ICANN Transcription

ICANN Whois Privacy/Proxy Abuse Study Findings Webinar

15 October 2013 19:00 UTC

Mary Wong: Hello, everybody. Good morning, good afternoon, good evening wherever you are. Welcome to this Webinar on the Whois Privacy and Proxy Abuse Study that was done by NPL and led by Dr. Richard Clayton who is here with us today to present the findings from the study which he conducted with NPL and a number of his colleagues, some of whom are on this call as well.

Before I hand things over to Richard I'd like to call on my colleague, Nathalie, to walk us through some of the ground rules and how this is going to work.

Nathalie Peregrine: Thank you, Mary. This is Nathalie from staff. So just a few housekeeping rules. If you've joined this Web conference via the audio bridge, so on the phone, your lines will be muted until after the presentation. Once the presentation is over we'll open all lines so you can ask your questions to the presenters directly.

If you are on the Adobe Connect room only, so you're listening to the audio streaming at the moment, you're very welcome to join the audio bridge towards the end of the call to ask your question directly. Alternatively, please write it in the Chat with question at the beginning of the statement so then we can pick the questions out and save them for the end.

Thank you ever so much and over to you, Mary.

Mary Wong: Thank you, Nathalie. And, once again, welcome everyone to this Webinar. First a few words of introduction. As many of you know ICANN has been doing a number of Whois related studies on various aspects of the Whois system for the past few years. This particular study was one of several that was commissioned by the GNSO Council a couple of years ago.

And on this slide we have listed some of the others, some of which are related in some ways to what Richard and his colleagues did for us on this particular study.

Without going through all of them let me just say that the terms of reference for this study as well as the others were developed with the cooperation of the ICANN community through a number of public comment periods and other sessions. And more precise details are - those terms of reference, the other studies their results and other information pertaining to ICANN's work on the Whois system can be found at the Web page that's listed on this slide.

I'll hold further comments until the question and answer session should that be relevant. And at this point I'd like to introduce Dr. Richard Clayton from the University of Cambridge and his team from NPL to present their findings. Over to you, Richard.

Richard Clayton: Hello, everybody. On the slide, now basically for this study I led the study but I was helped by a very fine project team in particular Tyler Moore from Southern Methodist who provided us with data relating to typo-squatting domains and also introduced us to a number of our data suppliers.

Nicholas Christin from Carnegie Mellon who has done a lot of work looking at fake pharmacies and the way in which they advertise and he provided us with a feed of domains involved in that particular activity. And then at NPL Tony Mansfield did - helped a lot with the - did a lot of the experimental design and also did all the statistics for the study, which I'm extremely grateful. And David Hindley who did the project management and kept us all on track.

We started the contract eventually back in April 2012. We put out a draft report a few weeks ago. And the public comment period ends at the end of the 22nd of October.

The - just to get everybody on the same page here in terms of definitions because we have been very precise in this study as to exactly what we mean and this preciseness is exactly in line with the preciseness which NORC did in their study.

Now as we all know when people normally register domain names then the registrant will supply their name and contact details for the registrant specifically. They may - you may also find within the Whois other fields giving other contacts such as admin, billing, technical, zone contact and so forth.

And if those details matched up then even if the particular Whois format that we were looking at didn't actually provide a phone number or an email contact for the registrant then if the address detail - name and address details in some - one of the other fields exactly matched then we treated that as if it had been supplied in the registrant field directly. And as we all know this data is public and is available for anybody to look at on the Port 43 Whois service.

Now when you register a domain you may choose to use what we call a privacy service. And in a privacy service then the registration will show the name of the registrant but the contact details, which will be put into that Whois record, are generic and they're essentially there the contact details of the people providing the privacy service.

Sometimes, of course, the email address is specific to the particular domain name so that they can automatically forward email to the actual registrant. But apart from that you learn nothing about who the registrant is or where in the world they might be located.

In the case of a proxy service then the domain is registered with no details at all of the person who you might naively think of as being the registrant. The registrant is the proxy service and all of the details which you find within the Whois are specific to that proxy service. Again, there might be a customized email which is specific to the person who is actually using the domain. But all of the rest of the details are for the proxy service.

And ICANN's technical view of this is that the registrant is the proxy service but there is a beneficial user of the domain name who is likely the person who has paid for it.

Now the original research hypothesis that ICANN wanted to have people look into and for which we tendered for this particular study was they wanted to know whether or not a significant percentage of the domain names used for illegal or harmful Internet activities were being registered by a privacy or proxy services to obscure the identity of the perpetrator.

Now we looked at that and we thought that that only touched on one part of the problem and that it would be much more useful to, in parallel, also examine the hypothesis that the percentage of domain names use in these illegal or harmful activities and registered via privacy and proxy services is significantly greater than the equivalent percentage for entirely lawful activities, i.e. that the first hypothesis says, "Are bad people using privacy and proxy services?" And the second hypothesis says, "Are bad people using privacy and proxy services more often than not bad people?"

And the final bit of the puzzle was that if people who maliciously registered domains, I loosely call bad people, these malicious registrants if they weren't hiding their entity by using privacy and proxy services maybe that they were hiding their identity some other way.

And in particular what we set out to do was to assess whether or not we could take the contact information from the Whois and if there was a phone

number there could we make contact with the domain registrant by using that phone number?

Not by using anything else so we didn't go and look at Websites to find out whether or not they had a Contact Us page or anything like that. We're just using the information from the Whois and seeing whether or not there was a phone number there which worked to reach the registrant.

And to rather spoil this here's the answers to what we found, which is, yes, bad people who cause domains to be registered maliciously, do use privacy and proxy services.

And - but the second hypothesis, do they use privacy and proxy services more often? The answer is yes, sometimes and no, sometimes. There are a number of legal and harmless activities which we identified and we studied which are also users of privacy and proxy services rather more than average.

And the final question we asked and basically we found that when domains are maliciously registered then one way or another contact details are hidden. So if they're not using privacy and proxy services they sure aren't providing valid phone numbers.

However, many other domain registrants, and we see all the detailed numbers later on, many other domain registrants don't provide working phone numbers either.

So to summarize what we actually did, before we get onto the results, is basically what we did is we got a list of URLs which were being used in various harmful ways, various sorts of activities from phishing to running child sexual abuse image Websites, to running frauds and so forth.

We took these lists of URLs, we picked out the domain names and then we studied only the domain names which were in Biz, Com, Info, Net and Org so

basically ignoring the other generic TLDs and also ignoring all the country code domains. And now in some cases almost all of the domains were in these - in these particular TLDs; in other cases it was a high proportion but by no means all. But we were only studying these five TLDs.

We then looked at the Whois data and used our skill and indeed a program which we developed especially for the purpose to assess whether or not the Whois data was for a privacy or a proxy service. And it if was not for a privacy and proxy service then we pulled apart the Whois data in order to determine whether or not there was a contact phone number present.

And from this we get precise statistics of exact counts, if you like, for usage of privacy services, exact counts for usage of proxy services, exact counts for Whois details which have no valid contact number.

Then for the Whois details where we did have a phone number then we did a sample. We didn't have the budget in order to look at absolutely every single one so we took - so we took a sample of these. Details of all the sampling and so forth are in the report.

So we took a sample from these registrants with a contact phone number and then we made a phone call - or at least our subcontractors made a phone call. We paid attention as to whether or not the registrant appeared to be a business or an individual and we chose appropriate times of day for - to maximize our chances of getting hold of somebody who was either a business or an individual.

And we paid obviously attention to the address they'd given us to where in the world they were and which time zone it was. And then our subcontractors gave them a one-question survey and in the registrant's native language so if they lived in China we talked to them in Chinese. If they lived in France we talked to them in French.

And the one question survey was, "Did you register Example.com?" for the - obviously for the appropriate value of Example.com - for the Whois data which we had pulled.

If we didn't get an answer then - so it was a valid phone number but it rang and rang and nobody was able to answer, alternatively if the person wasn't actually there just at the moment then we rang back again up to four times with a schedule which was designed so we weren't ringing at the same time everyday in order to maximize our chance of getting hold of these people.

And all of the details of this and the details of the flow chart we provided to our subcontractors for making these phone calls and the information we provided for them for the - asking the question are in the report.

Now the - we got a whole series of different sorts of results which we had to classify from our phone survey. Now first of all the phone number needed to be what we call apparently valid which means that it had to have sufficient digits to look like it would be a dialable phone number.

It had to be not all 9s or all zeroes because in many cases people fill in a phone number of all 9s or all zeroes because they're trying to persuade a Web form that this field, which is a required field, they're going to fill in and put in a number there so that they can move on and register their domain.

And the other requirement was that if it was a North American phone number then we checked that the area code was valid, in particular a number of people register domains in North America using a 555 area code because they've watched Hollywood movies and they know that the fake - a fake phone number so we treated that as indeed it is, it is a fake phone number.

Unfortunately it's not possible to do this all over the world so elsewhere in the world we weren't able to be quite as rigorous in checking that the phone numbers were in fact valid.

Now if the number was apparently valid then it was a candidate for being randomly sampled. And if it won the lottery and it was randomly sampled and we made a phone call to it then of course in some cases the number turned out to be invalid and we got a message back from the phone system saying that this number could not be dialed and therefore that was in fact an invalid phone number, we just hadn't realized it earlier.

The next possibility was we rang the phone number and it just rang and rang and rang. And that's kind of an indeterminate position because maybe if we tried more times or more days then we might have got hold of the registrant.

Alternatively we might reach voice mail or somebody answered who said, well, yes I know who Mr. Smith is but I have no idea where he is today or he's in a meeting and we never managed to get him out of a meeting so that's, again, a rather indeterminate result.

The next possibility is that the phone was answered and we said, "Could we speak to Mr. Smith?" And they said, "Who?" And we said, "The person who registered Example.com." And they said, "We have never heard of Example.com. This is a pizza shop. What on earth are you talking about?" Whereas clear that what has happened is that somebody has stolen somebody else's identity in order to register a domain.

And the final possibility of course is that we rang up the registrant or the company if it was a company - a business domain name, and the person answered the phone and said, "Yes of course we registered that domain name." At which point we get a positive response.

And just to summarize that we have - we treat the - if there is no apparently valid we classify that as being no phone number. If it - if we had a message back from the system saying that this number cannot be dialed - it cannot be

completed as dialed then that's a failure. Equally it's a failure if - in the identity theft case.

Obviously it's a success if the person says, "Yes, I registered that domain." But then we have this sort of indeterminate position in the middle which we treat as neither being a success nor a failure because maybe if we'd rung them up to tell them that they'd won the lottery, you know, provided we were convincing enough that it wasn't a scam, and then we convinced them we really - they really had won the lottery then maybe Mr. Smith would have magically come out of his meeting in order to talk to us.

So we're not sure whether or not - what the position is there. And maybe with a different message we'd have got a different result in terms of talking to somebody.

So let's look at some real results. And in particular we'll look at the first work package we did which was looking at phishing. This is fake Websites for stealing security credentials. And these are not just banks but these days people phish email services, online games, you name it they phish it.

And the reason for looking at this specifically is that this splits up into three different groups of data which are - which basically show the results of the whole - the whole of the survey that we did in one neat little package.

Now the source data we used this is typical of the sort of sizes of data we're working with so we had about - nearly 33,000 URLs from the people who make it their business to pass around lists of phishing Websites.

We picked out - from this we ended up with about - just over 5000 domains of which 57% were in the Biz, Com, Info, Net, Org TLDs that we were studying. All of the details of this are in the report of course.

And then we used specialist knowledge, understanding of how phishing works, to split these domain names into three different groups. And the three groups are, first of all, compromised machines. This is where somebody has registered a domain name many years ago perhaps. They're using it to run their business or to conduct their personal life.

And some bad person breaks into the Website and adds some extra pages which are phishing pages. And then they send out emails to attract people to come and visit these fake bank Web pages which are on this compromised machine.

Now clearly the domain name here has been registered by somebody who has no thought of the fact then their site is going to be broken into in the far distant future by some bad person for phishing so they're registering their domain name with a - just using whatever view they have of the world of whether or not they're going to provide a phone number or whether or not they're going to use a privacy and proxy service.

The second group are third party domains. So in this category comes free Web hosting sites, URL forwarding systems, various cloud services and so forth where the criminals buy or get for free a service from this third party and then they use that service for phishing whereas somebody else might use it for hosting pictures of their new baby.

So here again the domain has been registered by a company who has no thought to the fact that it's going to be used by bad people but they're just making their choice of how to register the domain on the basis of their world view.

And then finally the final group, which for this particular week, was what about sixth of the overall total, these are maliciously registered domain names. These are the domain names which look like (Barclays).com but there's an extra Q in there or something like that or alternatively they're just some

random chosen domain name which has been registered specifically for the - for malicious purpose and we can see no legitimate use for these domain names at all.

So when we look at the results from these three particular groups we see some striking differences. So first of all on privacy and proxy usage then the maliciously registered domains are - 31% of them have chosen to use privacy and proxy services whereas the compromised machines, so these, remember, are just innocent third parties who happen to be running an insecure Web server, their decision to use privacy and proxy service are - nearly 25% of them chose to use a privacy and proxy service.

Now interestingly this is slightly higher than the figure that NORC got when they did a survey across all possible domain names trying to work out what the percentage of usage of privacy and proxy services are.

The likely reasons for these differences are that NORC is - because they're looking at all domain names they're picking up a number of parked domain names which will have an impact on the usage of privacy and proxy services. But the difference is interesting that there is this difference there but it's not a huge difference shall we say. And the third parties are very much below average users of privacy and proxy services.

When we tried to make the phone call to people then again we got rather different results. The third parties we managed to get through to somebody 32% of the time; the compromised machines we got through to people 24% of the time and for the maliciously registered domains, rather some surprise, a handful of people actually answered the phone and said yes they did register that particular domain name and that worked out to 1.8%.

However, this turns out to not to be a particularly useful way of looking at it because we're getting a combination of effects here. And these indeterminate results mess up the way in which makes it easy to understand. So the - what

we feel is the best way of looking at these results is to calculate what the percentage chance is that you have no hope at all of reaching the registrant by phone and you know that up front.

I.e. for this particular activity if we add together - using a privacy and proxy service adding in no phone number, adding in calls which failed in the phone system, adding in reaching people who said, "No, I've never heard of that domain," what's the overall percentage?

And here we see the third parties it's just under 50% so half the time for these third parties you have no hope at all of ringing them up. For the compromised machines it's near enough 62%. Sixty-two percent of the time you have no hope at all of making a phone call and reaching the registrant.

And for the maliciously registered domains then 92.5% of the time you have no hope at all of reaching the registrant. And you'll the see the difference between these - the reason they don't - these things don't add up to 100% is because of this intermediate areas where you might or might not succeed.

We didn't happen to but with a different question you might reach somebody. So there's fairly startling differences there and we'll see that continuing when we look at other sorts of malicious registration.

In Work Package 2 we looked at data from a Website called AA419 which collates Websites which are being used in advance fee frauds, which are being used for various sorts of transport scam, fake banks, fake - completely fake banks as in not even pretending to be a real bank but just being a fake bank in order to run some sort of fraud, a whole wide range of different sorts of frauds. There we found 46% of people using privacy and proxy services but nearly 90% impossible to contact by phone.

Unlicensed pharmacies, we all know what we mean by this. These are the Websites which will offer you Viagra without a prescription, that sort of thing.

There nearly 55% of registrants are using privacy and proxy services. And, again, a very high percentage, 91.8%, impossible to contact by phone.

When we looked at - sorry, we didn't look at, we studied the data from child sexual abuse image Websites, then - and this was an entire year's worth of data, just over 800 - somewhere around 800 Websites. Then we found that 29.5% used privacy and proxy services.

Now for this particular category we didn't try making any phone calls. First of all we didn't think that anybody we did get hold of, by chance or whatever, was likely to give us an honest answer as to whether or not they'd registered the domain.

And secondly because we were looking back over a year in terms of our data set in order to get a reasonable sample size, we felt that this was an unreasonable question to ask people as to what the situation was a year ago and maybe phone numbers had changed, all sorts of difficulties with doing this.

However, when we talked to the experts in this area they said that their experience was - and their experience, of course, is in attempting to identify who has been registering these domains with a view to putting them in prison, then when we talked to these people they said that in their experience 100% of all of the contact details were invalid and that people never used their real names or phone numbers. And if they appeared to be real names or phone numbers then they'd been taken out of some directory and they weren't valid.

So basically what we're seeing here is depending on the sort of activity, the sort of malicious activity we're looking at, we're getting different rates of usage of privacy and proxy services but we're getting a fairly consistent result in terms of whether or not we have no hope at all of making a phone call and reaching these people.

We also looked at some legal and harmless categories so we looked at legal pharmacies from which we took a list from the LegitScript Website. These are all North American pharmacies both for humans and for pets. We looked at a group of law firms - an international group of law firms who are associated with each other.

We looked at executive search consultants, headhunters, as we might call them. We looked at whole range of banks, mainly North American but by no means all. We looked at some very large Websites from the Alexa Top 3500 and in particular we picked out the ones which we were studying where we were studying whether or not they had been typo-squatted. We'll come onto typo-squatting on a later slide. But these are the real Websites here.

And we also looked at adult Websites, which are - were mainly adult pornography Websites; there are a few other types of adult Website in this category which didn't necessarily contain pictures. Again, we didn't look at them, we studied the data.

And here, again, you see a wide range of privacy or proxy usage. The legal pharmacies are less than 9% of them are using privacy and proxy services when they register their domains. But the adult Websites it's up to 44% which is higher than the number of the malicious categories that we looked at. And indeed banks, at 28.2%, that's almost as high as the child sexual abuse image Websites that we studied.

If we look at the figures for the impossible to reach by phone here we see variation, legal pharmacies only 24% impossible to reach by phone but the adult Websites goes up to 55% impossible to reach by phone but still well below the 90% that we're seeing for the malicious registrations.

The numbers in the last column is the people we did reach by phone, which I put in for completeness. But there's a caveat here which is a number of these figures are - have quite large error bounds because some of these samples

are relatively small compared with some of the other categories that we looked at.

So the story so far, just to label this, average usage of privacy and proxy services according to the NORC study is 20%. The best - the most average sort of category which we had was the compromised Websites from the phishing category and that was - and we found that to be 25% so it's somewhere in that region.

Privacy and proxy services are being used more than average - more than just average when they're - by the maliciously registered domains. And we found a range from just under 30% up to nearly 55%.

But - and this is important - but some of the legal and harmless activities are well above average too, banks, as already mentioned, 28%, adult Websites 44%.

But the key measure here, the one that really shows what's going on is that if privacy and proxy services are not used then the phone number may be missing or it may be invalid or it may reach somebody other than the registrant.

So when we look at the impossible to contact rates then basically for the legal and harmless categories we've looked at so far we're looking at 24%-62% failure rates. But for the malicious registrations it's very tight range between 88%-92% with perhaps 100% for the child sexual abuse image Websites.

Now we look at a couple more complex data sets where we're getting rather a more mixed message. In particular in Work Package 8 we looked at a list of URLs, which we were given by StopBadware, where these are basically domains where malware may be present. These are mainly compromised Websites but there are some malicious registrations chucked into this list as well.

Here we found 20.4% of the registrants using privacy and proxy services and about 51% not possible to reach by phone suggesting that if these - if this category is similar to the other categories then basically the indication here is that most of this list is compromised Websites because the figures line up with those.

In WP8 we looked at the SURBL list. This is SURBL supplier list which indicates whether or not if you see this domain in an email then SURBL recommends that you don't accept that email into your inbox but either reject it or put it in the spam folder.

These are mainly maliciously registered domains but not always but most of them are. Here we found 44% using privacy and proxy services but the not possible to reach by phone was 58.5%.

Now we have some problems with the statistics on this particular category because it turned out, when we looked carefully at the data, an awful lot of the domains had the same contact phone number suggesting they were all registered by the same person.

And the likelihood - the reason this happens is because of the way in which SURBL constructs this list in that when they start finding one bad domain name they use various techniques in order to find lots of similar domain names and then put them on the list proactively and that causes, first of all, report inflation.

And secondly, it makes the statistics a little bit more difficult for us to work out and we end up with these very high error bounds because depending on what one particular phone call results in, yes or no as to what they do, as to what they - what the result is then that has a big effect on the overall numbers. So these two particular work packages take with a bit of a pinch of salt because they're a little bit more complicated to understand.

I've already talked about the domains which are typo-squatted, the genuine domains which have been - but what some people do is they register domain names which are nearly the same but different in the hope that somebody will mistype the domain name or visit their Website instead and will click on various links and they will make money from the advertising because they've - essentially you've clicked on advertising link.

And the - Tyler Moore, who's done a lot of research in this area provided us a list of typo-squatting domains. And we found here that privacy or proxy services were used by 48% of registrants. But interestingly we actually managed to reach 10% of the - over 10% of the registrants by phone. That's twice the number of the adult Websites that we managed to reach by phone.

But once again, there aren't all that many people doing typo-squatting so there's a relatively small number of registrants and therefore individual registrant's decisions as to how they should register domain names has a disproportionate effect on the results and that comes out in the stats as the high error bound.

We also looked - in Work Package 9 at domains which have been subject to the UDRP process for the - for Com, Net, Biz, Org, Info. And, again, a lot of these are in fact typo-squatting cases; not quite all but most of them are.

And here we found privacy or proxy services being used by just under 40% of registrants. We didn't make any phone calls here because, again, the - our data set went back a year or so into the past. Now clearly what's happening here is that some typo-squatters are trying hard to hide and other typo-squatters are perfectly happy not to hide.

And we speculate that this is because typo-squatting isn't actually a criminal matter but a civil matter and therefore it's a question of whether or not the typo-squatter wishes to hide the fact that they own a lot of domains because

maybe the brand owner will see a - sort of an economy of scale if they can go after one typo-squatter and get rid of large numbers of domains which are attracting traffic away from them. So typo-squatting, again, a little bit more complicated to understand.

It's important to say that we concentrated a lot here on statistical significance. As I mentioned originally the measurements of the privacy and proxy services are exact, exact counts. And for many of the work packages we've got really quite large samples; we looked at over 70,000 domains in total so we think our results are pretty robust.

And the effect of this is that pretty much any variation here over 3% is statistically significant at 90% or better. See the result report for the full details.

The phone calls to the registrants was done on a sample basis and that makes these statistics really rather more complicated. What I will say is that the measure which I've been plugging all the way through, impossible to consider making a phone call, have rather lower error bounds because of the way in which they're calculated and therefore these really are quite a robust way of looking at the data that we have.

A slide here more for the record than anything else, which has the numerical results of the study overall. And I'll move on, in the interest of time, to the conclusion.

And basically the conclusion is - is pretty clear which is when domains are maliciously registered privacy or proxy services are used more than average, no question about that. But some legal and harmless activities also use privacy or proxy services significantly more than average as well.

But the difference here is that when privacy or proxy services are not used but it's a malicious registration then the people doing the registration very

seldom provide valid phone - contact phone numbers. So at least 9 out of 10 of the registrants, you know going in, you're not going to be able to reach them by phone.

But, conversely, many of the people who register domains for lawful and harmless activities also fail to provide valid phone numbers either. And basically anything between about 1/4 and 2/3 of these entirely proper registrants are inherently unreachable by phone.

But equally you wouldn't necessarily just look in the Whois for a phone number for a way of reaching somebody who had registered a domain and was using it for a Website or whatever.

So now we're going to have a Q&A session. There's a link there to the public comment report page. And the - that's open until nearly midnight on the 22nd of October. And we really would love to have comments.

But I would like to encourage you to provide comments about the report and whether or not we - some things in the report are unclear or possibly even wrong or mistaken rather than using this space to provide philosophical comments about criminality, the Whois system or whatever. That's not very helpful. That's maybe helpful for ICANN but it's not very helpful to us; we're looking for comments on the report directly.

And the other thing I have to say to you before we start the Q&A session is that you have to dial in to the audio bridge if you wish to ask a question by voice; alternatively you can type it in.

Mary Wong: Thank you very much, Richard. And thank you to you and your team for a most comprehensive survey with all the findings that you've presented today. Everyone's lines are now open. And if you are in the Adobe Connect you can raise your hand. If you are only on the phone bridge please just say your name and say you'd like to be in the queue.

Are there any questions? Steve Metalitz?

Steve Metalitz: Yes, thank you. I put this in the Chat but it might be easier just to ask it. You've spent most of your time today - and sounds like most of the study was devoted to trying to reach these registrants on the phone which surprised me a bit since that wasn't really the focus of the question I thought you were asked to study which was who is using privacy and proxy services.

But since you did go there did you compare your results with those found by the NORC, which completed a multiyear study several years ago - I think it started about seven years ago for ICANN - six years ago - on that exact question of how - of whether domain name registrants were reachable. How did your results compare with theirs? Is the problem getting worse or were the results consistent or did you not look at them?

Richard Clayton: We've not really compared our data directly with that. It's - in some ways it's hard to compare the two studies directly. There are some subtle differences in the methodology which mean that comparisons are not as simple as you might make out.

As I indicated when just looking at the usage of privacy and proxy services where we got different numbers it's rather difficult to see how those numbers come out. And therefore we know they did and we think there's - just our way of selecting domains which essentially means that the domains have - effectively have to be - are being used for a real running live Website because all of the categories come down to that in the end. It just makes a difference when sampling all domains.

And therefore comparing stuff across the two isn't quite apples and oranges but it's - it's more like apples and pears.

Mary Wong: Thanks, Richard. And, Steve, I guess - I can add that, you know, one of the things that these results show and in context of the work that ICANN is doing is really looking at the question of contactability or otherwise of registrants who may use a variety of ways to avoid contact or detection.

James, you had some questions in the Chat but you've raised your hand so I assume you're going to ask those questions of Richard at this point? Please go ahead.

James Bladel: Yes, thanks you. James Bladel for the transcript. And I apologize, I wasn't aware that we were allowed to speak so I kind of filled up the Chat box with questions. So I'll ask the first one and then drop to the back to the queue just to ensure that that's fair.

But my first question - well first I guess I should thank the presenters, Richard, and the other folks for this very comprehensive piece of work. And my question is, even following the script that was laid out in Appendix A I'm still not entirely clear on the scenario when a proxy service provider was listed as a registrant for a domain name and the telephone number associated with it was answered by someone who identified themselves as a representative of that proxy service.

But...

((Crosstalk))

Richard Clayton: We tried - we were in fact trying not to ring up the proxy services. If you - this may not be entirely clear in the text but basically if we believed that they were a proxy service and somebody who call themselves Acme Proxy Services, we would deem to be a proxy service. And certainly we did some other due diligence to see how many other domain names were registered to them and that sort of thing. If they appeared to be a proxy service then we did not ring them up at all.

James Bladel: And how did that - how did those names then - how were they dispositioned in your statistics? Was that a no contact or was that an uncategorized?

Richard Clayton: No, no that - that goes into the category of privacy or proxy service or in this case a proxy service.

James Bladel: Okay so just so my understanding - and I'll drop my hand here - is that you were not necessarily testing whether or not proxy services or different proxy services had different rates of responsiveness?

((Crosstalk))

Richard Clayton: ...no...

James Bladel: ...when their information was used as the registrant?

Richard Clayton: Not at all. As soon as we had identified that this was a privacy or a proxy service then we just counted that and moved on. It's only the ones which were not privacy or proxy services where we attempted to make a phone call.

James Bladel: Okay. Thank you.

Mary Wong: Thank you, James. Thank you, Richard. And, again, James, that goes to the general question of contactabilty or lack thereof. Greg, you have your hand raised?

Greg Shatan: Hi. This is Greg Shatan. Yes. I observe, as a general matter and having read the draft study, that while the study was initially commissioned to look at a number of different forms of intellectual property infringement that in fact the study in the end only viewed, in essence, one sort of intellectual property infringement, specifically typo-squatting, and did not follow through on the

mandate to review software piracy, media piracy or cyber-squatting outside of typo-squatting.

And I was - and specifically with regard to typo-squatting the study says that you believe or, you know, that you consider typo-squatting to be far more prevalent than other forms of cyber-squatting.

I was wondering, first, how you came to the conclusion that typo-squatting was more prevalent than other forms of cyber-squatting and domain name-related trademark infringement? And why, more generally, you seem to shy away from intellectual property infringement as a part of the study? Thank you.

Richard Clayton: Well, the basic reasons here are - we do actually - no, we have a couple of pages on this. And, in fact, you'll find similar material I think published by ICANN because we submitted this as possible (alternatives) and so forth when we were negotiating what we were actually going to study.

The difficulty in this area is that there are very few lists of domain names which are involved in these things. So if you don't have any data and you can't particularly see how to get any data then it's rather difficult to study. So that was one area.

The second thing which is that we felt that in practice media piracy - that what was in the Whois was almost certainly irrelevant to the way in which people dealt with it because if you're going to set up a Website to live stream the England Poland football match tonight without authority then the easiest way of finding out who's doing that is to work via the hosting company because of the bandwidth requirements and so forth. You wouldn't necessarily worry too much about what the domain name - how that was registered.

Now that may be our naiveté in terms of how we dealt with this. But, again, there aren't very good lists and certainly not very long lists of these sort of sites.

And the other thing is, of course, the - we did study in WP9 all of the domains which have been subject to the UDRP which of course covers that whole range of different sorts of intellectual property infringement.

Greg Shatan: And how did you determine that typo-squatting was more prevalent than other forms of cyber-squatting? That's not my experience.

Richard Clayton: Okay, well essentially that was our view. That was Tyler's view when we talked to him because he provided this data. If we're wrong then - then we're wrong, I'm sorry.

You know, as, again, the nice thing about typo-squatting from our point of view was it was relatively straightforward to generate the sort of data sets that we needed to study this whereas a number of the other areas which were mentioned and not just the intellectual property areas but some of the other ones we really did not feel that it was practical to study them because it would be a major exercise in itself to obtain anything like a decent sample size.

Mary Wong: Thanks, Richard. And thank you, Greg, for the question. And, Richard, I believe you're referring to the research done by Dr. Tyler Moore and Ben Edelman, which is referenced in the study.

Richard Clayton: Yes indeed.

Mary Wong: And I see that Lisa has put a comment in that to the effect that some of these questions and any follow ups you may have would be excellent public comments to put in so that we can take them back and analyze them and do some follow up as appropriate.

(Adam), you have your hand raised. Please go ahead.

(Adam Scobo): Hi, this is (Adam Scobo) from Re/Max Real Estate. I have a question about - and perhaps this is an easy question with respect to categorization. But I'm just sort of trying to wrap my head around the categorization of the compromised Websites.

It seems like they - in a way they almost fall into one bucket from one point of view and the other bucket from the other point of view in terms of that in terms of the registrant's use of legitimate, you know, contact information I think that what seems to be reflected in your comments is that they're more like legitimate Websites because - legitimate domains.

Because, to some degree, there's a legitimate person who at the time of registration actually registered it and presumably is somewhat like the rest of the population in terms of their use of proxies or their use of real information.

Although one might also imagine that - I think you saw a slightly higher number than the NORC study and that might be because these are, by definition, folks who may be a little bit lax about their security and a little bit more sort of smaller business or likely to use a proxy service from that point of view.

But so from that point of view they may seem like the general population. On the other hand, if you're looking at it from the perspective of these harmful or, you know, bad Websites, how likely are they to use a proxy service did you treat those as domains where they weren't using a proxy service?

Because, in a way, by definition of the activity of the scammer having compromised the Website and gone in it seems - this is - from their point of view that's another way of using a proxy service. They're coming into someone else's domain where, you know, partly so that they won't have to register their own domain where they could possibly be tracked down.

So can you speak a little bit to the categorization of those from those two sort of points of view if that makes sense?

Richard Clayton: Well, yes. Essentially different phishing criminals do different things. Some of them use free Web hosting where the domains are, you know, one of the domains that we looked at was blogspot.com because somebody had put a phishing Website on blogspot.com.

And in fact we actually made a phone call to Google's legal department and said, "Did you register blogspot.com?" And they answered the phone and they said, "Yes." So that was one of the data points from our study.

So yes from the point of view of the criminal wants to hide then clearly, yes, there is a - the ones who choose to compromise Websites are choosing a different - I am not going to get caught - strategy from the people who register Barclays with two Qs at the end as a - as a bank name Website and then go and phish Barclays from that particular Website.

But I might say in fact for the period we studied most of the phishing Websites that were malicious registrations were in fact for online games such as World of Warcraft where they register long domain names with lots of words involving (unintelligible) and hyphens and US and that sort of thing in the hopes of fooling the people who play online games.

And so, yes, they're adopting different strategies. That's - the key thing here, which is why we split them into - we split Work Package 1 into three separate sets of data and presented and three separate sets of data throughout is because these are different strategies by different sets of criminals with different risk profiles in terms of whether or not they think they're going to get caught and how much they need to hide.

(Adam Scobo): I guess my question is - is this - and I forget the - how you phrased the premise that you were asked to test and the premise you then also additionally tested. But from the point of view of how many people who run legitimate - who run, you know, scam Websites, what percentage of them are using privacy and proxy services?

You know, you could look at that and say okay well here's a domain name that has a scam Website and here's - and it is not using a proxy service therefore, you know, the scam Websites actually have a relatively low percentage of privacy and proxy usage.

But it seems to me if that is the interpretation that would be made from the categorization that would sort of seem to me to be somewhat erroneous in terms of how to categorize those because they are. And maybe those are sort of, you know, a neither yes or no sort of category. That's sort of the question that I'm going after is...

((Crosstalk))

(Adam Scobo): ...were they sort of, you know, put in the wrong bucket because of that sort of...

((Crosstalk))

Richard Clayton: Very much the assumption - very much the assumption we made was that if we found a phishing Website for Barclays on momandpopshop.com and we looked at momandpopshop.com and it appeared to contain details of some small Kansas community store, then we took the view that the people who owned the Website were innocent bystanders and had just been insecure in terms of looking after their things.

And, therefore, we didn't treat that as being a malicious registration because when they registered momandpopshop.com they were not being malicious,

they were trying to register a domain name for a legal and harmless activity. And that's the - that's the distinction we're making. Which the distinction we're making is was this domain registered for a malicious purpose?

((Crosstalk))

Mary Wong: Yeah, and thank you for your question and comments, (Adam). The report does describe how the team split the three categories for this particular work package and to some extent some of the manual processes that we used that - to determine some of the borderline cases.

We're at 4:00 pm so as I've typed in the Chat if you have a question and you're in the Adobe please raise your hand. If there's anyone on the phone bridge who is not in the Adobe Chat room please let us know at this time if you have a question for Richard or anyone on this call.

Seeing no hands - oh wait, hold on. Amr, you can have the last question since I hear no other voices on the phone. And, Richard, if you take this question and make some concluding comments if you'd like we can then end the Webinar shortly thereafter.

Richard Clayton: Okay.

Mary Wong: Amr, go ahead.

Amr Elsadr: Thanks. This is Amr. I was wondering about the examples for domain names that were registered for lawful purposes. The samples you selected were basically - seem to me to be basically either individuals or businesses - businesses like banks, online pharmacies, adult Websites.

I'm wondering why you did not choose any lawful organizations that were not businesses or that were not commercial or profit-driven and whether that would have affected the - proving the truth or false of the hypothesis of the

study? I'm talking about things like maybe political parties, NGOs, activists Websites, that sort of thing.

Richard Clayton: Well, I understand the question. The original aim we had was that the lawful and harmless, in some sense, mirrored the malicious things that we were looking at so we were comparing adult Websites with child sexual abuse image Websites. We were comparing real banks with the fake banks and so forth. It got a bit blurred in the end because of that.

One of the difficulties the whole way through in that area was identifying lists of domains which were within Com, Net, Org, etcetera, etcetera, and also which were reasonably long.

There are a number of political parties, particularly in some countries, getting 200, 300 domain names would have been a bit of a challenge for some of the categories you mentioned. And there were no particularly obvious categories. We fell back on the Yahoo Directory for a number of these things which was in some ways a little bit unsatisfactory because of the nature of the (curation) of that part of the Web these days.

So basically we weren't trying to study all lawful and harmless; NORC did that. And what we were trying to do was to give some example categories where we felt that there was some reason to believe that people would either shy away from privacy and proxy services or go after them rather more enthusiastically.

Yes, there are other areas we could have studied and I'm afraid we didn't in this particular case.

Mary Wong: Thanks, Richard.

((Crosstalk))

Mary Wong: And - was that a follow up?

Amr Elsadr: Yeah, I just wanted to be clear on whether you think this would have affected the testing of hypotheses or not?

Richard Clayton: I think that we got enough data to demonstrate what we thought going in might be the case which is that there is wide variation in use of privacy and proxy services between different types of criminal and there's wide variation between different types of legal and harmless activity. And what we've done is we've shown there is variation here.

And having more categories would just say well - and some of these are more like each other than others. It's the variation we wanted to demonstrate not to produce some sort of overarching theory as to who chose privacy and proxy services one over another; that's for a different study I think.

Amr Elsadr: Okay thanks.

Mary Wong: And as Richard said, I mean, one of the aims of this study was not that latter objective but really to demonstrate with some empirical data. And given what they've done the figures do show some comparisons that are very useful and does contextualize the hypothesis to a very large extent.

So since there seem to be no further questions I'd just like to take this opportunity to thank you all once again for joining this Webinar. The slides as well as the recording will be made available shortly.

I'd like particularly to thank Dr. Richard Clayton and his colleagues for the study and for doing this Webinar. And as this slide says, please do submit your public comments. They will be very helpful to us in finalizing not just the study but to ICANN in developing further work on the Whois system.

So with that thank you all very much. Have a good day. Have a good evening. And thank you again.

END