the new gTLDs and a subset of the legacy gTLDs for the period of early 2014 up until the end of 2016.

One problem we faced using this data set was that due to the method of data collection, some gaps – I shouldn't say gaps – for some domains, no data could be found in this data source, so we needed to have an additional data source for these missing domains, and domain [pools] were willing to provide us with the missing data, which allowed us to cover all the abuse of domains with WizData.

We also used domain data and formal zone files for three years, one zone file per day for each TLD provided by ICANN. And for the differential analysis, we also scanned our own data. I should say we scanned domains and created our own data.

So we did active web and DNS scans. We scanned each of the new gTLDs at I believe February or March of this year, May I guess, of this year, and those were, at the time, I believe some 24 million new gTLDs and to create a sample of the legacy gTLDs to scan because there are too many legacy gTLDs to be able to scan them all in this short time span. And we came up with a representative sample of some 17 million domain names to scan.

And finally, we also used registry information from ICANN containing information such as Sunrise Period and information about registry operators and parent companies of registrar operators.

Okay, so now Maciej will take over.

WAUDO SIGANGA: Hello? Yes, [inaudible].

MAARTEN WULLINK: Yes?

WAUDO SIGANGA: Hello? Is that okay? [Inaudible] the previous slide.

MAARTEN WULLINK: The previous slide.

WAUDO SIGANGA: Yes, on the previous slide. Yeah, you say that the legacy gTLDs, you took out samples because there are too many. How many legacy gTLDs are there? I thought there were a couple of dozen or something. The legacy gTLDs, [there are] many. The new gTLDs are the ones that [inaudible].

UNIDENTIFIED MALE: Yes [inaudible] Maarten.

MAARTEN WULLINK: Waudo, I should [inaudible].

UNIDENTIFIED MALE: Please, go for it, Maarten.

MAARTEN WULLINK: I should have said we took a sample of domains registered for legacy gTLDs. So there are, of course, not that many legacy gTLDs.

WAUDO SIGANGA: Okay, sorry [inaudible].

MAARTEN WULLINK: [Inaudible] registered. Yes. Sorry for the confusion.

WAUDO SIGANGA: Understood, yeah.

MAARTEN WULLINK: Okay, so if there are no further questions, then Maciej will take over here.

MACIEJ KORCZYNSKI: Thanks, Maarten.

Yes, so to determine the distribution of malicious content or malicious domains across gTLDs and also registrars and privacy and proxy providers, we built on our previously proposed free security occurrence reputation metrics.

So first, we analyzed the number of unique domains. Although it is the most commonly used and most intuitive reputation metric, it also has its own limitations.

Why? Because it might not give an indication of the amount of abuse that is associated with a single domain name. Let's just imagine one maliciously registered domain name that is used by the [inaudible] extensively in different phishing campaigns against different banks.

So for this reason, we propose a second complimentary reputation metrics and we analyze a number of [actually fully] qualified domain names. But also, this metric, it actually has some limitations because it might not indicate the amount of abuse associated with single fully qualified domain names.

Let's just imagine a domain that is compromised to, let's say, vulnerable plugging of content management system and then the [inaudible] used

extensively to distribute malware, binary integration files, and so on and so on, through URLs.

So actually, this work, you might still remember from the finance work, we were saying that it stems from our collaboration with the Dutch National Police where we analyzed URLs that were used to distribute child abuse materials, and then we realized that one fully qualified domain name can be used to distribute just one, for example, malicious photo whereas others are used to distribute tens or even hundreds of malicious photos using distinctive paths or URLs.

And we also were close [inaudible] and manually analyzed other types of data sets such as phishing and malware, and actually, in those types of abuse, we also observed this trend.

So here, just to give you an example, here we present those three proposed reputation metrics and phishing domains, fully qualified domain names and URLs or paths based on Anti-Phishing Working Group and parent legacy gTLDs.

So what we did here, we aggregated these incidents on a quarterly basis, as you can see here. And here, on the y-axis, we actually present the number of incidents in a logarithmic scale.

So what we can see here is actually a very significant difference between those three reputation metrics. So here, on the bottom, we see a number of domains, in the middle here, the number of fully qualified domain names, and the top line corresponds to the number of URLs, or paths.

So actually, what we observed, for example, at some point in time, we discovered around from what I remember, more than 50,000 of fully qualified domain names, under one domain name, Amazon [inaudible].com.

Here, for example, in terms of paths, we observed something like 700,000 URLs related to just one fully qualified domain name, t.co which is a well-known, of course, URL, shortened from Twitter.

Actually, those three reputation metrics are very useful and very complimentary. And what is also very interesting here is that they reflect the other [constructed] maximizing behavior. The [inaudible] do abuse free legitimate services such as online storage web services and URL shorteners that I mentioned, and they affect the reputation of such associate services.

So from now on, we will concentrate only on the number of domains as a reputation metric. So here, we present phishing domains from Anti-Phishing Working Group, aggregated there new gTLDs and legacy gTLDs.

Again, we aggregate abuse counts on quarterly basis, and y-axis, we again, present the total number of blacklisted domains in a logarithmic scale. So what we can see here is actually that the number of incidents in legacy gTLDs is very stable and at the same time, we see the clear upward trend in a number of abuse domains in the new gTLDs.

Also, please note that we marked the total number of abuse domains and it's mainly driven by legacy gTLDs, and .com and abuse.com domains.

When we take a look at another phishing [feat], this time from a clean-mx phishing, then we see exactly the same trend, upward trend for new gTLDs and very stable number of abuse domains in legacy gTLDs.

Here is another graph from SURBL phishing and when we now analyze the number of malware domains, here we see SURBL malware. And here we see malware domains with clean-mx malware, we see exactly the same trend.

So while the number of abuse domains remains approximately constant in legacy gTLDs, we observe a clear upward trend in the absolute number of phishing and malware domains in new gTLDs.

So now, let's take a look at actually spam. So here, the results are very, very interesting, actually, quite alarming. What we see here is that at the end of 2016, we see that actually, the number of domains in new gTLDs, there are more abuse domains in new gTLDs than in legacy gTLDs. When you take another spam [feed], SURBL, [inaudible], we observe exactly the same trend.

What is also very interesting is that the total number of abuse domains remains very stable and what we actually observe there is that the [inaudible] switch from legacy to new gTLDs.

UNIDENTIFIED MALE: I'm sorry. There is a question from –

MACIEJ KORCZYNSKI: Please.

UNIDENTIFIED MALE: I saw a question from David Taylor in the other room that is maybe to [his end], so David [can you confirm]?

DAVID TAYLOR: Yes, sorry. It was a question. I was going to wait for a little bit. I couldn't remember whether there was a slide later on or we got that where – I was looking at the phishing – where we're looking at the total abuse in new gTLDs, legacy gTLDs, where we normalized it against numbers of registrations. There may be a [slide] elsewhere coming through.

UNIDENTIFIED MALE: Yes, we will discuss it [inaudible].

DAVID TAYLOR: I thought so, yeah. So I was jumping ahead. I got an answer to my own question, so [inaudible]. Thanks.

MACIEJ KORCZYNSKI: Sure. Okay. But, of course, what is very important here is if we want to really compare the concentration of abuse domains in legacy and new gTLDs, we need to take [inaudible] size into account.

So how we calculated size? We verified the number of domains in each zone, in the new gTLDs and in legacy gTLDs that we used in this study. And as you can see, also starting from 2016, we see a considerable

growth in the number of domains in new gTLDs while the number in legacy gTLDs remains quite stable.

How we calculated rates? We calculated the number of blacklisted domains divided by all domains involved in zone files at the end of each study, period, on a quarterly basis and multiplied by 1,000 registrations.

So now, we're moving to actually abuse rates. And here, we present a time series of abuse rates of phishing domains in legacy gTLDs and new gTLDs, and here, based on the Anti-Phishing Working Group feed.

This time, on y-axis, we have rates in a linear scale.

So what is very interesting here is that both lines related to new gTLDs and to legacy gTLDs are converging with time and we see almost the same rates at the end of 2016.

When we take a look closer, we see that [x minus] 82.5% of all abuse domains in new gTLDs are in .com. This is not surprising because, of course, there are many, many more .com registrations than any other. And what's interesting, the top five most abused legacy gTLDs are .com, .net, .org, .[inaudible], and .biz. And that is exactly the market share of the legacy gTLDs. And that actually gives the intuition about how they're abused. And the great majority of domains are compromised domains, but I will [break off it] a bit later.

Now when we take a look at the new gTLDs, we see, at least for Anti-Phishing Working Group, that the top five most abused new gTLDs collectively on 5.7% of all blacklisted and all new gTLDs, that those do

not reflect the market share and the top five most abused new gTLDs, they're changing depending on the analyzed data.

So when we take a look at the abuse rate of malware domains in legacy gTLDs based on [inaudible] feed, we see in 2016, an exponential growth in terms of these rates. And we see that almost entire 2016, we observe more higher rate of abuse domains in new gTLDs than in legacy gTLDs.

And what is very interesting, and of course, something to be expected, are the results for spam domains in legacy and new gTLDs. And this is a figure based on the Spamhaus feed. And here, the differences are really, really huge.

For example, at the end of 2016, we observe a rate for legacy gTLDs. It was around 50, and for new gTLDs, it was around 500, which is one order [inaudible] higher.

So now let's take a look at the top –

WAUDO SIGANGA: Question.

MACIEJ KORCZYNSKI: Please.

WAUDO SIGANGA: Hello. It was a question about two slides behind.

MACIEJ KORCZYNSKI: This one, or –

WAUDIO SIGANGA: I think it's this one [inaudible]. No, the next one. Yeah, this one. [Inaudible] abuse rates of malware domains [inaudible] and new gTLDs. So I wanted you to explain what's happening with the new gTLDs go below the legacy gTLDs in terms of [inaudible] domain [inaudible].

MACIEJ KORCZYNSKI: Yeah, [inaudible]

WAUDO SIGANGA: [Inaudible]

MACIEJ KORCZYNSKI: Yeah. I'm sorry. I couldn't hear the question at all, actually.

WAUDO SIGANGA: Hello?

MAARTEN WULLINK: Waudo, we can't hear you properly.

WAUDO SIGANGA: Sorry, I think just continue with the presentation. I have a poor connection. Sorry about that. Just continue.

MACIEJ KORCZYNSKI: Okay, thanks. Yes, so here, we present top ten new gTLDs with the highest relative concentration of blacklist domains according to SURBL at WS data set and Spamhaus. Those results are for the fourth quarter of 2016.

And what we can see here, for example, if you take a look at the dot-science, that is the most abused gTLD in terms of relative concentrations, we see that the rate is around 5,000. What does that mean? It means that actually more than 50% of domains were abused, and this is quite impressive number.

If we take a look at the second one, .stream, it's 4,700 and that means that 47% of all domains that we found in zone file were actually blacklisted by Spamhaus.

And when you take a look at the results for SURBL.us, then those trends are actually very similar. The most abused newer gTLD in the fourth quarter of 2016 is .racing and the rate is 3,800. And that corresponds, of course, to 38% of all domains that we found in zone files were blacklisted by SURBL WS by blacklist.

So now the question is does the problem only affect all new gTLDs? And the short answer is no. Spamhaus, for example, and SURBL Blacklist that were really, really [inaudible] blacklist reveal that as many as one-third of all new gTLDs that we studied that were available for registration did not experience a single incident in the fourth quarter of 2016.

On the other hand, when we analyzed the Spamhaus data, we revealed that in 10 to 15 new gTLDs in the fourth quarter of 2016, Spamhaus blacklisted at least 10% of all registered domains in each of them. So that's also quite alarming. Those are quite alarming statistics.

So far, we were analyzing the blacklist, but what is actually very important is to distinguish between compromised and maliciously registered domain names. And this is quite critical because they actually require different litigation actions by different intermediaries.

For compromised domains, it's more hosting providers and web masters that should clean the content of a hacked domain or hacked website. On the other hand, for maliciously registered domain names, it should be more registrar that suspends a domain.

So in our study, we used three heuristics. The first one, if the given domain name contains a [string] of brand name or its misspelled version, and the third heuristic was if it was involved in malicious activity within three months after its creation. And those are taken. We built our method based on previous studies, for example, by Anti-Phishing Working Group by Greg Aaron and Rod Rasmussen and their recent study and global study on phishing.

We actually see that the attackers, what is also quite interesting, tend to age domain names so that they get a higher reputation score from security organizations.

So now let's take a look at the results. So here is the graph for Anti-Phishing Working Group. So what we did, first of all, we filtered out all

third party domains. We labeled them legitimate, and those are legitimate domains of services that are misused by the attackers.

We buy ours so we maintain a list of those services that are exploited by the attackers. Those are, as I was mentioning before, URL shorteners, free hosting services, web storage online services, and so on.

We also experienced not more than 1.2% of unlabeled domain names and it was because of the missing data, like WHOIS data. And the rest we label as maliciously registered or compromised.

And here is another very, very interesting finding, is that with the time, the attackers tend to more and more maliciously register domain names. So, so far, until 2014, it was the great, great majority of domains were compromised and now, we can see that with time, it actually changes.

And we performed very similar analysis for [inaudible] data and the results are very similar.

So let's now take a look at the compromised domain names. So here on the top, we see the rates of all blacklisted malware domains based on [inaudible] data set. And here on the bottom right, we see the rate of compromised malware domains only. New gTLDs in the line is purple and for legacy gTLDs. And what we can see here is that the rate of abuse domains in legacy gTLDs are driven by compromised domain names. So please note that lines corresponding to legacy gTLDs in bottom and top figures are very similar.

Now when we take a look at maliciously registered domain names, so now on the right bottom, we see rates of maliciously registered and malware domain names, we see that the rate of abused domains in new gTLDs are driven by maliciously registered domains. And actually, those rates are relatively much higher in comparison to legacy gTLDs.

And also, an interesting finding is that at least in URL blacklists, we see that those can be driven by single campaigns. So what our manual analysis reveals is that there are campaigns that we find anecdotal evidence that those are single campaigns or multiple ones. We see a lot of domains with, for example, Apple, or iPhone or misspelled versions of those, registered by the same entity, registered with the same registrar, more or less within the same time or even within the same day. And sometimes those statistics can be driven by single campaigns.

JORDYN BUCHANAN: Just to clarify, like on that last slide, what do those two graphs represent?

MACIEJ KORCZYNSKI: Sure, absolutely. So here on the top, we see the rates of malware domains and those are two lines correspond to the green one, to legacy gTLDs, and the purple one to new gTLDs. And here, we distinguish on the right bottom, we distinguish only maliciously registered domain names. And while we are doing it just to show that the great majority of all blacklisted domains are actually maliciously registered ones.

JORDYN BUCHANAN: For the new gTLDs.

MACIEJ KORCZYNSKI: For the new gTLDs, yes.

JORDYN BUCHANAN: Okay, thanks.

MACIEJ KORCZYNSKI: Sure. So one of the main goals of this study was also to perform inferential analysis of abuse in the new gTLDs and to verify if actually there are certain – if we can distinguish certain drivers, certain features that will drive the abuse counts.

So I don't want to go too much into the details of statistics here because that's not the point, but just very briefly, we used negative binomial [inaudible] linear models here. And what we did, we pre-selected and we collected and measured several independent variables. And the dependent variable here was the number of blacklisted domains. So first, new gTLD sites.

And what's all the rationale behind? So, in fact, our previous study on the entire DNS ecosystem and of all gTLDs revealed that there is a strong positive correlation between the number of domains in registry and the number of abuse domains.

So in other words, larger TLDs have a larger attack surface. But that strongly corresponds to compromised domain names and if we analyze

all TLDs, then in legacy gTLDs and ccTLDs, abuse there is driven mainly by compromised domain names.

So second potential driver was DNSSEC. So again, in our previous study, we used it as a proxy for security airport. Why? Because some registries, they incentivize their registrar so that they sign their DNS with DNSSEC and the idea or rationale behind is if they are doing it, then most probably they perform many more measures to prevent abuse in their DNS ecosystems.

On the other hand, an alternative explanation would be that [inaudible] actually could be interested in deploying DNSSEC and signing their malicious registered domain names.

So the third driver that we analyzed was the number of parked domains. So here, our initial idea was to correct for the new gTLD sites. So how does it work? In fact, parked domains do not serve the same type of content as regular ones. They do not run software that could be easily exploited. So on the other hand, we know that parked domains may be used to scan users or to distribute malware.

So the next two features that we measured were no DNS, meaning that the domain did not resolve to any IP address or HTTP error meaning that the domain resolves to IP address, but the corresponding website did not serve any content. So the idea here is that domain starting content are exposed to certain types of [inaudible] and can be hacked. And if those domains do not serve any content, they cannot simply be hacked.

The next feature that we pre-selected was the type. And here, it was the proxy for strict registration policies.

So what we did, we labeled each TLD according to four groups: generic groups, geographic, community and brand of gTLDs. And we assigned numbers to them: generic, one, geographic, two, community, three and brand, four.

So what's the [intuition] here? If we have, for example, .stop gTLD which is a generic TLD, it's much easier than no registration restrictions there [formation] to actually register such a domain.

If we take .pharmacy which represents the community, the new gTLD, then there registrar needs to prove that he or she represents the legitimate pharmacy. So the type of TLD actually corresponds to strict registration policies.

And funny what we did, we also included registry operator, or to be more specific, parent company of registry operators, [inaudible] variable in the model, as the proxy for registration practices in general.

So the rationale behind was that maybe this feature will capture joint registration practices like [inaudible] potentially, registration in bulk, payment methods, and so on.

So before going actually to, and before explaining the results of this inferential analysis, I would like to spend one more slide on our active measurements. So I mentioned, which we found quite interesting also.

So what we did, as you remember, Maarten mentioned that we scanned, we used our active measurement platform to verify if the

domains are parked, if they do not resolve, meaning they are labeled as no DNS, if there is an HTTP error of the corresponding website, or if they redirect to another website.

And this is quite interesting because in legacy gTLDs, apart from the content, apart from the domain that serve content, we see that a quite significant part of domains redirect to other domains. And also, we see quite a significant number of parked domains.

In new gTLDs, the two largest groups, apart from, of course, content are the domains with no DNS report or those who serve on an http error. And this is especially interesting because a previous study from 2015 show that there are as many as 16% of all domains in the new gTLDs that do not resolve to any IP address.

Our study shows that there are as many as 24.2% of all domains that do not resolve even.

Okay, so now let's move to the results from our analysis. What we observed is first up, there is a very weak but positive correlation between the number of domains in a TLD, meaning the new gTLD size, and the number of blacklisted domains.

And this is something to be expected because the great majority of blacklisted domains are actually maliciously registered rather than compromised.

When we take a look at DNSSEC, we see a very weak, positive correlation. And again, this is something that we could expect that the

correlation would be very weak. Why? Because you remember last time, we used DNSSEC as a proxy for security efforts.

But the problem here is all new gTLDs are required to deploy DNSSEC, and in advance, to show the plan, how they are going to do it. So we cannot really treat it as a proxy for security efforts.

Another interesting thing, we see a very weak, again positive, correlation between the number of parked domains and the number of abused domains. As expected, no DNS and domain stuff, do not serve content, cannot be really hacked. And here, we see a very weak, negative correlation. So the more domains that do not serve any content, the less abused domains.

And one very interesting result, somehow also quite intuitive, we see negative, still not very strong, but the strongest in comparison to all other drivers, we see that there is the correlation between the registration strictness and the number of abused domains.

And finally, we did not really find any statistical significant results that would prove that there is a correlation between registry operators and the number of abused accounts. But what you would need to do here is actually to collect a lot more different potential driving factors, of course, such as pricing, so not only we would need to collect, for example, payment methods and other potential driving factors, and collect them actually, more or less, on a daily basis and to find really correlation between those and abused domain names at the domain level.

So I think now Maarten will continue with privacy and proxy services.

MAARTEN WULLINK: Thank you, Marciej.

Okay, let's see. So yeah. So we also took a look at the use of privacy and proxy services in relation to abuse.

And first, a little bit about privacy or proxy services. Well, these services are, by definition, not, I should say it differently. So these services can be used for legitimate reasons as well. So you can use them to protect your privacy, block spam or stop any unwanted solicitations. So there are many legitimate reasons why somebody would want to have privacy and proxy services linked to his or her domain.

And, of course, the bad guys, the criminals that abuse domains like privacy or proxy services for the same reasons, of course.

So in this study, we analyzed how many privacy and proxy services are actually used in domain registration and how are they used. So in order to find these services, we extracted all the registrant information from the WHOIS data we have and we did a keyword search for strings such as privacy, proxy, protect, etc. And combining this with some manual inspection, we somehow [inaudible], we found some 570 privacy or proxy services. There might be more, but these are used so infrequently that they're not worth – they might have like less than five domains registered.

Let's see. So the way privacy and proxy service works, of course, is the registrant information is substituted by information from the privacy and proxy service. Usually, you would hope that if there is an e-mail

address or a telephone number available and you send e-mail or you call the number, that you can reach directional registrant. But there are lots of privacy proxy services where these e-mails just end up in a big, black box somewhere so they're totally unreachable.

So what we did is first, we wanted to have some baseline overview of how often these privacy or proxy services are actually used. So we looked at every new domain that was registered over the study period from early 2014 to the end of 2016 and so the graph you see here is not for abused domains but for all new domain registrations, so abused and legitimate domains.

And what we see here is for legacy gTLDs, there is a fairly stable line. Somewhere in the order of 24% of all [fully] registered legacy gTLDs uses a privacy and proxy service at the moment of registration.

When we look at new gTLDs, we see that there is a lot more variance for new gTLDs. On average, there is some 19% usage of privacy proxy services, but because of the high variance, the average doesn't say all that much in this case.

Let's see. So when we switch to privacy and proxy use for abusive [newly] registered domains – so these are domains that are registered for abusive purpose – we see again that for the legacy gTLDs, the use is fairly stable, which is on average, 5%. There is a spike at the end.

And again, for the new gTLDs, we see that the line is much livelier. There is way more variance in the data for new gTLDs, so in these spikes, in November, I guess, 2014, and another in July 2015. So these

could be caused by single campaigns of maybe spam campaigns for new gTLDs.

JORDYN BUCHANAN: Sorry, I have a quick question about that graph. Is that the percent of domains that are used for abuse only, 5-10% or 5-15% of them have privacy and proxy, or 5% of [inaudible]?

MAARTEN WULLINK: Yeah, so these are –

JORDYN BUCHANAN: Okay, go ahead.

MAARTEN WULLINK: Sorry, these are the percentage of abusively registered domains that use privacy or proxy services at the time of registration.

JORDYN BUCHANAN: Right. Okay, so hardly any of them is the answer.

MAARTEN WULLINK: Yeah, exactly. So basically, the conclusion here would be that yeah, the use of privacy or proxy services by itself is not really a reliable indication of abuse and that the use of privacy and proxy services remains fairly higher for legacy gTLDs than it is for new gTLDs.

Okay, so continuing onto geographical location, part of the study was about mapping the geographical location of domain names. At first, we wanted to use the registrant's location for this, but as we all know, registrant's information is notoriously unreliable. So after discussion also with the Review Team, we decided to use the registrar location instead.

So again, we used the WHOIS data for this and we extracted all the registrar names. The problem here is that the WHOIS data we had is somewhat, I wouldn't say [polluted], but there are lots of different variants of registrar names. So for instance, a big registrar such as Go Daddy might have 50 or 60 different variants of the Go Daddy name in the WHOIS data. So a big task for us to map all these different variants to a single entity, which we could then use to map or to count all the domains registered to these entities.

We use the ICANN accredited registrar list to map the final registrar name to the country, and for the ones that were still missing, we used the manual look-up to find these. And in the end, we found some, almost 6,000 registrars, and together, these recovering 99.99% of all the domains in our WHOIS data.

So when we look at the results, we see that, when we look at the distribution of registrars to countries, we see that the majority of registrars, because of historical reasons, is located in the United States, followed by China and Germany, the long distance. So there is a very long tail following the United States.

These are all the registrars, and when we look at domains registered through these registrars, you would expect maybe that the same distribution. So this table shows, on the left side, we see the new gTLDs, and the number of domains, and the market share. And on the right side, we see the legacy gTLDs [inaudible], and the number of domains and its market share.

So if we look at the left side, we see that the top player here is China with some 8 million domains, and then the share of almost 28% followed by the U.S. and Gibraltar. And Gibraltar, tiny rock in the Mediterranean. The reason why it's number three spot here is because there's one [serving Dutch] registrar [inaudible] that [applies] to the Cayman Islands.

And when we switch over to legacy gTLDs again, we see that the order is different. We see that the U.S. is highly dominant at the legacy gTLDs followed by China and Germany so that the domain distribution for [inaudible].

UNIDENTIFIED MALE: [Inaudible].

MAARTEN WULLINK: Can you still hear me?

UNIDENTIFIED MALE: Yes. Maarten, go on, please.

JORDYN BUCHANAN: Yeah, go ahead.

MAARTEN WULLINK: Oh, okay. So the domain feedback from this table is that the distribution of domains across registrars are different when we compare new gTLDs against legacy gTLDs.

And when we look at abusive domain names – for instance, the SURBL distribution – then we see that some locations really stand out such as Gibraltar again, which for SURBL has an almost 50% count of abusive domains followed by Japan with some 20% and China, 14%, and then it goes down fairly sharp.

And when we look at legacy gTLDs in the bottom table – these are two tables. It might be hard to see because there is little space on the slide for two tables. So the top table is new gTLDs, each country, the number of incidents, percentage per country, and the rate, and the top parties, the legacy gTLDs per country.

For legacy again, we see the United States is a top player followed by Japan and China which is almost similar as in the previous slide. Let's see.

Okay, and when we switch to registrar reputation, what we had to do here was that we had to filter our registrars so they're actually designed to register abusive domains such as domains that are used to [sinkhole].

So we counted the number of incidents per registrar and calculated the percentage of total abuse linked to a registrar. And when we take a look at the SURBL distribution again for registrars, again, we have two tables on this slide. The top part is the new gTLDs, and the lower part is the legacy gTLDs.

And there are also some players that stand out. Mostly, they're the first player for new gTLDs, which is the managing [inaudible] technology registrar, which has almost over 93% of its domains blacklisted by SURBL, followed by some other abusive, bigger registrars take a lot of abusive domains.

And if we take a look at number four, which is [Alt] names which is the registrar located in Gibraltar. We can also see that this one here has also almost 25% of its domains blacklisted, and for the legacy registrars, we did a picture again. This is somewhat different, but we do see some of the same registrars.

For instance, for legacy, we see the [inaudible] technology company at position number three almost 31% of abusive domains. But overall, there are not many overlaps between the new gTLDs and the legacy gTLD registrars.

This chart shows the managing [inaudible] technology company again for abusive domains listed by SURBL and Spamhaus. And what we see here is that there is a very sharp increase in abuse starting early 2016 and then topping off at the end of 2016.

So, what happened here that because of all this abuse, the registrar accreditation has been terminated at the end of, I believe, somewhere

around November 2016 and from that point onward, the abuse drops fairly dramatically again. So, that's clearly physical in this chart.

Of course, we don't know where this abuse moved to because the abuse doesn't, of course, disappear. The criminals just move to other registrars and we don't know where the abusive domain has been moved to if [they have been moved].

Another example is the [Alt] names registrar. We see a chart that shows some fairly high volumes of abusive domain name registrations. We see that for SURBL and Spamhaus, there are also some overlapping peaks, which might suggest the campaigns. The two peaks at the end, they [inaudible] to lots of abusive in the .top and .science gTLDs, new gTLDs.

Interesting is that although we see two very big peaks in the abuse lists for these two new gTLDs, when we take a look at the domain zone files for these gTLDs, we don't see any peaks there. There are no registration peaks, so that suggests that the domain names have been registered over a longer period and then used altogether at a short time, so this concludes our presentation. So, if there are any questions, we'd be happy to take them.

JORDYN BUCHANAN: All right. I have a couple of questions but let me see if anyone else from the Review Team has questions before I take the Chair's prerogative. I see David's got his hand raised, so David, go ahead.

DAVID TAYLOR: Thanks, Jordyn. Could we go back to the slide, which was just before the proxy privacy slide set, which came? I'm trying to remember what my question was now. I didn't write it down. That was it. In the inferential analysis of abuse, where we got registry operator no statistically significant results, I wondered whether A, we shouldn't have something in here or is it not something we can deduce where we got registrar operator because we see that later on with the bad actor registrar, so it's definitely seeing statistically significant results from registrars.

And the second part of that question is where we've had got this issue where registrar is linked to a registry, if that is in the public domain improved that would, to me, be a statistically significant result because then we can see clearly a registry operator is or has a correlation with the abuse counts, and I just wanted to know what everyone thought of that and what presenters think.

MACIEJ KORCZYNSKI: So, if I correctly understood the question, so you were suggesting to eventually exchange, let's say, as a dummy variable in our modeling and putting instead of registry operator registrar, [inaudible].

DAVID TAYLOR: Yes, it was twofold. It was one we've got there no statistically significant results for the registry operator and I just thought well, if we look at registrar operators, then we can see that there is a correlation, so that when I was just looking at the various new gTLD size DNSSEC [path], etc., we do see some in the registrar, so where we see that. I just thought there could be a point included here and then specifically,

which is where we discussed in the CCT Review face-to-face, when we looked at it, there's the discussion where registrars are partly owned or entirely owned by a registry. The whole vertical separation discussion, which we've had, and which was debated a lot of the new gTLD applications, etc.

But if we've got that and that's a public fact, then we would then have a registry operator or operators, which are linked to registrars, which then show that certain registry operators and certain TLDs do show statistically significant results, so I was just throwing the idea out there to discuss.

MACIEJ KORCZYNSKI: I mean, I agree. That's an interesting point, especially that, for example, what we did, we picked up, for example, [Alt] name registrar and we made the registration there and we [inaudible].

Okay, it looks better. Yeah, so there are two things. First thing is like it might make sense, actually, to instead of including registry parent company, maybe child, this is first thing. Second thing including registrars might make more sense, also. Also, because we work with this, we for example picked up the [Alt] name that was very high in I think it was number four in terms of abusive domains for Spamhaus. And we checked the registration policies there. What we could see there is that it's, for example, possible to register at once 2,000 domains from selected TLDs and randomize domains using ZIP codes, random numbers, cities, and so on and so on.

So, there, it actually might make more sense that the attackers would pick up more registrars than registries. So, in this sense, in this case, it's, of course, a good idea to try registrants in the analysis instead of registries. But the more general problem here is that what we wanted to capture it's that the entire, let's say, registration policy of the registry. That probably would make first more sensible at the registrar level.

And second thing, actually instead of putting specific registrars as we call it dummy variable, we should take into account registration policies one by one, like pricing per registrar, per day, or some short period of time, we should take a registration payment – sorry, available payment into account because maybe it's not only the price that counts there, but also the that the way that the attacker, the available payment methods for the attackers, and so on and so on. That should be done at the domain level or at the registry level or registrar [inaudible] also you suggested there.

I'm not sure if that makes sense.

DAVID TAYLOR: Yes, that does make sense and I think it makes sense to look at that because I think there's some very interesting data for us. It's the data, which kind of gets hidden and we don't see it. I think your point you just made there about a registration policy allowing the possibility to register 2,000 domains across various TLDs, I hadn't picked up on that, I'd certainly like to know more about that because that's a very interesting policy to have and one has to begs the question why do you

have such a policy and I'd very much like to see what TLDs that goes across and whether there is a registry operator that is somehow behind that because that seems to me the bad behavior, which we don't want to be seeing, so that you mentioned that's [Alt] names is one of the registrars with that policy. That's something I think we should be considering in looking at.

And did you see the fact that that was happening? Did you have the data where you saw where those domain names were being registered and which TLDs and were those 2,000 or those sort of names generally abusive ones or you probably don't have that depth of data?

MACIEJ KORCZYNSKI: No. We do not have this data but we check if that's possible to collect this data and we would be actually possible using a different API an collection [inaudible] measurement collects any data per registrar per TLD on daily basis or even on an hourly basis. And then we could really link, for example, promotions to domains that were actually registered at that very time. So, this is interesting thing. I'm coming back to you.

DAVID TAYLOR: I was going to say and can we do that? I'd be very interested in the results of that. Is that something, which is feasible now or is that something, which is just not feasible at this stage and it's for a later state? Or is this data you've got and you can run a few calculations?

MACIEJ KORCZYNSKI: No, no. That is not visible now but it's a possible future work because the thing we would need to start collecting all this data and after some time, we would need to perform [inaudible] analysis there. And so far, we're not collecting the data but we, more or less, we check if that is feasible, actually, and that would require some effort but that would be a very, very interesting study. Once [inaudible].

DAVID TAYLOR: So, I'm just following then I'll be quiet. Sorry. Just following that and I'll be quiet. Maybe that can be a recommendation, which we have then, if we can't have that data in this report. We can tie in a recommendation to that because it is not a ridiculous recommendation. It's something, which we [inaudible].

MACIEJ KORCZYNSKI: Yeah, I don't know if [inaudible].

UNIDENTIFIED MALE: [inaudible], David.

MACIEJ KORCZYNSKI: [inaudible], okay, so [inaudible].

JORDYN BUCHANAN: So, this is Jordyn. I [inaudible].

MACIEJ KORCZYNSKI: I partially heard the question and I definitely think that it's cool to be a recommendation because we find, for example, a lot of anecdotal evidence that the [inaudible]. But the thing is we might actually miss some very important details that some registry, for example, offers much lower prices but at the same time, puts a lot of effort to verify the identity of the registrar or, for example, does not offer certain payment methods because that would be convenient for the attackers or does not offer this excellent from the attacker point of view, kind of domain name generation algorithm for him or her. So that's something very, very interesting.

Also, coming back to point before about [Alt] names, I definitely agree. We can just guess what type of clients they're trying to attract by generating randomly domains using the cheapest TLDs and offering such service as domain name generation using ZIP codes or cities and so on and so on. So that's an open question, of course, but everyone can guess what [inaudible].

JORDYN BUCHANAN: Thanks, Maciej. I'm jumping the queue because my question is related to David's. In particular, the discussion around [Alt] names. To the best of my knowledge, [Alt] names is basically the house registrar for a specific registry operator, and so we see pretty clearly that there's a lot of abuse no matter if it's measured by rates or by number of incidents in the [Alt] names registrar, so I guess I would have assumed that that also translated into a high rate of abuse in the Famous Four TLD.

So, then I guess I was surprised to see that you guys didn't find any statistically significant correlation with abuse counts. Do you guys have any similar tables to the ones that you have by registrar by registry operator? At least just see if there's any registry operators that stand out in terms of number of incidents and percentage of domains that were used for abuse?

MACIEJ KORCZYNSKI: For registries? Sorry, Jordyn.

JORDYN BUCHANAN: Yeah, by registry operators. So, you say the registry, there's no [inaudible].

MACIEJ KORCZYNSKI: No, no, no, no. What we did, we just aggregated abuse counts per TLD sub [inaudible] per registry operators. Registry operators, we use only as potentially a variable, explanatory variable. But the other one thing there is the thing that we haven't found statistically significant result, it doesn't mean that the relation is not there. It's just complex and it really needs time and this also this idea we came up while discussing in the [inaudible].

But the reality is it's much more complex simply because the registry operator might capture many more signals than we can potentially imagine. That's why one of the recommendations is instead of putting registry operator there as potentially driving factor, we should actually

distinguish and measure each driver separately and include it in the entire model.

JORDYN BUCHANAN: Yeah, sure. I mean, I think that would be interesting, as well, but there are registry… So, for example, there are some registry operators that believe very much in these sort of restrictive models. So, when you say like the type is negative, has a negative correlation, I know that there's some registry operators that have either all or most of their TLDs have these restricted models. There's other registry operators that believe very much in sort of open TLD models.

So, I guess it's just surprising to me that given that the correlations were shown elsewhere and that we see high incidents in some of the registrars that are sort of house registrars for particular registry operators that we didn't see the correlation, but once again, there's some other tables that might be interesting to see just some of the doing some of the comparisons like you guys did by per registrar on a registry operator basis might be instructive, but I guess we'll follow up in some of our thinking about follow-on recommendations. Thank you.

MACIEJ KORCZYNSKI: Agree, agree.

JORDYN BUCHANAN: Now, before I jump to – sorry to keep you on waiting, Laureen, but there is a question in the chat from Carlton from a while back that I want to make sure gets addressed. Actually, there's two, I think. But

let's go with the most recent one, which is about the overlaps in registrars for new and legacy abuse registrations, is there any qualitative difference in the appointment process?

And Carlton, by appointment process, do you mean is there a difference in how registrars – do some registrars have different processes for becoming accredited with ICANN or do you mean something else by appointment process? Carlton's typing.

So, is there a difference in accreditation process? I actually think maybe does someone from ICANN staff have an answer to that? I think I know the answer but I don't know if anyone wants to weigh in or Maciej or Martin know the answer, feel free to provide that answer. Yeah, Carlton, so why don't we, since it seems like we don't have the answer right here, so why don't we take that as a bit of homework? I think the answer is no. I think ICANN has a really consistent accreditation process across all registrars, but let's get that confirmation.

Laureen, you've been waiting.

LAUREEN KAPIN: Thanks, Jordyn. My question has to do with a reference in your paper on page six where you're talking about URLs used to distribute child abuse materials and your observation that some unique qualified domain names SQZNs can be used to distribute tens of hundreds of images. And I was wondering, it seems like an aside in the paper and there's no other mention of that particular type of abuse, which of course, as I read it, wasn't really the topic of study, but was that something that you actually gathered data on or was that related to prior work?

MACIEJ KORCZYNSKI: Thanks for the question. So, that was more related to our previous work that interesting thing there. I mean, it's a pity. We only could analyze the data from a [hospice] from the Netherlands. We tried the entire INHOPE network to get the data from the entire INHOPE network from [euro] but we didn't manage. But the short answer is yeah, that's our previous work.

LAUREEN KAPIN: I see. So, in this study, there wasn't any data that related to child abuse images as a type of abuse, correct?

MACIEJ KORCZYNSKI: Yeah. That is correct.

LAUREEN KAPIN: Okay. Yeah. I know that certain parts of the community, particularly certain members of the GAC, are particularly interested in this sort of information and study, so I wanted to clarify my own mind whether by way incidentally whether the data you collected for this study included this topic, as well, or whether that was something that was totally separate.

MACIEJ KORCZYNSKI: Maybe one remark here. If we could get this data, like global data, it would be very, very interesting because let's say that in the Netherlands, the problem somehow we manage to solve the problem,

but probably we speculate that we should observe that the attackers simply will move from the Netherlands somewhere else, and this is something that we would really like to study but it's unfortunately, it's possible for a moment because we do not have access to this type of data.

LAUREEN KAPIN: Okay. Thanks for clarifying that.

MACIEJ KORCZYNSKI: Thank you.

JORDYN BUCHANAN: All right. So, there's another question in the chat from Carlton, which is for Maciej, which is do you have a qualified set of data labels and definitions that you work from?

MACIEJ KORCZYNSKI: So, by labels, you mean for example malicious registered versus compromised versus legitimate and so on or some different labels?

JORDYN BUCHANAN: I see Carlton and Brian are both typing. So, yeah, Carlton agreed yes, that sort of label.

MACIEJ KORCZYNSKI: Yes. We have, of course, this data.

JORDYN BUCHANAN: All right, let's see. Carlton may have a follow-up but I see Drew's got his hand raised in the meantime, so go ahead, Drew.

DREW BAGLEY: Hi, yes. So, thanks for the presentation and all the answers. I have a question about some of the data you presented regarding privacy and proxy usage and abuse. It appears that looking at all the various charts you presented, that at the same time that there was, perhaps, an overall abuse on many of those charts broken down by now we're hosting overall abuse, phishing campaigns, etc., that at the same time, there was that dip toward the end of 2016. That's when there is a spike in the use of privacy and proxy services and I was just wondering if you had any theories as to why that may be, that the abuse that still did exist was more strongly correlated with the use of privacy and proxy servers.

MACIEJ KORCZYNSKI: Which chart? Let me go to the… In the slides?

DREW BAGLEY: Yeah. Yeah, some of the ones you just passed by showed there being some sort of dip, I'm sorry, [inaudible].

MACIEJ KORCZYNSKI: Back?

DREW BAGLEY: Yeah. So, here we have for privacy and proxy services – oh, I guess it was for legacy gTLDs. Okay, going off of memory of what I thought I saw in the presentation. Okay, so for legacy gTLDs, it appears to be a spike, and you use the privacy and proxy services, whereas some of those other trends we see related to specific types of abuse – okay, yes. I was thinking of this chart, okay. The usage for abusive newly registered domains, so the use of privacy and proxy services for newly registered domain names, there was a spike at the end of the year and then some of the other trends you showed with regard to specific types of abuse, there appear to have been a dip where, perhaps, there was some sort of campaign maybe in July of 2016 that showed right now off the top of my head.

I'm just going off of what I remember from the presentation. I don't remember which chart, if it was related to phishing or malware hosting or whatnot, but there seem to have been a spike maybe in July in overall types of other forms of abuse and then a dip by the end of the year at the same time we're seeing the spike in the use of privacy and proxy services. Yeah, so here with malware domains in particular must have been what I was thinking of.

MAARTEN WULLINKSEE: Yeah, so Maciej, do you have any idea why this might be the case?

MACIEJ KORCZYNSKI: Once more, Maarten? What [inaudible]?

MAARTEN WULLINKSEE: So, why do we have, for instance, at this slide, for privacy proxy services at the end, we see spikes. An increase in the use of privacy proxy services for abusive newly registered domains. When we go here, we see a decrease for abusive domains for [inaudible].

MACIEJ KORCZYNSKI: Okay. First thing is that those are different datasets. Here what I can say is for sure that those are not related. Those spikes are not related to privacy and proxy services because we manually check them and [inaudible] single campaigns and we could conclude that those are single campaigns because of the registrar names, among others.

And for spikes in privacy and proxy services, we actually, what we tried to find someone who go to evidence by manually analyzing this data and actually to – or as I was saying, find some anecdotal evidence that those belong to the same campaign. But if someone was using actually privacy and proxy services, they were also making sure that other anecdotal evidence is not there. So, they were, for example, registering those domains in longer period. But, yeah, but I'm not sure if that answers your question.

DREW BAGLEY: Yeah, I mean, perhaps, it's all just a bunch of separate campaigns. I just found it kind of interesting that there was that significant increase in the percentage of abusive domain names that used privacy and proxy services at a time when the overall numbers of abusive domain names,

perhaps, were going down because there was that spike in the summer and yet that wasn't the same time as the spike in the use of privacy and proxy domain names, so I guess, perhaps, that's related, yeah, maybe just two different types of campaigns.

MACIEJ KORCZYNSKI: So, the short answer there is that those are a different dataset, so we can, yeah.

DREW BAGLEY: Okay, and then one follow-up question – I'm sure there will be other questions – going back to what everyone else has been speaking about with regard to the common registry operators and correlations with abuse, and I know we've discussed this and how there are so many different factors that go into it. Is there anything in addition to what's already in the paper that you might be able to provide us with so that in the paper, we could at least acknowledge all the different factors that would go into abuse that you already have pointed out such as price, restrictiveness of registration policies, size of TLD, and whatnot, but where we could at least state that with these other variables in mind, that we'll drive all of these things, it still appears there is this common trend amongst registry operators and abuse. Or just something we could state that we could at least better acknowledge that potential common thread, especially in light of with Jordyn has discussed with the fact that some registry operators we know will have common registration policies overall, and so for those, we might be able to say, "Hey, even with a ton of registration policies, it appears other factors

such as, perhaps, price might be distinct factors in driving abuse." Or is there anything like that so we could at least draw some conclusions even though we would have to acknowledge all these other variables?

MACIEJ KORCZYNSKI: Thanks for this question. So, I believe to make systematic study, it would be impossible at this point because simply the collection of the data, we would need to collect the data over time and so on. But also, in the paper itself, there is quite a lot of anecdotal evidence, for example, for pricing.

If we take a look at the top 10, the list of top 10 registries, sorry, not registries, TLDs, that are the most abused TLDs, I think that the [inaudible] of the report, then it's very easy to verify under which a registry operator they are operating.

So, at least some anecdotal evidence about – and, of course, then verifying that [few] or like I'm not sure, for example, 5 out of 10 of those are under the operation of Famous Four. Or if you take a look at those, the list of top 10 TLD, then it's very easy to find the registration pricing at the end of 2016. That would provide some [inaudible] but evidence of the common and registration [inaudible] that provides in this case.

JORDYN BUCHANAN: Thank you. I know this is not present overall statistically because I think at some point we discussed that it was difficult to discern, but anecdotally, do you have any conclusions that may not be in the paper that you might be able to express about the correlation between

resellers and abuse versus registrars as resellers have no direct contractual relationship with ICANN?

MACIEJ KORCZYNSKI: No. It was very difficult for us. We got some data. I'm not sure if it was – yeah, I think that the data was from our first WHOIS data provider. There were resellers, chains of resellers, but we could not determine the order of the resellers. So it's really hard at this point to actually draw any conclusions, for example, about resellers.

JORDYN BUCHANAN: Thanks. All right. It looks like David Taylor has another question.

DAVID TAYLOR: Thanks, Jordyn. To the presenters, having done the data analysis and having been through all of this, is there any specific data that you wish you'd had and which ICANN collected always available elsewhere that would have enabled you to drill down into certain things, some of which we may have discussed?

You've looked at the data yourselves and you've gone through it. Have you come up with something saying, "Oh, I wish we had that data. That would be interesting"? Because I'm wondering whether we could collate that into a recommendation for future DNS abuse studies.

MACIEJ KORCZYNSKI: At the time of the study, of course, we were thinking of many different features, especially when we found that the great, great majority of domains in new gTLDs are maliciously registered. But I think now we would be able to collect [inaudible], and that would include – of course, we mentioned it several times – pricing but also with promotions. But also what would be very interesting in including the future analysis would be actually to verify the correlation with [inaudible]. If actually reactive security measures somehow influence the choice of the attackers, that could be interesting to add on the top of registration policies.

Yeah, that's what I can think of. But apart from that, there's all kinds of interesting data that could be related to a registration [inaudible], including promotions, including the availability of free WHOIS data, a registrar, free DNS services, and so on and so on. But we checked it and we would be able actually to collect this data at this point. Of course, that would take time. Really, I don't think this data actually exists on a daily basis, for example, per registrar. But we could be able to collect this data and potentially perform some follow-up analysis.

DAVID TAYLOR: Thanks.

JORDYN BUCHANAN: All right. I'm going to try to wrap this conversation up so we can have some follow-up discussion ourselves, but I do notice that Carlton had put in some follow-up question in the chat a while ago, which was about whether you could share the labeling that you'd done with ICANN and

your overall data. My understanding is that you've already given a copy of your data set to Drew. Is that right?

MACIEJ KORCZYNSKI: We gave aggregated statistics to Drew. The problem with us sharing the first domains and labels is we would need to actually cross-check with data providers if they would be willing to share this data. On our side, there is no problem. We could share this data, but that would need to be discussed with data providers.

JORDYN BUCHANAN: Sure. So you've just given various forms of aggregation to Drew.

MACIEJ KORCZYNSKI: Exactly.

JORDYN BUCHANAN: All right. My guess is that's probably more easier for the Review Team to work with in any case.

I see both Carlton and Jamie are typing. I'll give them just a second to type and make sure there's no final questions. But I do want to spend a couple minutes before we end – whoa. All right. Jamie provided a lengthy clarification about ICANN's enforcement on resellers. As opposed to reading that out loud, I'll have folks refer to the chat.

In any case, it sounds like ICANN does have some enforcement of relevant policies on resellers through the registrar agreements and

through requirements on the registrars' path through to resellers. But no direct relationship with the resellers.

All right. As Drew points out. I'm going to thank both of you for your time and for this instructive call. I know you guys are doing webinars as well, so certainly it'll be interesting to see the rest of the community's response. Then I think we'll wrap it out, let you guys go, and we'll have a brief discussion of Drew's paper and next steps for the Review Team. Thank you again.

MACIEJ KORCZYNSKI: Thank you so much.

MAARTEN WULLINK: Thank you.

JORDYN BUCHANAN: All right. Drew, you send out a new version of the DNS Abuse paper today or yesterday – I don't remember. Very recently. Do you want to give us a brief synopsis of any changes that you made as part of this last round of revisions?

DREW BAGLEY: Yes. Very briefly, the changes incorporate what was discussed on last week's call. However, I did not change the recommendation at all because there didn't seem to be any consensus on changes that needed to be made to those in their current form. However there was at the

time a proposal to possibly add an additional recommendation. Instead, what I did with that after our group discussion where there didn't seem to be consensus on creating that additional recommendation, I still took those points and incorporated into the body. Basically – do I scroll? Okay, I do have scrolling. Okay. Basically what I did was acknowledge that there may be other forms of abuse identified by the community. I acknowledged that in a few places.

Going up to the top, I changed the introduction of this chapter a bit, and I imagine we'll develop this even better before the final form. But at least before this goes out for public comment I made sure that we are acknowledging that DNS abuse can be defined much more broadly than technical abuse by different communities, but also acknowledge that there's no consensus in those definitions due in large part to the local nature of laws enforcing various forms of criminal activity that may constitute abuse, as well as subjective interpretations, especially with regard to anything that is much more content-dependent and in the fact that these types of abuse may actually fall largely on other infrastructure providers, such as hosting providers, and therefore there isn't necessarily a consensus.

However, I use that an intro to the fact that the ICANN community of course did reach some consensus on technical forms of abuse that they were concerned about prior to the introduction of the New gTLD Program and that informed the safeguards that were developed as part of the New gLTD Program, which therefore consequently became our focus. So that's just a lot more context to that.

Going along those same lines of identifying other forms of abuse and also allowing for our recommendations to apply to other forms of abuse in the future, I acknowledged toward the end – I apologize; I put a sentence at the end of the body of the text that says, "These recommendations may be applicable to curb other misuse of domain names to the extent the community reaches consensus on other forms of DNS abuse." So I put that in place instead of making a completely distinct recommendation that the community reach consensus on these forms of abuse. So that's more so to acknowledge that that's going to organically develop, presumably as DNS abuses are tackled in different arenas, but to say that these recommendations are not to be construed so narrowly in the future to the extent that a consensus may develop.

Other than that, in the body I added more data from some of the findings, such as some of the findings that may deal with intellectual property infringement – the overlap between technical DNS abuse and intellectual property infringement, where we see that the researchers are pointing out with regard to the use of Apple-related trademarks and DNS abuse. I think that feeds directly into what we're concluding and even the recommendations are making with regard to proactive practices that may be undertaken to prevent certain forms of technical DNS abuse.

I added a bit more context about some of the other things – nothing that I would say is earth shattering. With that said, I wrote this very late and I probably have to go back and even fix typos with some of this stuff.

So those are the big high-level things. I know I got an e-mail from Waudo about some stylistic changes I should make. So my plan would be to go back through and clean this up in that way, but not substantively change the recommendations, unless there is any opposition to our recommendation as is and we reach consensus on what they should be like going into the public comment round.

JORDYN BUCHANAN: All right. Thanks, Drew. I see David has raised his hand. I'll point out that Carlton noted in the chat that there's three types of named categories of abuse in the contract that aren't considered technical, so it might be good to explicitly acknowledge those.

Carlton, I just suggest perhaps that the most useful thing to do would be to just drop Drew an e-mail with that information so he would get – or an edit to the paper so he can get that incorporated.

Why don't we go ahead? David's got his hand raised.

DAVID TAYLOR: Thanks, Jordyn. Drew, I noticed in the chat that you asked whether I could recommend some language for that for a recommendation or a potential recommendation, so I'll try to [thrush] that out a little bit and send you an e-mail on that. Or we can speak.

Having been through a [inaudible] amendment, I did skim through this paper and I'm obviously quite keen on the whole DNS abuse side of things. So I think it is fundamentally important.

One of the recommendations I wondered whether we needed to call it out in a very specific way was where we've got abuse rates clearly correlating [inaudible] stricter registration policies. If we're seeing a registry operator that's been identified or we're able to identify it as experiencing significant abuse – we see quite a few of these up there. The ones where there are significant abuse: should we not be recommending a requirement that they either clean up or adopt stricter policies; something along those lines? Is that something which we should consider doing? Because that just struck me and again struck me further listening to the presentation this time as we did in the face-to-face. It seems to be something we should go for, but I don't know what you think. I'm happy to discuss this further.

DREW BAGLEY: Yeah. I guess the problem with the doing that would be that we would just be calling out the registrars or registries that we know about in the sense that there could be, because of all the different variables that go into play with all this, some that we would be missing. So on the one hand, I get your point. If we have this evidence, why not call them out specifically and say, "You need to do this"? I think exposure through transparency to the extent we can describe it is better.

In terms of specifically requiring stricter registration policies, instead, the way the recommendations are tailored – I can't remember if you were on the call when I went over the [inaudible] last week or the week before, but basically I've tried to tailor these recommendations to acknowledge that a free and open Internet is invariably going to have registries with low registration requirements that are very open, as well

as perhaps low prices. In the midst of that, we could still have other practices that they could undertake – it might not change the registration policy – that would reduce abuse, especially when we're looking at the use of Apple trademarks registered by perhaps even the same registrants. There's things like that in the screening process in the beginning that, if they were looking for those thing, that type of abuse would not happen or would happen at lower rates, even if they have open registration policies and low prices, because they were at least scrutinizing things proactively and perhaps stopping obvious offenders and whatnot or [sliding] them or maybe even putting some sort of [inaudible], not allowing those domains to go active right away if they're flagged in a certain way and require manual review. So there are other practices instead. So practices I think we should be recommending, but perhaps not policies that would place very strict restrictions.

But with that said, obviously I am completely open to any sort of language you can come up with or any ideas you might have about how we could call out what we know and what we've seen but not be, I guess, too inclusive to where we're mistakenly only calling out three operators when maybe there's 100 other bad ones and instead what our recommendation is construed as being is sanctioning the 100 other bad ones and only calling out the three.

DAVID TAYLOR: Thanks. We can probably take that offline and discuss it a bit more, Drew. I suppose part of my thinking really just comes back to, if we put an analogy with the trademark world and the PDDRP, when that was put in place to deal with abuse of trademarks by a gTLD operator. We

don't seem to have an equivalent in the DNS abuse world, and there seems to be a heck of a lot more abuse going on in the DNS abuse than in the trademark abuse. So maybe arguably we're saying that, with the PDDRP, having a mechanism in place has helped reduce the abuse of trademarks, and that would be a good thing to have something similar in DNS abuse.

DREW BAGLEY: Yeah. Let's absolutely continue this discussion today or tomorrow.

JORDYN BUCHANAN: All right. Thanks. Since we're overtime, based on the conversation in the chat, I think what we're likely to do is try to reform tomorrow's – is it tomorrow when the Safeguard and Consumer [inaudible] meeting? When is that scheduled for? Is that for 10:00 tomorrow? 10:00 Eastern?

UNIDENTIFIED MALE: That's tomorrow at 1:00 p.m. UTC.

JORDYN BUCHANAN: 13:00 UTC?

UNIDENTIFIED MALE: Yeah.

JORDYN BUCHANAN: Is that the same time as the Competition call?

UNIDENTIFIED MALE: Oh, sorry. You mean the webinar?

JORDYN BUCHANAN: No. I mean the Safeguard and Consumer Trust Subteam call.

UNIDENTIFIED MALE: No. the Safeguard is scheduled just after the Competition Subteam call, so at 14:00 UTC.

JORDYN BUCHANAN: 14:00 UTC. All right. So we'll have an offline discussion, but we'll probably try to turn one of tomorrow's subteam slots into a plenary slot to continue this conversation, since almost all of the interesting topics prior to the publication of our supplemental report are in the safeguards side of the equation.

So we'll work that out via e-mail and send an update to everyone very shortly. It sounds like we'll have to see at least one more revision of Drew's paper before we can finalize it as well based on today's discussion. So look for more discussion on these interesting topics around DNS abuse very soon.

Thanks, everyone, for joining the call today.

[END OF TRANSCRIPTION]