[MUSIC PLAYING]

MILES PLANT: Hello, and welcome back from lunch. My name is Miles Plant, and I'm an attorney with the Federal Trade Commission's division of Privacy and Identity Protection. I'm here to moderate the next PrivacyCon 2021 panel, which is on advertising technology.

Today's panel on ad tech consists of three speakers, each of whom will be presenting their research on different areas in the advertising-technology space. First, we will hear from Imane Foyad, who recently earned her doctorate in computer science with a focus on detection and measure of web tracking from the Centre Inria Sophia Antipolis. Imane will be presenting her paper, which is entitled, Missed by Filter Lists-- Detecting Unknown Third-Party Trackers With Invisible Pixels."

Next, we will hear from Janus Varmarken, a graduate student earning his doctorate in network systems at the University of California Irvine. Janus will be presenting his paper, which is entitled, "The TV is Smart and Full of Trackers-- Measuring Advertising and Tracking on Smart TVs."

Our third speaker is Miranda Wei, a graduate student at the University of Washington. Miranda is earning her doctorate in computer science and engineering with a focus on security and privacy. Miranda will present her paper, which is entitled, "What Twitter Knows-- Characterizing Ad-Targeting Practices, User Perceptions, and Ad Explanations Through Users' own Twitter Data."

After each presentation, I will pose a follow-up question to the presenter. And actor all three presentations, we will have a general discussion on the future of advertising and technology. And then we will try to take some audience questions as part of that conversation. With that, I'll turn it over to our first presenter, Imane.

IMANE FOUAD: Thank you for the presentation, Miles. Hello, everyone. I'm Imane Fouad, and today, I'm happy to present to you part of the work that I've conducted during my PhD on the detection of unknown third-part trackers with invisible pixels. This work was done at Inria under the supervision of Nataliia Bielova, Arnaud Legout. Next slide, please.

Let's start with this simple example. Let's say that you want to build your own website, and so you open your browser, and you go to the [INAUDIBLE] to learn more about [INAUDIBLE]. When you visit the website, the website will include content directly provided by the website itself but will also include content provided by other domains that we refer to as third-party domains.

Thanks to Disconnect, we can visualize the domains that are included in the website, and so when we visited [INAUDIBLE], we found that additional online domains are included in [INAUDIBLE] and can potentially track your activity on their website. And so what is web tracking? Next slide, please.

Web tracking is a technique that allows its operator, in our case, third-party domains, to recreate a user's browsing history. So coming back to our example, using web tracking, the domains, including in [INAUDIBLE], can track your activity within the website. So they can track repeat visits to the website. But they can also track users' activity across different websites, and so they can record users browsing history, or what we refer to as a user's profile.

And so the question now is how these trackers are able to recognize or identify a user across different websites. Well, we have several techniques that allow the identification of a user across websites. In the context of this work, we focus on safer tracking that relies on the browser's storage.

And so what happens is that the third-party tracker associates an identifier for the user and then [INAUDIBLE] part of the browser storage, let's say, in a cookie, for example. And so next, when the user visits a third party-- or website that includes the content from the third-party website, the browser automatically sends the request to the third-party domain to fetch the content. Along with the request, the browser automatically attaches the cookie with the corresponding identifier. When the third-party domain receives the identifier, it recognizes the user and tracks her across the websites.

In practice, tracking is much more complex than that. Next slide, please. And so the goal of our work is to uncover these tracking techniques deployed in the web. I will not go into details of the methodology, but I'll just give a high-level overview.

So as a result of our analysis, we identified six categories of tracking behaviors that goes from the analytics behavior, which is not that harmful for the user's privacy. In fact, analysis behaviors only enable tracking within the same website. It is useful for the website owner in the sense that it allows them to get statistics about their websites, to know which pages are most popular, and number of visits to their websites.

The most-known analytics service is Google Analytics. Additionally, the [INAUDIBLE] tracking that relies on cookie synchronization-- so we've seen earlier that each third party's able to recreate part of the user's browsing history, depending on the websites, where this third party's included. And so every third party will create its own profile, or browsing history of the user, that is not necessarily similar to the other third parties.

And so, ideally, these third parties will want to merge all the data collected about the user. The way to do that is to deploy cookie-synchronization techniques. As I said earlier, I will not go into details of each tracking technique, but you can always refer to our paper to read more about the six categories, and I'll be happy to exchange with you further if you have any questions. I have to mention that, as a result of our analysis, we found that at least one type of tracking is found on 92% of the domains that we have visited. So in our work, we studied over 100,000 pages.

So to recapitulate, what we've seen so far is that when you visit a website, the website may include content from a third party. So the browser will send a request response to the third party to fetch the content. The third-party service will use this request response, set a cookie on your browser, and deploy different and sophisticated tracking techniques to create your-- recreate your browsing history.

And so as you might all know, we have a number of browsers' extensions that are built to help protect your privacy. How efficient they are-- well, next slide, please. To answer this question, we evaluated the efficiency of the most known browsers' extensions. We studied Adblock, Disconnect, Privacy Badger. And Ghostery. We found that Ghostery is the most efficient among these browsers' extensions. However, it still failed to detect over 24% of the tracking requests that we detect.

We've additionally studied the combination of all four browsers' extensions. And so we studied the efficiency of these four browsers' extensions combined. And we found that even combined, these four browsers' extensions miss over 20% of the tracking request that we detect. Now, to have a better understanding of the efficiency of the existing tools, we evaluated the efficiency of the filter lists that are the core of the browsers' extensions. Next slide, please.

So we studied EasyList, and EasyPrivacy and Disconnect, that are not only the core of main browsers' extensions, such as Adblock, Adblock Plus, [? uBlock ?] Origin, and Disconnect, but they're also widely deployed in the research community. EasyList and EasyPrivacy. for example, in the last few years, was deployed by 16 papers published in top conferences tech trackers. Disconnect, on the other hand, is the core of the Firefox-tracking protection.

So we evaluated the efficiency of these two filter lists. Next slide. And we found that both EasyList, EasyPrivacy, and Disconnect filter lists miss over 25% and 30% of the trackers on 69% of the websites that we have visited. And so here, the main question that we had is why these filter lists are missing this big percentage of tracking requests. I mean, 25% and 30% versus obviously a big percentage of tracking requests.

We found that one of the main reasons behind the fact that filter lists miss trackers is that these requests missed by EasyList, EasyPrivacy, and Disconnect are deployed with functional content. So by looking into the content type that you served with this request missed by these filter lists, we found that it was mainly functional content, such as script, big images, and HTML files. In the following, I'll explain in detail what we mean by functional content and how this fictional content can be used for tracking purposes. Next slide, please.

So let's take this example. Let's say that visit google.com or just open your browser. If you're using Chrome, for example, then google.com will be your page by default. So if you just open your browser, you'll be directed to google.com. And so you're forced to visit the page google.com.

What happens is that google.com, which is the first-party website in the context, is the website with which you are directly interacting. We store a cookie on your browser, let's say an ID cookie with the value 1234. So according to a Google policy, an ID is an advertising cookie. And here, we assume that google.com associated and identifier 1234 to [? its ?] user.

So in that context, the cookie stored by google.com is a first-party cookie. And here, it's not problematic, let's say. Next, please.

Now, let's say that the user revisits [INAUDIBLE] website. In fact, [INAUDIBLE] will have or include content from google.com. Why? Well, [INAUDIBLE] include the Google-search engine, which is the functional content in [INAUDIBLE], as it allows you to make a search within the website. So in the context [INAUDIBLE], the Google search-engine content is useful, or functional, or what we call functional content.

And so what happens now is-- next slide, please-- the browser will send a request to Google to fetch the content from Google search engine. Along with the request, the browser will automatically attach the NID cookie with the identifier with the value 1234 that is already stored in your browser. Google.com received the request with identifier 1234, and it's going to know that the user with the identifier 1234 who was visiting google.com earlier is now visiting [INAUDIBLE]. It's all on the server side.

In the example, Google is recreating your browsing history, or what we refer to as a user's profile. And so clearly, the request sent to Google is a tracking request. However, this request can not be blocked by the filters lists or a browser's extension because it's used to serve, or to bring functional content to the [INAUDIBLE] website, which is the Google search engine.

Now, to recapitulate what we've seen, you visit a website. The website stores a cookie in a first-party context. Next, you visit a different website that has content, functional content, from the same website.

And so what happens is that the browser will send a request to [INAUDIBLE] content and will automatically attach the cookie that was first stored in the first-party context. But here, for example, in the [INAUDIBLE], the cookies used in third-party context. And using this request, the third-party domain can recreate part of your browsing history. However, the browser's extensions are not able to block this request because it's serving functional content to the website. In conclusion, we found that filter lists do not block tracking requests sent to Google search engine on 329 different websites. Next slide, please.

Now to conclude, in this work, we have uncovered new tracking techniques. We've also shown that the state-of-the-art tracking-protection mechanism, such as browser's extensions, miss over 25% of the tracking opportunities that we detect. We've also found an explanation of why these tracking requests are missed, and we showed that it's mainly due to the usage of functional content with tracking requests, which leads us to the conclusion that between protecting our privacy or keeping the functionality of the website, we clearly need a more fine-grained approach to the tracking techniques. Thank you for your attention, and I'll be happy to take questions.

MILES PLANT: Thanks, Imane. I think it's particularly helpful you focused on how Google integrates this tracking across different sites and cookies. And I was curious if you could talk a little bit about how you think Google's efforts to deprecate third-party cookies will impact tracking behaviors and how we detect them.

IMANE FOUAD: Thank you, Miles. That's a great but also, a fundamental question. I will give you a high-level answer. And if you have any technical questions, don't hesitate to get in touch with me. I'll be happy to exchange further.

So let me explain how the dep will impact tracking techniques. My short answer will be that even after deprecation, a significant number of tracking will be still possible. Now, to give even the most likely detailed answer, after the removal of third-party cookies, first-party cookies can still be deployed to track users' activity and recreate her browsing history, thanks to the techniques that we detect in our work.

And so the [INAUDIBLE] idea is that the first-party cookies will not be blocked. And in our work, we uncovered that first-batch cookies are shared with third parties and synchronized with the third-party cookies. Therefore, after removal of third-party cookies, the third-party trackers can still rely on these first-party cookies to track users' activity across websites.

Additionally, in our new work that is under submission, we even show that you can still perform cross-site tracking using first-party cookies in a persistent way, meaning that even if the user tried to clean her browser and delete all the cookies, including first-party cookies, the tracker can actually link a user's activity before cleaning the browser and after cleaning the browser and track her activity across websites, relying on these first-party cookies and without usage of third-party cookies.

MILES PLANT: I'll be really interested to see that follow up research then, especially as we see this evolution of this space. And I look forward to hearing more from when we have our conversation at the end.

IMANE FOUAD: Thank you, Miles.

MILES PLANT: Thank you. I am going to turn to our second presenter, Janus Varmarken. And he will be presenting his paper on smart TV.

JANUS VARMARKEN: Thank you, Miles, for the introduction. So I also want to start by saying thank you to the FTC for putting this conference together every year. I think it's a great initiative.

Today, I want to give you a brief overview of our work on measuring advertising and tracking on smart TVs. And this is a joint work with my great coauthors, Hieu Le, Anastasia Shuba, and Professor Athina Markopoulou, and Professor Zubair Shafiq. Next slide, please.

So this work was motivated by the fact that smart TVs are now widespread in the US with at least one smart TV in four out of five US households. So naturally, you can see that this represents a platform with a very large audience for targeted advertisements. And the smart TV is uniquely positioned in terms of tracking, as it can literally observe your viewing habits and then base the ads of of this information. Furthermore, smart TVs were found in a paper in IMC 2019 to be the IoT device to contact the most third parties, which also suggests that a significant amount of tracking is going on in this platform. Next slide, please.

So with this work, we wanted to answer two research questions. First, we asked, what does the Advertising and Tracking Services, or ATS for short, ecosystem of the smart TVs look like? And to answer this research question, we take a network-measurement approach, meaning that we look at what network traffic these smart TV devices generate.

And we looked both at traffic generated from smart TVs used by real users and smart TVs that we instrument in a test-bed setting. Now, due to time constraints, I will focus this talk on the results from our test-bed setting. Second, we also ask, how effective are existing privacy-enhancing tools? And here, we analyze the effectiveness of DNS-based blocklists, as these are universally applicable across all smart-TV platforms.

For anyone in the audience who might not be familiar with these blocklists, you can think of it as a tool that you can deploy in your home networks to prevent your devices from contact with certain service on the internet. And we also use these four blocklists that I list here in the lower-right corner, to label domains in our network-traffic data set as advertising and tracking services, or just regular domains. And we used this labeling throughout our data analysis. Next slide, please.

Now, to answer these research questions, we needed a data set with network traffic from many, many smart-TV apps. So to automate the generation of such a data set, we instrumented through smart-TV platforms, namely the Roku platform and Amazon's Fire TV platform. And by instrument, I mean that we wrote some software that automatically interacts with the apps of these two platforms in a way that mimics how a real user would interact with the apps. We picked these particular two smart-TV platforms because they are the top two-most widespread in the US, and they're also the top two in terms of number of ad requests sent.

Next, I will present our instrumentation tools, how they were designed, and some insights from the analysis of the network traffic we collected using these tools. Next slide, please. We instrument the Roku platform by configuring a Raspberry Pi as a routed wireless network, and we then connect our Roku device to this network. So this means that all of the traffic from the Roku flows through with the Raspberry Pi.

We then run our instrumentation software, which we call "Rokustic" on the Raspberry pi. And what Rokustic does is it runs tcpdump, which is a tool that captures all the Roku's network traffic, and then it also sends virtual key presses back to the Roku device in order to interact with the app on the test. Now, one thing that's important to note here is that the Roku operating system does not allow us apps to execute in the background. So this effectively means that all the network traffic that we locked during interaction with a particular app will pertain to that app or the operating system.

So with this setup, we could test apps in very large batches. We did as many as 500 at a time in one go. And we're only limited by the storage capacity of the Roku device. Next slide, please.

Our instrumentation tool for Fire TV, which we call "Firetastic," is a combination of existing tools. So here, we use [? A ?] Monitor to [? block ?] network traffic. [? And A ?] [? Monitor ?] is a VPN-based traffic-interseption tool from some of our prior work. [? A ?] [? Monitor ?] runs directly on the Fire TV and locks all the network traffic going in and out of the Fire TV. And it also has capabilities to label each and every packet with the responsible app.

Furthermore, [? A ?] [? Monitor ?] can actually decrypt traffic, as well. So by doing so, we get additional insights on Fire TV that we can't get on Roku. To automate interaction with the Fire TV apps, we use a tool that's called TRiBot. And what TRiBot does is it maps the user interface of the Fire TV app and then explores this user interface in a strategic manner.

And since TRiBot can run on one laptop and control multiple Fire TV devices in parallel, we can actually paralyze our testing across multiple Fire TVs here. So with this setup, we managed to test 1,000 apps in just one week by paralelizing testing across five Fire TVs. Next slide, please.

So to ensure that we collect traffic for the most relevant apps, we determined the top 1,000 apps of the-- for the platform and 1,000 apps for the Fire TV platforms-- Fire TV platform-- sorry-- by calling the respective app stores of these two smart-TV platforms. We then used the Rokustic and Firetastic to test each and every one of these 1,000 apps for approximately 15 minutes each. And the resulting network traffic data set contains approximately 2,200 unique domains across all of the Roku apps and approximately 1,700 unique domains across all of the Fire TV apps.

We then used the four blocklists that I introduced briefly earlier to label all of these domains as advertising and tracking services or just regular domains. And the interesting thing we saw here was that when we only looked at the ATS domains, we saw that there is actually a large number of domains that are only present on the Roku platform and a large number that's only present on the Fire TV platform. And this is even the case if we reduce these ATS domains to their effects second-level domains. You see that there is still a large fraction of domains that appear to be exclusive to one platform. Next slide, please.

We also wanted to identify the key players of the advertising and tracking services' ecosystem of Roku and Fire TV. So what we did here was we determined the top 20 most prevalent third-party ATS domains and then identified the owner, or the parent organization that was the owner of that domain. And what we found here is that while Alphabet has a substantial presence on both platforms, we see that the remaining key players of Roku and Fore TV actually appear to be different.

So in total, we believe that these findings suggest that the advertising and tracking services' ecosystems of Roku and Fire TV appear to have substantial differences. Next slide, please. So this concludes our brief overview of what the ATS ecosystems of what smart TVs look like. Next, I'll proceed to our second research question-- how effective are existing privacy-enhancing tools? Next slide, please.

So to address this research question, we analyzed the effectiveness of DNS-based blocklists. We studied these particular four sets of blocklists, where the tick mark here on the right indicates if the blocklist contains a SOP list that specifically attempts to target smart TVs. Now, I should say that the results that we present in the following slides are based on the union of all of these blocklists, but our paper has additional details where you can see a comparison of one block list to the others. Next slide, please.

So as can't cover everything here, I'll focus on how well the blocklists prevent tracking. And for this, we search a payload of the packets in our [INAUDIBLE] dataset for Personally identifiable information, or PII for short, and then examined to what extent the blocklists are actually capable of blocking packets containing PII.

However, before we dive into the results, I want to take a step back and recognize that not all PII exposures are necessarily bad. So the view we take here is that exposures to the first party, meaning the app [? developed ?] it's own service, are generally warranted, as this could be for personalization purposes, such as keeping track of what episode of a TV show you're about to watch next. On the other hand, we view all exposures to third parties as strictly training related because there is really no functional reason to send PII to these external entities.

Finally, exposures to the platform operator, meaning Roku, Amazon, in this case, on this binary, as this could be for software updates-- sorry-- software-maintenance purposes. But we also know that both platform operators engage in advertising and tracking. So that could also be the nature of those exposures. Next slide, please.

So here, we show a subset of the PII values that we consider now [INAUDIBLE] and a number of apps that expose each PII value to each respective party. And the block right here is the percentage of domains that were blocked among all the domains that received the respective PII. And what we observed first is that there are very few exposures to the first party, and the blocklists seem to capture this functional nature well, as only a few domains actually block here. Next slide, please.

We found that hundreds of Roku apps, and the advertising ID, and the serial number to a third-party service-- but the blocklists actually do a reasonable job here at preventing this. A substantial number of Fire TV apps also send PII to a third-party service. And here, the blocklists seemed to struggle with preventing exposures of the serial number and the device ID. Next slide, please. We also find a very large number of apps that send PII to the platform operator. And again, the blocklists seem to struggle with [? preventing ?] exposures of the serial number and the device ID.

Now, very alarmingly, we also find almost 700 apps that send all of these three PII values together in one request to what appears to be an Amazon app-related endpoint. And this is especially concerning because it effectively eliminates the user's ability to opt out of targeted advertisements by resetting their advertising ID, as Amazon can now simply link any old advertising ID to the new one by joining on the serial number, which is a static value that the user cannot change.

So in total, we find that hundreds of apps expose PII, and the blocklist seem to be reasonably effective at preventing this for Roku but less so on Fire TV. However, we should note that the numbers for Roku represent a lower bound, as we can only analyze pure text traffic on Roku, whereas we can actually decrypt traffic on Fire TV. Next slide, please.

So in summary, we showed that the ATS ecosystems of Roku and Fire TV appear to differ substantially. We also showed that blocklists provide some means to prevent tracking on smart TVs, but there is definitely room for improvement here. Finally, I want to refer you all to our project side, where you can find our data set, and our tools, as well as our full paper and full conference presentation with additional details. Thank you.

MILES PLANT: Thanks so much, Janus. So looking at your research, which is really focused, obviously, on smart TVs and tracking across that, I think a lot of folks are more familiar with tracking across mobile and web. And I was wondering if you could talk a little bit about how there are similarities between them and differences.

JANUS VARMARKEN: Yeah. that's a great question. Thank you for that. So we actually did perform a comparison to Android in our paper. And what we did was we looked at the top 20 third-party domains of Roku, and Fire TV, and then on Android. And what we found here was that there is actually a substantial overlap between these domains for Fire TV and Android.

And we believe that this likely has something to do with the fact that Fire TV is built on top of Android. So this, in turn, means that all the ad tech that's available for Android is effectively available out of the box on Fire TV, as well as you-- as a developer, you can directly apply these libraries on Fire TV, right?

Now, on Roku, the story was different. The overlap there was much smaller. And here, the explanation is the opposite of what we saw on Fire TV because Roku is like a standalone platform that has no relation to other platforms. So you have to develop these [INAUDIBLE] specifically for Roku, right?

Another thing that's more widespread on smart TVs than the mobile platform is this notion of automatic content recognition. And this is a technology that enables the smart TV to identify what you're watching by essentially taking screenshots and audio snippets of what's on screen and then comparing those with a reference database. And if I recall correctly, this was actually the technology that Vizio was using back when they were sued by the FTC in 2017. So in our way, we didn't really differentiate this type of tracking from the more traditional types of tracking, but I believe that this is definitely a topic that deserves more attention from the research community.

MILES PLANT: And just one quick additional follow-up-- some of these TVs-- obviously, Apple TV is a large area, and then you talked quite a bit about Android. Is it your understanding that, typically, if somebody's signed in through their Google ID on their Android interface on their TV that Google can collect information, both from the TV, and mobile, and web and connect that all to somebody's identity?

JANUS VARMARKEN: I don't have data to support an answer for that, but it seems like the case because if you are literally logged in, then they have some kind of token on that device that identifies you, right? So my immediate answer would be yes, but I don't have any data to support that. Sorry.

MILES PLANT: It wouldn't surprise you, I guess, is the-- yeah. Great. Thank you so much, Janus. And we will be back in when we have a conversation at the end. Third, I'm going to turn to Miranda Wei, who's going to present her research on what Twitter knows.

MIRANDA WEI: Thanks so much, Miles. And also, hi, I'm Miranda. I want to say a thanks to my coauthors and collaborators-- Maddie, Sophie, Nathan, Justin, Margot, Dorota, Ben, Michelle, and Blase. Without them, this wouldn't be possible. So next slide.

If you're here, you probably targeted advertising allows advertisers to use detailed information, like consumer demographics, preferences, and more to target ads. And taking this real ad from Heinz ketchup as an example, the types of ad targeting people might expect are factors like gender, age, or general interests. But the reality is that ad-targeting criteria are far more detailed because social-media companies collect highly detailed information.

So this ad was shown to me on Twitter in September of 2018. And in reality, it was targeted to me because I matched 16 criteria set by Heinz, including being on their list of organic and natural ketchup buyers, as well as, apparently, interacting with the hashtag #parenting And I was able to learn this information because of GDPR and similar data regulations.

So through the data-access requests, as required by these laws, Twitter provides users with data about how ads were targeted to them. And we used this data to conduct a study that measured targeted advertising on Twitter, as well as user perceptions and preferences of that targeting. Next slide.

Our research questions and study protocol were as follows. First, we asked, what are the discrete targeting mechanisms offered to advertisers on Twitter, and how are they used to target Twitter users? We answered this by using real consumers' Twitter data to measure how advertisers were actually targeting ads on Twitter.

And second, we asked, What do Twitter users think of these practices and the existing ad-transparency mechanisms? The study procedure for our research was as follows. First, we recruited participants from the prolific crowdsourcing platform to request their own Twitter data. We then invited them back a few days later when they had received their Twitter data, and we automatically extracted and uploaded only three advertising-related files. We then use their data to create surveys customized to them to pass ads that they had seen on Twitter, and we solicited their thoughts on ad-targeting criteria that had been actually used, as well as their opinions of ad explanations, which are a common transparency mechanism currently used. Next slide.

We received valid data from 231 participants for our study. The average person had spent 6.6 years on Twitter and saw over 1,000 ads in the last-- in the prior three months. Across all of our participants, we collected over 240,000 ads that used at least one targeting type. And an example of a targeting type is targeting by location. We also observed over 45,000 different instances of targeting, indicating a huge diversity of criteria used by advertisers to target ads. For example, Boston would be one instance of location targeting. Next slide.

Through our analysis of Twitter-advertising data, we found 30 specific targeting types that we categorized into three groups. The first is demographic targeting, characteristics about users and their devices either provided by the user or inferred by Twitter. We saw language targeting used over 350,000 times. And other demographic targeting included location targeting or targeting based on whether you're using a new device, which is something that Twitter infers.

The second group is psychographic. So this has to do with users' lifestyles, behaviors, or attitudes, as provided by the user, or it also could be inferred by Twitter. And the more well-studied in the academic literature are types like behavior, or interest targeting, which were used tens of thousands of times in our data. But we found types that were used more than interest targeting.

For example, follower look-alikes allows advertisers to target who-- to target people who are similar to followers of on account, even if they don't follow that account. And that was used half a million times in our data. Conversation targeting targets users based on conversations they've engaged with on Twitter, and that was used five times more than interest targeting.

Finally, the third group of targeting types is advertiser-provided, which uses information that advertisers collect about users online but outside of Twitter. And so these targeting types can be very opaque to users, for example, tailored lists, which use lists of users as determined and uploaded by the advertiser; mobile targeting, which lets advertisers target users of their own mobile apps; and tailored web, which allows advertisers to target people that have previously visited their website. Our work uniquely captures this whole range of targeting types and how they were actually used by advertisers. Next slide.

So Twitter has advertising policies that prohibit targeting by sensitive or illegal categories, yet because we had real consumers' ad-targeting data, we discovered that there were established companies who may have violated these policies. For example, we observed keyword targeting to gay, #African-American, and Latinas. And we also observed targeting to people that had engaged in conversations about liberal Democrats, a UK political party.

Some of the names of the tailored-list targeting also suggested violations, such as account status-- balance due, Christian audience to include an LGBT-suppression list. These names were selected by advertisers for their own reference, as advertisers upload lists of user identifiers, like Twitter handles, to specify the users that they want to target. So I'll return to the potential violations at the end of the talk and the implications, but know that even though these were in the minority, these are just some examples of the types of targeting that could be perceived as sensitive. Next slide.

Because we found a range of targeting types, we also wanted to measure consumer understanding of these types. And we found that consumers' mental models about some targeting types was correct, but others were not. Over 95% of our participants understood age, location, gender, and keyword targeting. However, only portions of our participants understood types like language, platform, and conversation targeting.

For example, one participant thought that platform targeting meant political platform, but "platform" actually refers to the device, or the operating system. And another participant referred to the common misconception of phones listening in for conversation targeting, saying, given what I know about how phone microphones are always on when ads pop up, based on what I've said in a conversation. But as I mentioned earlier, conversation targeting is actually about Twitter conversations and which topics you tweeted or interacted with.

And finally, very few participants understood targeting types, like tailored lists, behavior targeting, and mobile-audience targeting. Tailored-audience targeting is the advertiser-uploaded lists that I mentioned before, and mobile is about targeting to people who installed an advertiser-owned app. But one participant thought it was based on your phone network. So though a few targeting types were well-understood, in general, consumer awareness of ad targeting and the ad-targeting types that are currently being used is very far from ideal. Next slide.

Our user study also elicited participants' opinions about ad targeting. And notably, these survey questions asked about criteria that had actually been used to target ads to that user. Participants tended to be OK with the use of targeting types generally, such as targeting based on events, but were less comfortable with the use of specific instances, like the 2019 Women's World Cup. We found that only 25% of participants did not want to see ads that used event targeting. However, when participants saw a specific event they were targeted on, then 65% said that they did not want to be targeted.

And we also found that when participants agreed that an instance of targeting was accurate, they were significantly more likely to agree that it was fair and comfortable. For example, accuracy was highly correlated with many other perceptions. But this doesn't justify collecting more data to increase accuracy at the cost of privacy. And our free-text responses indicated that there is an upper bound where increasing accuracy is no longer comfortable. Next slide.

So looking now at ad-transparency mechanisms, we found that participants valued having more detail about how ads were targeted. And in our study, we looked at six different ad explanations, which are info boxes that appear when you look at an ad and you can click in the menu, why am I seeing this ad? And you can see the full text of the ad explanations in our paper, but here are the results.

The bar chart will show the percentage of participants that strongly agreed or agreed that a specific explanation that we tested was useful. And maybe not surprisingly, control was the least useful ad explanation. We simply said that they might be seeing the ad because an advertiser paid for it.

The next group of ads that were the next-most useful were the Facebook and Twitter ads that we're currently being used when we ran the study. So we replicated Twitter's exact ad explanation, and that describes targeting only by interest, age, or location, if those were used, because we discovered that Twitter only ever mentioned three types of targeting-- a total of three types of targeting in their explanations, even if more were actually used.

And participants rated our three speculative ad explanations-- the detailed text, detailed visual, and quote, unquote "creepy"-- as the most useful. In our qualitative responses, we found that participants perceived ambiguity and explanations to mean that information was either missing or intentionally omitted. And so that might explain why they preferred our more-detailed explanations.

For the explanation that we labeled creepy, we use declarative language, and we listed as many targeting types as possible based on how that ad had actually been targeted. And we thought that this would be overwhelming and creepy, hence the name of what we called it. But participants actually thought that this was the most useful of the three that we created. Next slide.

And in conclusion, I have four key takeaways. So first, mandating advertising transparency is extremely important. GDPR's right-to-data access was the only way we were able to do this study, and there's still more transparency to be had, given that it took us over four months to receive clarifications from Twitter about how to correctly interpret the consumers' ad files. And we're researchers.

Through this work, we uncovered that well-studied targeting types, like interest targeting, were some of the least important to our participants. And instead, lesser-known types, like tailored audiences and look-alike targeting or perceived as more concerning but were poorly understood, and thus weren't further studied. And as I mentioned earlier, advertisers uploaded and named lists of user identifiers for targeting, which is how we uncovered some potential Twitter-policy violations. But without advertisers naming their list descriptively, we never would have known, for example, that there was a list of Christians that the advertiser wanted to exclude from their campaign. And this raises serious long-term questions about enforcing and maintaining transparency, given that advertisers can change list names and use other ways to hide how they're targeting users.

And finally, we found that existing ad explanations should be much more detailed. Our three speculative ad explanations perform significantly better than Twitter and Facebook's existing ad explanations. Next slide. Well, thank you so much for listening to my talk.

MILES PLANT: Thanks, Miranda. You talked a little bit about-- and just talked about it again, about tailored audiences and about companies uploading lists of users and giving those lists names. And effectively, it allows the company that's using Twitter's advertising to target those folks with ads or to create a look-alike audience based upon them. And you noted a couple of examples of companies uploading lists that seem to violate Twitter's own rules about what can be the basis for those. What can you tell about what Twitter knows about those lists and what information they receive? And what, if anything, did they do to understand what's in the lists?

MIRANDA WEI: So as far as we can tell, Twitter chooses not to know the origin of these advertiser lists. I'm not aware of any public information about how Twitter audits these lists. And certainly, there's no way for end users to the basis of these lists. If you download your data through the way that our participants did, which is something that you can do if you have a Twitter account, you might see the name of the advertiser list, but I'm not aware of any way to get more information about the origin of these lists.

And so as you mentioned, the violations that seem to violate Twitter's own policies-- the privacy implications for consumers are pretty serious because that means that Twitter has policies, but maybe they're not checking them, or maybe they don't have any means to check them right now. But that also means that Twitter's passing the buck on to advertisers who are uploading these lists and saying, you have to make sure that these lists are OK and that the privacy is good, whatever that means. It's all very vague.

And so this is not a good situation for anyone to be in. In fact, Twitter actually added a question, like a Q&A in their health pages, that inform advertisers that the list names would be public around the time that our paper came out. But as far as we know, they didn't do anything besides put that up. And so I don't really believe that this is something that individuals should be having to track down. And they should be able to if they want to, but instead, this is a systemic issue that regulators need to address and with the cooperation of the advertisers and of Twitter themselves.

MILES PLANT: And just to follow up on that, Twitter's not the only one that uses this list in advertising technology, right? I think Facebook has custom audiences, and it's very, [INAUDIBLE] very similarly. Are you aware of any other companies that use similar technology?

MIRANDA WEI: Google is the big player in this space. Google would be on web ads and not necessarily their own platform. But Google is a big one. And I'm sure TikTok. Any social-media company probably has similar things, yeah.

MILES PLANT: Well, thank you very much. Before turning to the general questions, I really want to just thank all three presenters and really encourage the audience to read their papers. I think that the research on each area was really eye opening and just begs a lot of additional questions too. And so I look forward to seeing additional research from each of them on these topics.

So all three of our presenters come to the ad-tech space from different angles, and they tackled different areas of it. But something that always comes to my mind and-- it is, what can consumers do, and what can policymakers do to provide greater transparency and privacy? So maybe we can take those one at a time. And what can a consumer do when faced with each of these issues? I'll just open the floor and let you all tackle that as you see fit.

IMANE FOUAD: Maybe I can start. Thank you, Miles, for the question. So what can consumers do to prevent web tracking?

Well, for consumers, I will say it's complex and even impossible to block all tracking. No consumer can fully prevent tracking. However, as we have seen in the presentation, one can reduce the tracking on the website using browsers' extensions, for example.

And so Miles, you were perfectly right to ask what both consumers and policymakers can do because one of the keys to protect users' privacy is to focus on regulation. Well, I'm not an expert, but I worked with legal scholars on compliance with the European regulation. Namely, we studied the compliance with the GDPR and ePrivacy Directive. And it turns out that it's really complex to design a regulation that that's going to ensure better protection of the user.

I believe that the main reason is that the policymakers do not have enough technical knowledge, and so my gut feeling is that it's extremely important that regulations should be designed. And there are collaborations between policy makers and computer scientists. So in summary, my answer would be that the collaboration between the different communities-- computer scientists and policy makers-- is the key to design a strong regulation that can help protect users' privacy.

MIRANDA WEI: I'll just say I love that answer, and plus one to everything you just said. One other thing that I might add is that existing transparency mechanisms or regulations are getting there. And it's a sign that this is-- we're moving in the right direction.

My work was only possible because of GDPR and allowing the right-of-data access. If that hadn't been a law in the EU, I might not have been able to see my own Twitter data. But that was something that Twitter did, and I guess they just decided to implement it for all of their users.

Another thing that I'll mention is that consumer awareness of these issues is also very important. It's not sufficient by any means, and it shouldn't be on any individual person to have the responsibility to demand anything. But collectively, and being aware of these issues, and talking about them-- I think, that helps policymakers understand what's important and also, companies themselves.

MILES PLANT: Yeah, I just want to follow up real quick with MIranda on that. In terms of downloading your data from Twitter, how complicated was that, and what was the process, just a real quick explanation?

MIRANDA WEI: Yeah. So it is buried relatively deep in your settings. I think they have changed it recently, and so I don't have the exact steps listed. But if you just search for "Twitter data download," you should be able to find it. You go through, and you have to enter your password. And then it takes a few days for them to compile the information, and then they'll send you a notification saying that it's ready. And you can download it.

One very interesting, or bad consequence is that the data downloads are very hard to understand. It took us, a team of very qualified researchers, I'd like to think, many, many weeks to understand these JSON files and go through, like, what does this attribute mean? What does this characteristic mean?

And there's ongoing work by some people, some of my coauthors, on how to interpret this, and what are the best ways for consumers to understand it? But if you are really interested in the nitty-gritty details, you can certainly find a lot of JSON files right now if you go to your Twitter sites.

MILES PLANT: So I want to turn to Janus in posing that question that we asked before, which is in the smart-TV space, what can consumers and policymakers do to either-- well, to protect consumers' privacy and ensure greater transparency?

JANUS VARMARKEN: Yeah, I just want to first echo what the other speaker said. I think those are great answers. In terms of concrete things that you can do as a consumer here, well, we already saw in the presentation that if you do actually apply blocklists, you gain something. And every little bit is something, so I would definitely recommend that as the immediately obvious thing to do, right?

Then this is kind of like platform agnostic, irrespective of what smart TV you have. But then if you have, say, an Android-based one, you could actually also deploy this tool that we used that's called [? Amon. ?] It lets you do even more fine-grained blocking. So because it can actually decrypt traffic, you can define certain values that you don't want to be exfiltrated from your TV and then either have this tool completely deny those requests or change them to other arbitrary values. But that's only specific to the Android-based technology, basically. So you have to look carefully for specific tools if you want to go down that route.

And then like a good everyday thing to consider, I would say, is that the thing about what apps you use and try to like limit your usage-- so one interesting thing we saw was that there was some apps from some developers that were very aggressive in terms of the number of trackers that they contacted. And a general theme here was that it was apps-- that all of these apps, like, looked similar to one another, and you could see it was the same developer. And maybe they were more like cooking apps, and then you had one with Indian recipes, one with Italian recipes, and so on and so forth, right? And you saw that all the same trackers were used in all of these 20 apps from the same developer, essentially.

Now, to focus on the policy makers, I would say that if we can provide some strong incentives to the manufacturers of these devices, such that they get a competitive advantage through privacy, I think that that would change a lot of things, right? So if we could offer some kind of official recognition of products that adhere to certain standards, like some independently audited label for, say, a smart TV or any other device, where you have an independent agency or governmental agency auditing what kind of private information these devices access and send out, I think that would be a good way to move forward and actually make manufacturers also join in on this.

MILES PLANT: I really appreciate that input. I think that both for the consumers' perspective and the lawmaker perspective, it's helpful to think pragmatically about how to tackle these issues and what's actually going to enable consumers to make informed choices and protect their own privacy. In particular, Janus, your mention of shopping around the apps is something that the FTC talks a lot about, about comparison shopping, where there are multiple options and really thinking about trade-offs in terms of what data you're handing over and what services you're getting back. And so it's good to hear that echoed.

And turning back, I'll just say that during Miranda's presentation, she mentioned, and in my follow-up question mentioned, that Twitter has, in her view, abdicated responsibility back to the advertisers. So framing it that way, or taking that as a starting point, what can companies do and what should they do to provide better privacy protection for consumers, as well as better transparency on the collection, use, and disclosure of consumer data? Janus, do you want to go first?

JANUS VARMARKEN: Sure. So in terms of privacy, protection, what companies can do here, right? I would say that if you could actually deliver these products with functionality that allows users to block certain kinds of traffic on the device itself so that you don't have to go out and set up some sort of, like, specific device in your network to block this traffic, this would be similar to-- I guess most people are familiar to browse-- familiar with the browser's and ad block [INAUDIBLE], right? But at the system-wide level here, if the manufacturer could deliver that out of the box, such that the user had the ability to say, no, I don't want you to send my email, or I don't want to send my serial number to anywhere, and block that just out of the box, that would be great, right?

In terms of transparency, I would say that there's a problem here that researchers like us-- we want to look at the network traffic, and we want to see it in clear text to understand what the devices are doing, right? But on the other hand, you have the manufacturers that think of this as a security risk if they provide this functionality to us. So there's a bit of tension there, right?

So I believe there's a need for some work where you can give researchers access to these privileged programming tools, where you can actually see clear-text traffic but do so in a responsible way, by only giving access to select [INAUDIBLE], for example, without getting too technical. And then I also believe that if you could use a tool similar to the one we proposed in this work from the manufacturer's side to vet the apps that they publish in their app store, that would be good. Then you could have the manufacturer itself assign privacy labels to the ads that are available for the devices with the number of trackers they contact during the vetting process, and number of personally identifiable information they expose through the vetting process.

MILES PLANT: Great. Miranda, do you have any thoughts on this?

MIRANDA WEI: Yeah. So your question about, what can companies do, reminds me of some of the things that were mentioned earlier in the Algorithmic Bias talk. And specifically, one of the speakers mentioned a lot of companies maybe not looking intentionally at their algorithms in the hopes that they wouldn't be held responsible if they didn't really know what the algorithms were doing. And I think there's something similar here, where if Twitter doesn't know what the violations are, then maybe it's not their fault.

But I think instead, both, in my case, the platforms where the ads are happening and the advertisers themselves, everyone needs to be proactive about caring about privacy and thinking about what consumers might want or what they need because just because you can do something does not mean that you should do it. And I think that's especially true with ad targeting.

MILES PLANT: [INAUDIBLE] thanks. And Imane?

IMANE FOUAD: Thank you, Miles. So in the context of web tracking, what companies can do-- my answer would be it depends on the company we're talking about. If we're talking about the publisher, which means the website owner, then I would say that tracking techniques are so sophisticated that today, it's even hard for the publishers themselves to detect all the tracking happening in their website.

And so, if, by companies, you mean the advertising companies, then given that tracking is their core business, then I believe that the main-- or what would be the most efficient answer, once again, would be to go through a strong regulation that can force these companies to comply and to be transparent. And so to conclude, once again, the solution is to be transparent so you can really come through strong regulation, that trackers should be forced to publish in a transparent way the techniques that they deploy and with whom they're exchanging the data that are collected about you.

MILES PLANT: Thanks so much. Well, we are running out of time. So I want to thank all of you again. Thank you to Imane, Miranda, and Janus for presenting their research, or presenting your research to us, and driving this field forward. Next up will be a panel on the Internet of Things, moderated by my amazing colleague, Linda Holleran Kopp. And so I hope you all will join us for that. Thanks so much.

[MUSIC PLAYING]

LINDA HOLLERAN KOPP: Hello. Welcome back, and thank you for joining the fourth panel of the FTC's sixth annual PrivacyCon, where we all learn about some very interesting research that is being conducted on privacy and the Internet Of Things, or IOT. My name is Linda Holleran Kopp, and I'm an attorney with the FTC's division of Privacy and Identity Protection.

Today's panel will first discuss research related to the privacy compliance of apps that are used with voice-personal assistance, like Alexa or Google Home, that so many of us have in our homes. We will then look at how a privacy label on IoT devices can impact consumers' perception of the privacy risks of using that device, as well as their willingness to purchase the product. And a final presenter will talk about an interesting new dynamic tool that is designed to conduct a real-time privacy analysis of IoT apps.

At the end of these presentations, we will have the opportunity for questions and answers. If you have any questions, please email them to privacycon@ftc.gov, and we'll try to get to them as time permits. So let's get started.

First, we will hear from Anupam Das, an assistant professor in the Computer Science department at NC State University. He will present his paper, Hey Alexa, is this Skill Safe,"? taking a closer look at the Alexa skill ecosystem. Anupam?

ANUPAM DAS: OK. Thank you. Thank you, Linda, for the nice introduction here. So today, I'm going to talk about our recent work, where we looked at identifying the overall vetting process for third-party applications on the Alexa skill ecosystem.

So what next? So let's see Alexis. So Alexa is Amazon's smart voice assistant. And there are multiple vendors out there. And statistics say that there's almost-- over 4 billion active devices that we have voice assistant built in them.

Now, one of the things that is driving this adoption is that a lot of the smart devices that are coming out of the market right now are-- can easily integrate with the device assistant, which means that users can easily control smart home [INAUDIBLE] and devices through their voice assistant. And other neat features that are coming out from a lot of the voice-assistant platforms is that a lot of these vendors are opening up their platform to enable third-party developers to develop applications on top of the voice assistant.

So in the case of Alexa, these third-party applications are typically known as skills. And all, you need to do in order to install and interact is use the voice interface. So for example, if you want to install or interact with one of the third-party skills, all you need to say is, "Alexa, open," followed by the invocation name for the particular third-party that you are interacting with. And from a developer's point of view, that also means that you can always-- you can develop new applications and submit it to the Play Store for other people to use. Next.

So in terms of our research questions, we were trying to look at this overall vetting process that goes into this Alexa skill ecosystem, so we're looking at existing limitations that might be exploited by attackers. So one of the things that we looked at is that for the Alexa skill ecosystem is that you can have applications that have duplicate invocation names. So basically, you can have different developers develop skills that have the same identical invocation name. And now, this can create confusion in the sense that if a user wants to activate one of those skills through the voice interface, a user might not necessarily know which particular skill is being activated. So we looked at this particular aspect in which attributes might be used to activate a particular skill.

Then we also looked at whether a developer can actually register any skill under any arbitrary developers' names. Now, this is important because you can launch, potentially, phishing attacks through this limitation. The other thing we looked at is that, in the context of skill, is that-- whether Amazon was properly vetting all the different data types a particular skill was trying to request. And so those data types are typically captured through the term "intent" in the context of a skill.

And the other thing we were looking at is like any other platforms. Amazon also has a permission API that is made available to third-party applications. But we were also looking at whether a lot of the skills were actually bypassing this permission model and actually invoking the sensitive data types to the voice interface instead of using the permission APIs.

Next, we looked at-- looking at-- or looked at whether you could actually do-- perform squatting attacks in these kind of ecosystems, where, given that you can invoke the skills through the voice interface, it's easy to think of it as doing a skill that sounds very similar to an existing skill. So we looked at whether that is physical, and if so, which particular patterns of squatting was more effective. And lastly, we looked at the privacy policies of a lot of those skills and whether the privacy policies were properly disclosing the data types that the skills were requesting. Next.

So before I go into the findings, let me briefly highlight how we actually conducted this whole analysis. So we crawled the top seven Alexa skill stores. So this covered skill stores from Canada, US, UK, France, Germany, Australia, and Japan.

And this resulted in a total of over 90,000 unique skills that we were able to identify. And once we identified the skills, we were basically scraping metadata from the public pages. So this metadata includes various information, like the skill name, the developer, the ratings, if there's a privacy policy, what is the privacy-policy link, and so on.

Now, in order to test which particular skills might activate, we used a semiautomatic approach where we used Amazon text-to-speech engine, Polly to activate particular skills. And to analyze the privacy policy of skills, we used existing tools like POLICHECK, which is a tool that enables you to find inconsistencies within the privacy policies themselves. And lastly, we also developed our own skill and published skills to validate some of the findings that we found.

So let me go to the first point. And so the first point is, given that there are multiple skills with identical invocation names, the very first question that came up-- came to us is that, how does Amazon select which skill to activate? And can this activation process lead to activating the wrong skill?

So we tried identifying various public attributes associated with a given skill. So for example, if you go to a skills page, you can actually look at the various number of ratings that has been given to this skill, the average rating that the skill has, whether the scale has an information requests. And you can even identify the age of the skill and whether the skill has any content-advisory label.

So we tried testing with various attributes that are publicly available. And we then did some statistical analysis to figure it out if one of the particular attributes is strongly correlated, in terms of which particular skill gets activated. And we found that skills typically that have high ratings or average ratings has a strong correlation in terms of getting activated. Next.

So now, that shows correlation, but that doesn't necessarily mean causation. So we want to test this thing out, right? So this is where we developed our own skill, and pushed those skills, and published those skills, and wanted to see if we can manipulate some of these attributes and see if the activation process changed.

So we registered two different skills with the same identical invocations. And so we published one at first, and we wanted to check if that was properly activated. And it did properly activate. And then we waited for one week, and then we published the next skill with the same invocation phrase. And we then checked which particular skill was activated.

It turned out that the new skill that was registered was activated instead of the old one. So this indicates that there is a particular choice that Amazon makes in terms of which particular skills to activate. Statistically, we found that reading was one of the influential factors, so we tried to then check if we can increase the rating of the old skill to manipulate the whole selection process. So we did that, but, unfortunately, that did not result in the change in the selection process. So that indicates that Amazon is probably using some internal parameters which is not publicly available to select the skills.

So the takeaway message here would be that there are duplicate invocation and duplicate skills with the same identical invocations. And given that Alexa now has this auto-enabled feature, which means that if you ask for a-- if you invoke one of the skills and if a match is found, it will automatically install it in your account, this can potentially lead to activating the wrong skill.

Sorry. One second. OK. So then the next thing that we looked into is the registration of particular developers names. For example, can an attacker register skills under any arbitrary developer's name? So for this, what we did is that we tried to register stills with different developers' names, well-known developers' names.

In our case, we tried to register skills under the name of Microsoft, Ring, Samsung, and Withings, and Philips. So a lot of the times, we did succeed. As you can see here, we were successful in registering with Microsoft, Ring, Samsung, and Withings. But we did not succeed when we tried to register a skill under the name-- of the development name of Philips.

And so one thing that we found-- and the reason for this could be is that when you submit a skill, the skills are vetted, and there's both automation and manual process in this vetting process. So we believe that the one case that was not successful may be the case where a [INAUDIBLE] might have realized that this is a skill being pushed under different developers' names. And that could be the reason that it was flagged.

But one other interesting thing that we also found is that once-- if you do succeed-- so for example, we're showing a screenshot here for the Ring-- if you do succeed and if you click the Rating tab, then it basically hotlinks the developer's name to any other product that Amazon has under that developer's name. So this basically can create even further confusion to a consumer thinking that this is actually developed by this Ring developer.

So the take-away message here that we found is that it is possible to register a skill under well-known company names. Next. So the next thing that we analyzed is the change of back and forth. And so when you say back and forth-- so this is in the context of a third-party application where the third-party application controls the back end of how they process the data and how they want to interact with the user.

So one thing that we tested out is that since skills have to register for the particular data types that they want to interact with-- that they want to interact with-- and typically, these terms and data types are termed as intent, is that you can register any arbitrary number of intent, whether they're technically used or not. So what we did-- we registered multiple intents, and some of them were sensitive in nature. For example, we used [INAUDIBLE] type of phone number.

And we then basically tested the skill and vetted the skill. Whoops, I'm sorry. We submitted the skill, and the skill was vetted, but it did not trigger any of the sensitive intern types. And it was eventually approved. And when I mean, approved it was publicly available at that point.

So once that was publicly available, what we did is we went to the back end because this is the back end we control as a developer. We changed the interactive dialogue to activate one of those dormant intents. So basically, at this point, we were asking for the user for their phone number where, previously, this was never used-- or triggered during the vetting process. And so this whole process means that an adversary can potentially register [INAUDIBLE] intent and after approval, can retrigger this intent to retrieve sensitive information from the end users.

So next, we looked at whether some of the skills were actually bypassing the permission model. So before I start that, Amazon does explain what the permission model in the context of an Alexa ecosystem is. So Amazon does have a permission API model, which enables, basically, developers to request sensitive information from the user using this permission model. And if you do use this permission model, what that means is that Alexa basically forwards the end users to the companion app to approve, or give consent, to use this sensitive information.

But, again, this voluntarily has to be declared by the developers. So we wanted to see if there were skills that would actually bypassing this permission API and actually asking the sensitive information through the voice interface itself. So we looked for sensitive [INAUDIBLE] permission, private data types, like phone numbers, location, email, and name. And we searched for these terms within the skill-description page. And so this [INAUDIBLE] description page is the information that is available publicly on the website. And we basically had [INAUDIBLE] go through this and extract, and basically, find skills that were actually accessing some of these sensitive information types.

So this resulted in us finding 358 skills that were actually describing, at least in their description of accessing the sensitive information. So our next task was to see-- manually vet and activate the skills to see, actually, whether they were really accessing the sensitive information. And so, through our manual vetting process, we did find that 52 of them were actually false positives. So they really didn't ask for that information, even though the description page did mention the sensitive information. But we did find 169 skills that did request the sensitive information. And so that basically indicates that the Alexa may not be properly mediating the intent type based on the sensitive information that has been requested.

And so lastly, we also looked at the privacy policy. And so one unique attribute about privacy policies with the Alexa skill ecosystem is that by default, skills don't need to have a privacy policy. And this is a bit different from Google ecosystem.

But Alexa does mandate the privacy policy for skills that request permission and that access the permission APIs. So any skill that is going to be requesting a permission [? productivity ?] data type will have to submit a public-policy link. Otherwise, it will not appear publicly in [INAUDIBLE]-- sorry-- in [? skills ?] store.

And so we basically analyzed the skills that were accessing this types of permission. And we also downloaded the privacy policies for those skills. And we wanted to see if the privacy policies were accurately reflecting on the permission data types that they were capturing.

And so we analyzed around 1,224 skills, and we found that a lot of them were not properly explicitly addressing the sensitive data type they were accessing. So in general, the data types, like full name, phone number, and locations, had the most number of inconsistencies, in the sense that they did not mention anything in the privacy policy about this particular data types. So that's one other interesting finding that we did.

So there's a few more other interesting outcomes on the paper, but we will not have time to cover them. But then we did make some observation and did try to make some recommendations based on what we found. And one of the indications and recommendations is that a lot of the users did not be-- did not understand the difference between a native skill and a third-party skill just by looking at it in the public skills store page. And so we think a skill-type Indicator is something that would be useful.

We also feel like validating the developer affiliation is an important factor here, especially as it [? enables ?] phishing attacks. And the other thing is that, given when the skills are submitted and you have to submit the various intents, this is where I think that Alexa or Amazon can do a better job in validating the intent types. And the other thing is that we saw that the enforcement of a privacy policy is not always useful. And so we believe that a better approach would be to have a privacy-policy template built into the submission web page where the developers will have to explicitly say the type of data they're accessing, the purpose, maybe whether users have opt-in and opt-out choices.

And so that's some of the recommendations we made. And so we did disclose our findings with them Amazon and had multiple interactions with them. We did see some of the skills removed. But again, we don't know if that is the result of our research, or any other research that is [INAUDIBLE] happening, or whether the developer just simply removed the skills themselves. But we have made our data set public, and the website also has some demos of some of the things that I did discuss today. So if you're interested, I would encourage you to visit that website and look at some of the other findings that we found.

So with that, I'll end. And I'd also like to thank all of my coauthors listed here. So if you have any further questions about any of the findings that we have, we'll be more than happy to answer any of the questions you may have. With that, I think I'll end.

LINDA HOLLERAN KOPP: OK. Thank you, Anupam. That was great. Next, we are going to hear from Jeffrey Young, who is pursuing his PhD in computer science at Clemson University. He will present on his paper, "Measuring the Policy Compliance of Voice-Assistant Applications. Jeffrey--

JEFFREY YOUNG: Thank you, Linda, for the introduction. And yes, today, I'm going to be talking about our paper, Measuring the Policy Compliance of Voice-Assistant Applications, so if you can click to the first slide.

So the first thing I'd like to do is just set up some definitions. So what is a Voice Personal Assistant, or VPA? VPA actually is the software that runs on smart speakers, such as Amazon's Echo and Google's Home. The VPA itself is usually named separately, like Alexa. And for this work, we analyzed the two most popular VPAs, which is for Google's Home and for the Amazon Alexa system.

To show the popularity of these devices, currently, there are over 128 million smart-speaker users in the United States alone. And the VPA itself runs on other software similar to cell-phone apps. So these applications are called "skills" for Amazon, and they're called "actions" for Google. But for the remainder of this presentation, I will just refer to voice apps as "skills" just for simplicity.

So skills, like cell-phone apps, are created by third-party developers. So both Amazon and Google do not put stipulations on who can create and publish skills. But these skills do go through a vetting process on both platforms. And for skills that pass-- they're publishers to a skill store. So on the internet, there's a lot of different information regarding the developed skills.

For the purpose of security, VPA providers have created a set of policies that skill developers should adhere to. And just for a quick example, there are over 50 content policies for Amazon alone. So in this work, basically, what we did was we conducted a large-scale testing of skills on these policies to see their compliance to these different policies that were set up by the VPA vendors. So if you could click, basically, here, it's just because recent research has shown that the VPA vetting process can be weak and that security and privacy issues exist, we are interested in the policy compliance of deployed skills currently running on the Amazon and Google skill store. So that's what we looked at in this work. Next slide.

So here are some examples of policies put in place on the Amazon platform. We consider these policies to be high-risk because of the subject matter that they deal with, that being kids and health. If you could see here, in the kids' section, the policy is collect any personal information from end users, which would be children. Of course, it's not allowed. This is for the Amazon platform-- promotes products, contents of service, or directs end users to engage with content outside of Alexa.

Most data collection for kids is prohibited, by the way, over this platform. Content presented to children is also of particular interest, as well as promoting content outside of the platform-- so anything that is promoted not within the Alexa ecosystem.

Most health data is prohibited from collection. Also, giving health advice or information must come from a disclaimer-- or must come with a disclaimer-- I'm sorry-- stating that this material is not suitable for actual medical advice. Next slide.

So just to give an outline of how skills work, this is a basic skill interaction. So this is a high-level view of how a skill works. So skill interaction-- it starts with an utterance. And that, for example, would be, Alexa, open ABC, some skill. For some skills, permissions need to be granted via an app on a cell phone. Here, I would like to mention that permissions do not need to be granted for every skill. They can be accepted once for many skills of a certain type.

Next, the device captures the audio files of the user-- so whatever's verbally spoken into the device-- and sends it to the cloud to be processed. It is here that a skill's front-end code is actually housed on the cloud. From the cloud, the interaction data can be sent to the skills back-end code, which can be hosted outside of the VPA platform. This is what makes traditional static code analysis for [INAUDIBLE]-- or for skills not possible and basically serves as a black box. So we really don't know what's going on the back end of these skills.

Also, it is the back end that provides most of the content of the skill, which can include audio files, images, and text. So because a large portion of the skill's source code is hosted externally, all we have to analyze is the skill's functionality-- so how the skill interacts. What the skill actually does is what we look at. Next slide.

So here are the contributions, some of the contributions, of this work. So we designed and develop a dynamic testing tool name "Skill Detective." it's basically a chatbot model that interacts with the Echo device and the Google device. We conducted a comprehensive dynamic and static analysis of skills to detect if they follow current policies of VPA platforms. So from the interaction data that we collect, we could determine policy violations.

And after over a year of development and testing, we have tested 54,055 Amazon Alexa skills and 5,583 Google Assistant actions. I would like to mention that there are many more Amazon Alexa skills published than that of Google actions. Also, Google makes it more difficult to interact with its platform using web drivers for some reason. So it makes it more difficult to actually test Google actions. Next slide.

So this is the Skill Detective overview. I will briefly go over this. Let's see. Here, I would like to mention that both Amazon and Google have created a developer's testing console that takes then a textual input and transcribes the voice of the device into text. So you will receive a textual output back from it.

Unfortunately, it will not transcribe audio files. If the output is strictly audio file, then it just gives you a message saying that it's just an audio file. So we have to transcribe those separately.

First, we created the simple utterances from the different skill stores. So these are used to begin the skill interaction by invoking the device. So we mined the skill stores and collected a data set of all the sample utterances.

Second, the Skill Detective collects the initial skill output and sends it to the interaction model. So here, Skill Detective determines if the output contains a question and if so, will predict the answer to the question. So this is done because these are verbal machines. So you have to be able to talk back and forth in order to gather as much output as you possibly can from each individual skill.

So next, the answer would be sent back to trigger the next output, and then this process goes on until we've exhausted the skill. We call this back-and-forth skill "navigation." So we're navigating the skill. So basically, the Skill Detective just carries on a conversation with the smart speaker and records all the interactions.

So during skill navigation, Skill Detective creates a map of the skill by keeping track of the outputs and the output types so that the skill can be thoroughly explored. And during all this skill interaction, the system collects all the data types of output by the skills. This includes images, audio files, and text. And lastly, the interaction data itself is analyzed for policy compliance. So we look at all the data that we've collected from each skill separately. Next slide.

So automating the policy-violation detection-- so we check-- basically, well first, this is a page taken from the skill store for a specific skill, wealthy nutrition. We check the category to determine which policies apply to each given skill. So first, we want to know, is the skill, like, in the kids category, where certain policies would apply that would not apply to other skills?

Next, we check if the skill has any missing permissions. So should the scale have a permission set, like collects data or something like that? And we can analyze the skill's output and then determine if it does collect this data. And if so, are the permissions handy on the website?

We also check for inconsistencies between the privacy policy, skill description, and the skill's behavior. We transcribe all the audio files to check for policy violations. And we also analyze images for potential policy violations. So overall, we check for over 40 different policy violations on the different platforms. Next slide.

So here's some evaluation results. So we identified 6,079 skills and 175 actions, potentially violating at least one policy requirement. 590 skills and 24 actions-- or of those actions violate more than one policy. In the kids category alone, we identified 244 policy-violating skills.

And 80% of the skills and 68% of actions in the Health category potentially violate at least one policy. And 623 skills and 25 actions potentially violate policies related to personal-data collection. So these are some of the results that we've found so far in our testing. Next slide.

So here, we have some examples of some policy violations that we found. You will see that this is the Alexa testing console that we're interacting with. So the first one here is collecting personal data. And as you can see, it says, please tell me your name. Now, this skill normally would not be in violation, but it's in the kid's category. So because it's in the kid's category, it asks to collect-- it ask for the child's name, or the user's name. Next on the examples.

Also, here's an example of explicit or mature content and toxic content. So I won't read what it is, but yeah, if you could read it for yourself. The next example here is requesting a positive rating, which is actually a violation. And that is done, too, because the ratings that are used-- or so it's thought that the ratings are used to determine which skill is selected for and during invocation. Next slide.

So here's a few key just evaluations. So we analyzed the entire kid's category on the Alexa platform, and these are the results. So collecting data in kids-- we found that 34 out of 3,617 skills collect data. 21 skills directed users to outside of the Alexa platform. 12 skills had explicit or mature content for children. 177 skills requested a positive rating.

We found four skills that had toxic content, and we found 244 skills totaling-- or 244 total policy-violating skills for the Alexa kid's section. Now, if you see in contrast, the Google action section-- we actually-- some of the, like requesting a positive rating does not apply, and directing users outside of Google does not apply. And we actually found no violations in the kid's category. But again, if you look at the difference, between 3,617 skills and 108 actions is all that there was at the time that we collected the data.

Next slide.

So here's the non-kids evaluation results. So we found totally 3,464 skills collected-- or requested a positive rating. 1,709 skills in the Health category were lacking a disclaimer. That's 79% of the Health category. In contrast, 151, or 66%, of actions lack a disclaimer in the Health category.

146 skills collect health data, whereas 13 actions collect health data. Three were directing users outside of Alexa. This was actually found In Image or Audio files. Two were lacking privacy policy. This is also in the Image and Audio. And incomplete privacy policy-- we found two skills, no actions.

So if you could click to the next slide-- so this is basically the end. We reported the results to Amazon and Google and got their acknowledgment. We have found that Google confirmed that 43 out of 175 actions were immediately removed from the store because of our reporting. And the remaining actions were deemed to have not been in violation, or the policy violations only wanted a warning rather than a takedown. So we did report everything to Amazon, and they have-- we have not validated anything like this with them.

So that's it. Thank you very much. And next slide, I guess, yeah, that would be the end of it.

LINDA HOLLERAN KOPP: [INAUDIBLE], thank you, Jeffrey. Now we will hear from Pardis Amami-Naeni, who's a postdoctoral scholar at the School of Computer Science at the University of Washington. As part of her doctoral research, she had develop a privacy and security label for smart devices. She has continued this research in her current paper and will present on which privacy and security attributes most impact consumers' risk and perception and willingness to purchase IoT devices. Pardis.

PARDIS EMAMI-NAENI: Thank you very much, Linda. And hi, everyone, and thank you for joining my talk. I'm Pardis Emami-Naeni, a postdoctoral scholar at University of Washington

Today, I'm going to talk about our project to identify how privacy and security factors impact IoT consumers' risk perception and willingness to purchase. This work has been conducted while I was doing my PhD at Carnegie Mellon University and it's a joint work with my colleagues Janarth, Yuvraj, and Lorrie at CMU. Next slide, please.

Many of us have the experience of purchasing a smart device for ourselves or others. Now, I want you to remember the time when you were in the physical store or searched online for the smart device to buy. You probably saw the information on the price of the device or its technical specification, for example, its size or internet connectivity.

But do you remember seeing any information about the privacy or security practices of the device at the time of purchase? The answer is probably no because, usually, this information is nowhere to be found. When purchasing a smart device online or in a store, consumers are not able to make an informed-purchase decision, as information on the security and privacy behavior of the smart devices is not readily available to them. Next.

To help inform consumers' purchase decision-making process, we designed an IoT privacy and security label, somewhat similar to nutritional labels for food items. We talked about the details of this label in our paper at Oakland 2020. Our label has two layers-- a primary layer and a secondary layer that can be accessed through the primary layer by either scanning a QR code or typing in the URL. We designed this label mainly based on inputs from a diverse sample of privacy and security experts from academia, industry, government, and NGOs.

From the literature, however, we know that experts understand it could be different from consumers. And it's important that the label conveys risk to consumers accurately and potentially impact your willingness to purchase the smart device. To assess this, we conducted a large-scale online study, which I will focus on for the rest of this talk. Next.

In our survey, we asked participants to imagine purchasing a smart device with a privacy and security label. The label talked about only one privacy or security attribute value. Each participant was assigned to answer questions related to three purchase scenarios.

This is an example of a scenario that we presented to participants. The [? words ?] inside the brackets are the factors that [? were heard ?] in the scenarios. Imagine you are making a decision to purchase a smart speaker with voice assistant for yourself. This device has a microphone that will listen and respond to your voice commands. The price of the device is within your budget, and the features are all what you would expect from a smart speaker with voice assistant.

On the package of the device, there's a label that indicates the following privacy and security practice-- purpose of data collection, tailored advertising, and monetization. At the end of each scenario, we ask participants to specify on a larger scale how each attribute's value would change their risk perception and willingness to purchase and why. Next.

At the beginning of my talk, I showed you the layered label that we previously designed based on inputs from experts. In this study, we selected a subset of the label factors to assess. Next slide, please.

For each attribute, we selected two values which we hypothesized to be the most protected and the least protected. For example, for the purpose of data collection, videos providing mainly device function as the most-protected value and monetization as the least-protected value. We hypothesized that the most-protected values of attributes should significantly decrease the perceived risk and increase the willingness to purchase. We expected the opposite for the least-protected values. Next.

We designed and mixed between subjects and within subjects' surveyed on Amazon Mechanical Turk. The statistical models indicated that the impact of almost all tests of privacy and security-attribute values are risk perception and willingness to purchase, per as we hypothesized, meaning our hypothesized most-protected values significantly decreased the perceived risk and increased the willingness to purchase. And the least-productive value significantly increased the risk perception and decreased participants' desire to purchase the smart device. There were a few exceptions though. Next.

Based on the factor coefficients from the regression analysis, we ranked the effectiveness of attributes in terms of changing participants' risk perception and their willingness to purchase. Started with the risk perception, the top-three attributes that significantly increased the perceived risk were sharing data with third parties, providing no control over access and selling data to third parties. On the other end of the spectrum, the top-three attributes that decreased the risk perception were behind multi-factor authentication, sharing data with nobody, and having no cloud retention. Next slide.

In terms of willingness to purchase, we found very similar trends. The top three factors to increase participants' desire to purchase a smart device were no device retention, no cloud retention, and sharing data with no one. And the top three to decrease their willingness to purchase were sharing data with third parties, having no access control, and selling data to third parties. Next slide, please.

The open-ended explanations help us identify several reasons as to why some participants' attitudes toward risk perception and willingness to purchase were different from what we have hypothesized. Starting with the average time to patch, based on the UL guidelines in our study, we selected the most-protected value of this attribute to be one month and the least-protected value to be six months. In our models, both values significantly increased their perceptions and decreased their willingness to purchase.

As you can see, this is a jitter plot of the average time to patch. The x-axis indicates the impact of the attribute values on the risk perception. And the y-axis shows their impact on willingness to purchase. The blue points represent the most-protective value, in this case, one month. And the red dots represent the least-protective value. Here's six months.

And as you can see, most of the blue and red points are clustered in the bottom right, showing an increase in risk perception and decreasing willingness to purchase. As a best practice, manufacturers should patch the vulnerabilities into shortest time. But sometimes, depending on various factors, it could take longer to issue a patch. Our findings suggest that manufacturers need to provide consumers with justifications as to why it takes them a specific amount of time to patch a vulnerability and why it might not be practical to patch their vulnerabilities faster. Next.

Purpose of data collection was another factor that did not change some participants' risk perception and willingness to purchase as expected. Although we hypothesize that providing data for device functionality should decrease the perceived risk, that was true for only 12% of participants. Other participants stated that this information would not impact their risk perception or would even increase the risk, mostly due to their lack of trust in manufacturers. A participant mentioned the companies who collect data are incredibly untrustworthy and do not have consumers' best interests in mind and are utilizing the data they collect. Next.

We also noticed a few privacy and security misconceptions. For example, some participants thought that no security updates indicate better security, as the device does not need to be updated. A participant said, if there are no updates, then the system must be providing maximum security already. Next.

Another misconception was related to mentioning the average time to patch. Some participants believed that even mentioning the word "patch" indicates that the device is not secure, as it needs to be patched. A participant said, on the label, it advertises that patches are even needed. That is why there is a perception of decreased privacy. Next.

To recap, we explored the efficacy of our previously designed IoT privacy and security label in conveying risk to consumers and influencing their willingness to purchase. To do so, we assessed the impact of the subsets of label privacy and security factors and conducted a large-scale online study. We found that in most cases, the label was effective to change consumers' risk perceptions and informed their willingness to purchase.

However, we found a few exceptions. By qualitatively coding the open-ended responses, we surfaced participants' privacy and security misconceptions that impacted their risk perception and willingness to purchase. If you want to get more information about our label project and see how you can help with this effort, please visit iotsecurityprivacy.org. I'm Pardis Emami-Naeni, and thank you very much for your attention.

LINDA HOLLERAN KOPP: Thank you, Pardis. And now last but definitely not least, we will hear from Genevieve Liberte, a graduate student in computer engineering at the Florida International University, where she also works at the school's Cyber-Physical Systems Security Lab. She will present on her paper, "Real-Time Analysis of Privacy (un)Aware IoT Applications." Genevieve.

GENEVIEVE LIBERTE: Thank you, Linda. And good afternoon, everyone. My name's Genevieve Liberte, and the title of this presentation, as Linda said, is "Real Time Analysis of Privacy (un)Aware IoT Applications." This research was a joint effort between Florida International, Purdue University, and Penn State. It was also recently presented at the Privacy Enhancing Technology Symposium of 2021. Next slide.

So in the world of the Internet Of Things, or IOT, users install IoT applications to manage and control smart devices, like thermostats, smart locks, and cameras. Apps necessarily have access to sensitive data to implement their functionality, communicate with external servers, and send notifications to users. However, this access to sensitive data can have negative privacy implications. Some IoT apps have been shown to leak sensitive information to unauthorized parties, and many apps also transmit user data to remote servers for data visualization and behavioral profiling.

Despite entrusting their apps with this data, users have little knowledge or control over what sensitive data leaves the app or is shown to third parties. On the left of the slide, we see a portion of an example IoT application source code written in the Groovy programming language. App source code typically includes a description block. And you can see at code point 1 that this app-subscription block is pretty vague. It doesn't do a great job of describing how the user's sensitive information will be utilized or shared.

During the installation process, the user will grant permissions to the app and authorize specific recipients for notification purposes, which will populate the permissions block variables as seen at code point 2. To support its functionality, the app may include functions which transmit data over the internet, despite not receiving explicit authorization from the user to do so. This sort of behavior can be seen at code point 3.

It's also possible for an application to hard-code additional recipients to whom sensitive data may be sent without the user even knowing. This can be seen in the behavior of the "leakinfo" function down here at code point 4. Next slide.

Now, from the example, on the previous slide, we can extrapolate four main privacy challenges that constitute the threat model we're focusing on-- first, privacy behavior from apps. Apps may access private information without the user's consent. Second, sensitive leaks. Sensitive data may be leaked to unauthorized recipients through malicious or carelessly developed IoT apps.

Next, undisclosed or malicious-app content. Apps may not properly inform users how their data will be used or why it's required. Sorry about that thunder. Finally, unprotected data flows. Apps may not protect external data flows, leaving sensitive data vulnerable to eavesdroppers while it's being transmitted over the wire.

When looking at ways to protect against these threats, at present, none of the major IoT platforms provide a way to analyze app privacy risks or inform users how their sensitive information will be utilized. And most existing third-party tools for analyzing source code for privacy risks use static analysis, which may not catch information leaked through hard-coded recipients or through variables and device states that the user hasn't explicitly defined. Next slide.

in thinking about a solution to these issues, we wanted to ask existing IoT users how they could be better served by potential IoT privacy tools. We designed a survey asking about three main subjects-- the experience and demographics of the participants, the privacy concerns they have with IoT, and the need for IoT privacy-analysis tools and their usability requirements. We had 112 participants in our survey, many of whom belong to an educational institution and had completed at least a bachelor's-level education.

We found that over 2/3 of participants were concerned about their personal data, habits, location, and device states being handled and shared by IoT apps. The majority of users expressed concerns about the privacy of using IoT systems, and many were aware of privacy issues and IoT through news stories or other media. When asked about the idea of a tool to uncover privacy [INAUDIBLE] in IoT systems, almost 97% of our participants found this idea to be highly desired and expected. Regarding their expectations for this idea, participants expressed a shared desire for a user-friendly tool that can be configured with various privacy preferences and provide real-time notifications when privacy threats are detected. Next slide, please.

So taking into account the feedback we gathered from our survey, we arrived at our proposed solution, IoTWatch. IoTWatch is a dynamic analysis tool to uncover privacy risks that IoT apps pose to the privacy preferences of users in real time. IoTWatch works in three stages-- instrumentation time, which happens before the user configures their preferences; install time, at which the user defines which privacy labels they wish to see notifications for; and runtime, when the app is actually running on the user's smart device.

During the instrumentation process, the original application source code is spent to the IoTWatch instrument or tool. IoTWatch analyzes this source code and builds an intermediate representation of the app to determine how and when each function is called and what variables may be passed to them. IoTWatch uses this intermediate representation to then insert bits of its own code into the app so that data sent out of the app during runtime will also be sent to IoTWatch with servers for analysis and matching with the user's preferences. Once the code has been instrumented, the user configures their privacy preferences with IoTWatch's graphical user interface in their app. After this process, the user can begin to use their app normally.

During runtime, IoTWatch collects all of the function calls that result in data leaving the app and uses natural-language processing to determine whether any of these function calls result in sensitive data being leaked according to the user's own configurations of what data they want flagged as sensitive. If data lakes are found, IoTWatch immediately sends a notification to the user, detailing the type and source of [INAUDIBLE]. Next slide.

So how does IoTWatch's source-code instrumentation work? It begins by generating a genericized version of the original source code, called an intermediate representation. Most IoT apps follow similar structures, even on different programming platforms. And using the intermediate representation allows IoTWatch to be applicable to these different IoT programming platforms.

From this intermediate representation, we're able to determine places in the code where sensitive information is passed to the functions. We also identify all of the places in the code at which data exits the app to be transmitted, which are known as syncs. IoT watch specifically focuses on internet syncs and messaging syncs, where data is sent out via the internet or via SMS, respectively.

Knowing all these locations allows IoTWatch to construct the app's Abstract Syntax Tree, which is then analyzed by a custom node-visitation algorithm to produce an intra-procedural control flow graph of all the functions in the source code. Once we have the graph of how data and functions flow throughout the app, IoTWatch can identify user-defined inputs in the app's permission block, as well as the recipients and contents of the sink-calls. This control flow graph is also what enables IoTWatch to add its own code to the app. This extra code is used to collect and transmit sink-call data to IoTWatch's server and implement real-time push notifications to inform the users about IoTWatch's analysis results. In the next slide, we can take a look at how IoTWatch's data collection works in more detail. Next.

So on the right here we have our example IoT app after being instrumented by IoTWatch. At point C1 in the code, we see the user-defined input of the user's phone number in the app's permission block, which IoTWatch has identified. It uses these inputs to determine that the phone number is a legitimate sink-call recipient, which has been provided by the user. At points in the code where sink-call occur, like line C2 and C4, IoTWatch adds additional code to analyze sink-call and inform the user if these sink-call lead to violations of their privacy preferences.

For instance, on the left is an example of the user's configured privacy preferences. We see that they have chosen to be informed whatever device information or location information is transmitted but not daytime information or user behavior. Correspondingly, at point C3 and the code, IoTWatch has inserted a notification event to let the user know that this sink-call involves device date information. Similarly, point C5 in the code represents how IoTWatch instruments this "leakinfo" function, informing the user of both the kind of information being handled and the privacy violation due to the info being sent to an unauthorized hard-coded recipient. Next slide.

So once the source code has been instrumented and the user configures their privacy preferences, the user installs the instrumented app to their device. The app transmits its data to the IoTWatch server whenever a special data [INAUDIBLE] flagged. This information permits IoTWatch to identify the type of sensitive information an app uses and combines this info with other app data, like the description block, to uncover sensitive data leaks. IoTWatch leverages a REST API to securely exchange this data between the app and the IoTWatch analysis server.

IoTWatch classifies the content of sink-calls according to four main labels-- device info, date-time, location, and user behavior. These labels were chosen based on our survey results. And a given string could be assigned multiple privacy labels, depending on the specific information it conveys, like the string, "the door will remain open for another five minutes," which pertains to both device info and date-time info. Next slide, please.

The classification of sink-call contents in IoTWatch is achieved through Natural Language Processing, or NLP. NLP is a machine-learning technique to process and analyze constructs inherent to human language. To build our model, we collected the sink-call contents of 380 different IoT apps from the Samsung SmartThings marketplace. We filtered out any punctuation or extraneous words in these strings and then manually labeled each one as belonging to the four different privacy labels.

We then built a classifier using several NLP and machine-learning frameworks. To train this classifier, 80% of the 380 apps were used as a training set, and the remaining 20% were used for the test set. We found that overall, our classifier could achieve 94.3% accuracy when classifying sink-call content as the four privacy labels. Next slide.

We collected a total of 540 Samsung SmartThings apps. 380 of those were used to develop our NLP model, which I explained on the previous slide, and the remaining 160 of these apps were used to evaluate IoTWatch's performance. Of these 160 apps, 120 were taken directly from the Samsung SmartThings marketplace, and the remaining 40 were from malicious apps from the IoTBench repository, an IoT specific corpus of apps used to evaluate systems for IoT privacy and security.

We first tested IoTWatch's ability to classify sink-call contents into the four privacy labels, the result of which can be seen here at the table in the top right. IoTWatch converted strings into privacy labels, with an average of 93.98% accuracy and 97.33% specificity. We observed the highest accuracy for the date-time, and location-info categories, likely because these are the most obvious types to identify using NLP because they're more likely to contain strings with numbers.

Next, we evaluated IoTWatch's ability to detect sensitive data leaks. We split the leaks we were testing into leaks-via-internet sinks and leaked-via-messaging sinks. We found that IoTWatch was 100% effective at identifying sensitive-data leaks-via-internet sinks for both the market apps and the malicious apps. For messaging sinks, IoTWatch was also 100% effective, though only the malicious apps contained sensitive data leaked via messaging. Next slide, please.

In conclusion, we developed a privacy-analysis tool, IoTWatch, to perform source-code analysis and instrumentation of IoT apps to collect data and enable privacy analysis using natural-language processing. IoTWatch demonstrated 100% effectiveness at identifying data leaks and privacy concerns in accordance with 540 Samsung's SmartThings IoT apps and was able to classify privacy-related data flows with 94.25% accuracy. IoTWatch is able to achieve all this while only adding an average of 105-millisecond latency and around 1 kilobyte of extra data to the app.

Overall, we hope that our success with IoTWatch can demonstrate that it is possible to develop tools that provide this granular of level of control to IoT users. We learned that users have strong desires for more transparency and control over their privacy-sensitive information in IoT apps. And IoTWatch is proof that these desires can be met. While IoTWatch is the first dynamic analysis tool that achieves this, to our knowledge, hopefully, this can be a good first step towards tools such as these becoming widely available to and expected by IoT consumers in the future. Next slide.

So thank you so much for your time. I just wanted to thank my Professor, Dr. Selcuk Uluagac, and the main author of this paper, Dr. Leonardo Babun, as well as the National Science Foundation, for supporting this research. Also, I apologize if my connection was unstable. It started raining very hard, so sorry about that. But thank you so much.

LINDA HOLLERAN KOPP: Thank you, Genevieve. Your connection was great.

GENEVIEVE LIBERTE: OK, good.

LINDA HOLLERAN KOPP: As a reminder to our watching audience, if you have any questions, please email those questions to privacycon@ftc.gov. So I wanted to open up the discussion to our panelists. And I think it is interesting that one of the themes that [INAUDIBLE] emerged today-- and it's certainly true in the area of privacy more generally-- is a desire for more transparency. And two the tools that were discussed today are designed to provide that greater transparency. What is happening to our data?

So my first question is for Jeffrey and Genevieve in your Skill Detective and IoTWatch tools. Are they still in development? And do you have any plans to make them available to researchers or others? Maybe we can start with Jeffrey.

JEFFREY YOUNG: Yes, yes, Linda. Actually, the source code is still in development. But our paper is currently in evaluation. Once the paper gets published, then yeah, we will be publishing the source code to make available publicly to anybody who wants to use it. In fact, yeah, we're getting everything together now, and we have a GitHub repository set up, and so everything should be available to the community.

LINDA HOLLERAN KOPP: That's great. And Genevieve, what about IoTWatch?

GENEVIEVE LIBERTE: It looks like we may have lost Genevieve. Her connection problems occurred. Hopefully, she will be able to join us back. Jeff, I did want to ask a follow-up question on your presentation. Can you speak a little bit more about some of the effects your work has had so far on the voice personal assistance, like Alexa and Google Home?

JEFFREY YOUNG: Sure. Well, like I said in the presentation, Google did take down some of the actions that we had reported. But we have spoken with Amazon, as well. But we have not been able to really show that any action has been taken as of right now. But we are in communication with Amazon, and we're talking with one of the developers there who's in charge of the Alexa platform.

And so we're hoping that some of our work may lead to some changes or something along those lines later on down the line. But right now, we're just not really sure because they-- for a while, we contacted them, and we didn't really hear very much back. And then all of a sudden, we started talking, and then now we're back into that stage where we're not hearing very much back again.

LINDA HOLLERAN KOPP: Do you know if the skills that you had identified to them as containing policy violations are still active?

JEFFREY YOUNG: Yes, yes. There are quite a few of them that are still active. We have not tested all of them as of yet, but we are planning to go back through and test everything again. Also, we have an updated version of our system, so we're making the system smarter and can analyze deeper and deeper into each skill.

It's not a very easy task to talk back and forth, have two computers basically talk to each other. But yeah, we're planning on going back through and doing it all over again and keeping a database of the interaction models between each skill to see how skills change over time. That's one of the details, too, about having third-party developers and the source code on the back end is that skills can be updated, and they do not have to go back through a vetting process for that update. And so we have-- we can only see the skills-- how the skills interact, what the skill does. And so we're planning on keeping track of that and developing a data set down the line.

LINDA HOLLERAN KOPP: And Anupam, I thought you would probably be interested in Jeffrey's paper given your own research and a little bit of the overlap in some of the privacy issues described. You gave some helpful observations or suggestions at the end of your presentation about ways that Amazon could better protect its users' privacy. Did you have any additional thoughts about that, especially in light of what you heard from Jeffrey?

ANUPAM DAS: No. That is a great question, actually. And so we, actually, also had interacted with Amazon when we actually found some flaws, and so we had a similar communication pattern, where we had a flurry of exchange at the beginning. And then suddenly, there's no communication, but it revived again.

We did have a long conversation, about a one-hour meeting, about some of the findings and our recommendations. And they were interested in looking at those recommendations, but I think that one of the things that they're still looking [? for ?] as whether the recommendations really work. And so this is where, I think, we need to do a little bit more research in terms of what recommendations we're making and whether that's really impacting the users and making the right choices, or even making them aware about some of the gaps that might exist.

So I think we're currently focusing on some user-oriented study in that aspect also. And I think it would be great once we've done some of this, conducted some of this research, and then go back to them saying, OK, these things really work. At that stage, we were just making recommendations without any proof in that context. So I think that would be the best way to get the ball rolling again with them, after we've done some of those user-oriented studies.

LINDA HOLLERAN KOPP: And do you have any suggestions for how consumers, the users of the voice assistance, can be more vigilant about privacy with the voice-based skills?

ANUPAM DAS: Yeah, so from a user's point of view, whether the system supports any additional awareness or indicators, I think, from a user's perspective, there are certain things we can always adopt. And some of the things basically is that the [INAUDIBLE] feature, I think, could potentially lead to [? this ?] activation. I kind of explained that.

And we tried that ourselves. And then many of the times, users really get it wrong based on what they think is going to activate and what really activated. So one thing I would suggest is that if you do interact with the skills and you see certain skills being installed or activated in your account, do go back because you can go back and see what's activated right now, and see whether that actually matches what you really wanted to install.

And the other suggestion would be that if, at any point in the interaction, you feel like the skill is asking for something that doesn't make sense with the functionality-- suddenly they're asking for your phone number or location or ZIP code-- then that's another great pointer to stop and rethink, why is this requiring this information? Because there are many other alternative skills out there which might not ask this information, but you might still get the same services, right? So that's another thing as an end user you can do.

And the last thing is that, basically, a lot of the times, we interact with a skill or activate a skill just for fun. And this tendency also happens in mobile apps. We install apps for fun. We never use them in long term.

So I think a periodic checkup is something that we can also get ourselves used to, saying that, OK, in a month, let me see what skills I've activated on my account. Do I really use all of them frequently? If not, then this is another point when we probably want to deactivate or disable that skill. I think these are some of the practices or guidelines that we, as a consumer, can follow. And I think those will definitely help in reducing the risks to some extent.

LINDA HOLLERAN KOPP: And Pardis, I noticed that you were nodding in agreement to Anupam and his recommendations for users. I was wondering if, given your studying related to your privacy label and your research about consumer's risk perceptions, if you have any thoughts about ways users can protect themselves and, in particular, educate themselves about those perceptions, those risk perceptions.

PARDIS EMAMI-NAENI: Yeah. That's a really great question. And I think I just really don't want to put a lot of blame on the consumers here. I think the projects that we're conducting for IoT label, it was like, really, the idea here is manufacturers should really disclose what they're doing. And then it is on users to read, for example, that information, if it's, for example, in a usable format because we know that privacy policies-- people are not really reading them.

So I think people should really educate themselves. However, this information should be available somewhere. And I think this is a very [? serious ?] gap, this huge gap-- that we would like manufacturers to really tell consumers in a understandable language what they're doing and how they're protecting them and, basically, all their privacy and secure practices that are relevant to consumers and consumers' data.

But I think if that information is available, yes, they need to read that information. Consumers need to educate themselves. And if they see something that is questionable to them, they really have to, for example, maybe contact a manufacturer, or maybe contact privacy-security experts if they know them, or somehow raise that concern so that others who are more expert in this issue can really help them or can at least help them protect themselves if they cannot make it better or anything, at least help consumers protect themselves.

LINDA HOLLERAN KOPP: Out of your research, did you find any instances where risk perception was not aligned with the consumer's willingness to purchase the product?

PARDIS EMAMI-NAENI: Yeah. That's a great question. So yeah. So as you said, we basically assess both risk perception and willingness to purchase. And in most cases, these two were aligned, which, basically, I think, makes sense. But there a few exceptions. For example, consumers understood the risk, but that risk was not enough for them to change their willing-- their desire to purchase the device.

For example about multifactor authentication-- people knew that this was going to decrease their risk, and they perceived lower risks then the device had, for example, multifactor authentication. But then they told us that this information is not going to, basically, help them purchase a device. And in some cases, purchases are not going to purchase the device because of multifactor authentication because of its usability challenges.

Another example was about a security update. People knew that automatic updates are better than no update, for example, or better than manual updates in terms of risk. However, they said that they still would like to have control over the update, and they still prefer manual updates over automatic updates.

So I think it shows that risk is not enough. The label should not just be designed such a way to communicate the risk and that's it. The privacy and secure practices should be designed in a usable way so that not only it conveys the risk, basically not only decreases the risk, but also, they are useable so people are interested in using them and are, basically, going to purchase the device because of those features, not going to turn away because of the usability issues.

LINDA HOLLERAN KOPP: [? I thought it ?] was interesting-- your paper reported on the self-reported purchase behavior of the consumers. And there's always a lot of debate about what people say and what they actually do as it relates to privacy. Do you have any thoughts or expectations on how your IoT label could impact real purchasing behavior?

PARDIS EMAMI-NAENI: Great question. So I think the main reason that we did this study, like this line of study on IoT label in self-reported fashion was that we do not have the labels. We can not have the labels in-- [? for, ?] like, the real purchase behavior because devices do not have labels, basically. So we can not really test that in realistic purchase behavior.

But if you look into other literature, for example, food literature, we know that consumers who are more interested in having better health, for example-- those are the ones who would look into nutrition labels, or those who have more knowledge would look into nutrition labels. So there are different factors that might impact how interested you would be in looking for that information and how that information would impact your preferences and desire to purchase, for example, the device, such as interest, such as knowledge. So we don't really know for sure how our label is going to impact real purchase behavior, but based on what we heard from the consumers in all the studies that we've conducted, we know that the label is understandable to them, but we don't really know whether that would change their willingness to purchase if being presented with actual purchase behavior.

LINDA HOLLERAN KOPP: Genevieve, I'm glad you worked out your connection problems. Welcome back.

GENEVIEVE LIBERTE: Thank you.

LINDA HOLLERAN KOPP: So I wanted to go back to a question before you dropped off about whether or not the IoTWatch tool was still in development and if there are any plans to make it available to researchers or others.

GENEVIEVE LIBERTE: OK, great question. So IoTWatch is currently only implemented for the SmartThings platform, but we would like to make the tool publicly available down the line and implement it for other platforms, like Open [? Hub ?] as well, following a complete privacy audit of the tool just to make sure that it's working properly and protects data properly.

At present, we do have a demo version of the IoTWatch instrument publicly available, and that can be found at iotwatch.appspot.com. And this demo allows users to input their IoT app-source code and returns the source code after having its [INAUDIBLE] calls and important strings flagged and instrumented. And so that instrumented source code won't be able to be used by anything because the portion of IoT, the analyzer isn't publicly available yet. But yeah, that little demo is available so that people can try out the instrumenter and see what it would do to their own code.

LINDA HOLLERAN KOPP: That's great. And I understand-- and correct me if I'm wrong-- that the way IoTWatch works is it sends some of the data to its own servers for its analysis. Are there measures in place that protect the data?

GENEVIEVE LIBERTE: Yeah, so I IoTWatch uses TLS to secure the in-transit app information. And IoTWatch doesn't collect any personally Identifiable Information in addition to the information that's already included in the strings that's being sent to it from the app. Also, our tools don't fingerprint or expose any user activity whatsoever because IoTWatch doesn't actually collect information and store it. it just is sent the information, analyzes it, and then responds right back. And as part of IoTWatch, we also include a tutorial that explains to the user what we do with that information we collect. And that can be found in the paper, as well.

LINDA HOLLERAN KOPP: If a developer used encryption, would that allow them to defeat or evade IoTWatch's analysis?

GENEVIEVE LIBERTE: So the way that IoTWatch works right now, it doesn't account for encryption. But if an app were to be found encrypting its strings, there are some ways around this. So IoTWatch works by extracting IoT strings via flagging and instrumenting sinked calls in the IoT apps. And in the case of an app implementing encryption, the apps would need to implement the encryption functions before sending the communication, and therefore, the IoTWatch analysis could be easily modified to just extract the app information before the encryption step, exposing the IoT string in plaintext to our tool anyway.

However, we found that that's actually a moot point from what we've seen so far because out of the 540 SmartThings apps that we looked at, none of the apps encrypted the IoT strings further than just the expected TLS encryption layer. So if we had encountered an app that was encrypting data, that in itself would have flagged it as suspicious to us.

LINDA HOLLERAN KOPP: OK. And I'm interested in the thoughts of everybody on the panel about future areas of research regarding IoT that you think would be particularly useful. And maybe Genevieve, I'll start with you.

GENEVIEVE LIBERTE: OK. To echo what I heard Pardis saying, I definitely think that one of the biggest areas of research, or just development in general in the IoT space, needs to be some sort of regulation for describing the privacy impacts of IoT devices to consumers. I think that if something like that was in place, we wouldn't have had to-- IoTWatch wouldn't be necessary because the apps would be exposing what they do with private information themselves.

And so I think that, in terms of future research, we need to just be looking at better ways of analysis, maybe more dynamic-analysis tools, as well as static analysis, just to analyze how apps are running and what they do with the information, not only that users put in but also, the information that the apps themselves are dealing with, just in terms of, like, whether a door is open, or things like that.

LINDA HOLLERAN KOPP: Great. Pardis?

PARDIS EMAMI-NAENI: So I think, what I feel is missing here is, really, the realistic purchase settings and really understanding that their consumers would understand privacy and security information when purchasing devices at the time of purchase, at the point of sale. And I think in other countries, for example, in Finland and Singapore, we already have IoT labels. We don't have that in the US for sure, but I think if manufacturers are, for example, going to adopt a label, maybe going to just be in a pilot study to just adopt a label, and then we can study them and see when their consumers are understanding this information, that they understand the risk, and things like that, I think this is really important here because this can push this effort forward a lot. We cannot just continue working on self-reported responses. We really want to understand the realistic behavior.

And I think the new White House executive order asked NIST as well as FTC, to work together to basically conduct a pilot and look into the efforts into labeling smart devices. So I think this is now gaining some interest in the US's bill. And I'm actually very optimistic about this, that this is-- basically, US is also going to look into this effort. And maybe in the very near future, we are going to have these labels for devices. And then a whole new set of new research studies, basically, is going to be conducted after we have those labels.

LINDA HOLLERAN KOPP: And Anupam?

ANUPAM DAS: Yeah. So for the context of voice interfaces, I think one of the interesting research question is going to be, how do we design effective indicators, or interventions, for voice interface, because in many ways, when users are interacting with voice interfaces, they typically think whoever's the vendor, that's the company that's doing all the work. But when you open up platforms to third parties, that becomes really more tricky.

Designing effective indicators or voice-based interventions-- I think that is going to be challenging, and is also going to be interesting because we need that because users, we know, at some point, are not always thorough enough to check all of the information by themselves. So we need to place interventions or indicators as much as possible throughout their interaction through the various platforms. And I think that's going to be an open research problem that we're going to see in the very near future.

LINDA HOLLERAN KOPP: What about you, Jeffrey? What are your thoughts?

JEFFREY YOUNG: Well, one of the things I think we've found over testing this far is that, actually, developers, a lot of times-- we don't believe that they understand that they're actually violating policies. They're just developing code in a bedroom or something on those lines. Some of our future research, actually, is in the field of being able to test source code before it even goes to the platform-- so some sort of tool that you can test your own source code for policy violations, that sort of thing.

Also, a dynamic-permission model-- so because it's voice activated-- or it's voice interaction, it's very difficult to ask permission. So how can you design a permission model that's accurate that would actually inform the consumer? Do you give permission for this particular skill to collect this particular data at this particular point? And that sort of question-- that's going to probably be an open question for a while just because of the platform itself.

LINDA HOLLERAN KOPP: Great. So we are about at the end of our time. I want to encourage everybody to go to the PrivacyCon page of ftc.gov where you can read our presenters full research papers. They're well worth your time.

We are now going to take a short break, but please stick around for our panel on privacy in children and teens. I want to thank our panelists for really interesting work and research and sharing it with us today. And thank you, everybody, at home for joining us. Thank you, everyone.

[MUSIC PLAYING]

Welcome back, everyone. And welcome to panel 5 of PrivacyCon 2021. My name is Manmeet Dhindsa, and I am an attorney in the FTC's division of Privacy and Identity Protection.

I'm joined today by two fantastic panelists who are here to talk today about privacy issues related to children and teens. You can all find our full bios posted online by visiting ftc.gov. But for purposes of this panel, I will just quickly introduce both presenters and their work so we can just hop right into the discussion.

Our first presenter today will be Mohammad Mannan from Concordia University in Canada. And he'll be presenting on his papers titled, "Betrayed by the Guardian-- Security and Privacy Risks of Parental Control Solutions" and a second paper titled, "Parental Controls-- Safer Internet Solutions or New Pitfalls"?

Our second presenter will be Cameryn Gonnella from BBB National Programs. And she'll be presenting on her paper titled "Risky Business-- The Current State of Teen Privacy in the Android App Marketplace." I want to take a quick second to make a correction regarding a previous iteration of Cameryn's bio. For clarification, the research. we'll be discussing today was conducted wholly in-house by BBB National Programs and did not include any outside funding.

After each presentation, we'll have a brief Q&A with each individual panelist, and then, hopefully, we'll have some time at the end to come together as a group for a group discussion. If anyone in the audience has any questions as we're moving along, please feel free to email them at privacycon.ftc.gov, and we can hopefully get to them, time permitting.

And with that, like I mentioned, we have a fantastic panel ahead of us, so let's just jump right in. And Mohamed, with that, I'll pass it over to you for your presentation.

MOHAMMAD MANNAN: Thanks, Manmeet. Thanks for the introduction. Hello again. My name is Mohammed Mannan. I'm going to present our work on parental-control-solutions analysis. This is a joint work with my collaborators here, Suzan, Mounir, Quentin, and Amr from Concordia University Canada. This work was presented at [INAUDIBLE] last year. Slide 2, please.

Parental-control solutions are seen as a necessity by many parents to keep their children and teens safe online, which has become a very significant issue, even before the COVID catastrophe. Many products are also out there to help parents in this regard, and these products also come with a lot of safety promises. I'm quoting here one product which claims that parents do not need to worry and dance over their children's shoulder. And the product will take control over all internet activities. And these type of products are also recommended by some trusted government sources, such as the US FTC and UK's Council for Child Internet Safety. Slide 3, please.

From a very high point of view, parental-control solutions work as follows. The solution will sit between the children's devices and all external web services. The solution can be an app or an application installed on a device, or it can simply be a browser add-on, or implemented in a separate independent network device. The solution will check all outgoing network connections and in some cases, messages and then allow the ones that are deemed to be safe. Slide 4, please.

We analyze representative solutions from multiple platforms, including Android, and Windows operating systems, Chrome browser add-ons, and independent network devices. We did it for multiple platforms just so that we can have a comprehensive view of this domain from a security and privacy perspective. Slide 5, please.

So to enable parental-control functions, these solutions generally require some powerful privileges. For network devices, they generally monitor all external domains and plaintext traffic from a network vantage point, but usually, they don't intercept DLS encrypted traffic, which is done by some Windows applications. In Chrome add-ons, they need to see all browser data.

For Android apps, they require some really serious permissions, including Device Administration, and Device Management, and in some cases, even super-user permissions. Some of them also require to have access to Accessibility Services to monitor all user actions on the device, window content from other applications, phone calls, SMS messages, and real-time location information. Some of them also require authentication credentials for other social-media platforms, like Facebook and YouTube, to monitor the comments and messages in those platforms.

So because these platforms, these solutions, are highly privileged and they also deal with children's data, we wanted to know if they are secure enough to prevent simple attacks and if they themselves violate user privacy by collecting unnecessary personal data or by exposing personal data to third parties. For this, we designed a test framework and analyzed several selected solutions. Please see our paper if you're interested in the test framework, which I'm not going to discuss much here. In summary, we checked the solution [? code, ?] traffic generated by them during their usage, and also, their online interfaces. Slide 6, please.

Our result summary for these solutions isn't quite pretty. Among 54 solutions that we tested, we found 172 privacy- and security-related vulnerabilities. The majority of these vulnerabilities are in Android apps, but several network devices, Chrome add-ons, and Windows applications are also similarly vulnerable. I'm not going to discuss a whole lot of these results, but I'm going to present a few example vulnerabilities. Slide 7, please.

The first example I have here is an insecure update mechanism in Blocksi network, a parental-control device. In this one, it blocks the server, sends an updated firmware in plaintext, although the server attaches a cryptographic hash value of the binary firmware. But anyone in the network can replace the firmware and the hashcode with anything they like, including malware, because the hash value does not require any secret to compute, and the binary is also not signed. Slide 8, please.

The second example I have here is an Android app called SecureTeen. Like several other apps of this category, SecureTeen stores children's activities on their server side and allows parents to check the data at a later point in time. But unfortunately, they provide an API while you only need a parent's email address to have access to the parental account without knowing the account password. Slide 9, please.

Some other notable results from our analysis include the following. We have seen 13 solutions that allow unauthenticated access to their server-side data similar to the SecureTeen example that I discussed in the previous slide. Eight solutions-- they send personal data over HTTP in plaintext. And 16 others-- even though they use HTTPS, they can be easily downgraded to HTTP.

Six other solutions allow easy parental account takeover. And we have analyzed some solutions that are certified under the FTC approved KeepSafe program. And we found that they use third-party trackers and in some cases, also expose personal data, including even account credentials. Slide 10, please.

Now I'll discuss some examples of what an attacker can do with the vulnerabilities that we exposed. The control of a network device will enable the attacker to monitor all network devices and activities and even use the parental-control device in other malicious attacks. By having access to the parental account, this will be quite devastating because this may enable full control of the child's device to the attacker. The attacker, in this case, can install or remove applications from the device allow or block phone calls and internet connections and even access real-time location data from the device.

The unprotected servers that we have found from them-- an attacker can access the data collected from over half a million users, most of whom are teens and children. The use of HTTP can allow an attacker to drop or modify some very sensitive messages, like an SOS message, which is supposed to be sent when the child is in actual danger. Slide 11, please.

So overall, most parental-control solutions that we have analyzed do not meet privacy expectations and often introduce new attack avenues and make their users vulnerable. Also, as these products are seen as essential, parents cannot simply delete them, just like a game or other unessential applications, which, if that they are not meeting your privacy or security expectations, you can simply get rid of them.

So we suggest that these products should be designed in a way that even if they don't provide perfect functionality, they should do no harm in terms of privacy and security exposure. And if and when there is a bridge, the solution-- providers must accept liabilities for those bridges. And only strict regulations can make that happen. Slide 12, please.

Regarding the vulnerabilities that we found, we contacted all companies multiple times and still could not get a response from some. Few of them actually fixed their products, but still, many vulnerabilities remain open several months after we first contacted them.

Finally, I want to [INAUDIBLE] OPC Canada for their support in this project. I want to thank you all for your time and attention. I'll be very happy to take questions. You can also email me afterwards if you have any further questions. Thanks.

MANMEET DHINDSA: Thank you, Mohammed. That was fantastic. And thank you for both, not only the presentation but also, your research. Your research provides some really interesting findings about these parental-control solutions, although many of the findings are quite scary, especially for parents. Do you have any advice on what parents can do to choose a safe and effective parental-control [? solution? ?]

MOHAMMAD MANNAN: So for parents, I mean, I don't believe that most of them are tech savvy, so it would be difficult for them to choose something by understanding their security and privacy consequences. Our report can help to some extent. We also have, actually, a website with details, information on each product. But, of course, we only analyzed some selected set of products, not all products that are available in the marketplace.

So generally, what I suggest is that parents should avoid the products that come with some invasive features because those features, if not well protected, eventually can cause some serious issues. And parents can also check how much data is collected by these solutions in their online-account interface. So whatever data they may see, they should consider that the consequence of that being leaked at some point in the future. So instead of using these third-party solutions, what I would also suggest is that to use parenteral-control functionalities which are now built into most operating systems, even though they are not fancy, but they may be sufficient for most parents.

MANMEET DHINDSA: That's really helpful. And just as a follow-up to that, you mentioned that one of the things that parents should avoid are those solutions with what you called invasive features. What are invasive features?

MOHAMMAD MANNAN: So invasive features are, you know, if the solution can install or remove any app, if it can block phone calls, or SMS messages, or internet connections, which may be important for the child. And if those features are actually compromised, then the child may be harmed in the real world. So parents should be really aware of these features. Whenever you see that they are talking about, OK, we can give a lot of control to you, as the parent, there is a dark side of that control that-- I mean, those features can really backfire at some point.

MANMEET DHINDSA: Thank you. That's really helpful. So if we look on the other side of the coin, perhaps we shouldn't place all of the responsibility on the parents. And I'm curious if you have any opinions on what the developers of these parental-control solutions, or even the app-market operators that offer these solutions, for example, Google and Apple, what they can do to improve some of the issues that you found in your research.

MOHAMMAD MANNAN: So app developers in the marketplace-- they compete with each other, right? So they want to provide as many fancy features as possible, whether those features are necessary or not. So they just want to claim more features than their competitors.

I think they should see it more from the other side of it, that, OK, if we use these invasive features, it can actually cause some issues for us in the future. So they should avoid using powerful privileges if they are not necessary for the functionality of their solution. And they can also avoid using software-development kits or libraries that contain third-party trackers. And to avoid simple mistakes in the design and implementation, they can also try out our test framework.

For the market providers, like Google or Apple, as you mentioned, we know that they do a lot to keep their marketplace malware free, but I think they don't do enough to make it as privacy friendly, especially-- I mean, they should really consider here that when children's safety and privacy is at issue, they should really take it more seriously. And these providers, in fact, they can restrict apps from using powerful features, like Device Administration or Management that I mentioned before, which were designed for some other purposes, not for parental-control purposes. They can simply block these features when an app really doesn't need to use these features.

MANMEET DHINDSA: Thank you. That was fantastic. I'll now turn it over to Cameryn for her presentation.

CAMERYN GONNELLA: Thank you, Manmeet. So as you said, I will be presenting the findings from our white-paper study, "Risky Business-- The Current State Of Teen Privacy in the Internet Marketplace." This white-paper study was conducted by myself and our team with the impetus for a new program we're developing called TAP, the Teenage Privacy Program. And that has the goal of ultimately creating a set of self-regulatory framework and standards for industry best practices regarding teen privacy online. If you have any questions or you'd like to know more about that program, I have all the contact information at the end of my presentation. Next slide, please.

So why did we focus on teenagers as a unique group in this study? We focused on teenagers for two reasons, the first of which, as you can see from the numbers on the slides, is that teens are very online. They're very engaged with mobile apps and social-media platforms. They're downloading apps at least once a month, using social media multiple times a day, and often owning their own smartphone devices.

This has led to teen generations often being referred to as "digital natives" because they grew up with this technology, and they use it so often in their day-to-day lives. It also creates the impression that because they're using this technology so much, that teens are fully aware of the risks that might be involved with online engagement and that they're able to adequately protect themselves from any potential risks online. And in fact, you can see 72% of teens do believe that tech companies manipulate users.

However, the second reason we chose to focus on teens is because of this impression that they're able to handle themselves online, is they're excluded from important policy discussions about privacy. So right now, the focus remains on COPPA, the Children's Online Privacy Protection Act, which applies to children under 13, and protecting young children online. So there's no recognition that teenagers have their own unique needs online, even though we recognize this in other parts of life.

For example, we have age-based safeguards in place for activities, [? like ?] driving. And in the United States, you have to be 16 years old to get your driver's license. Or to vote in elections, you have to be 18. But again, there's nothing like that online.

And this is reflected by the proposed legislation that you can see on the screen that relegates teenagers to the same types of restrictions that are imposed on younger children. And it's unrealistic to think that teens will get parental consent, for example, for their online activities, like COPPA requires for children under 13. So again, the same measures are not going to be effective for teenagers because they behave and engage online differently than younger children. Next slide, please.

So our study ultimately showed that teens do have a greater attack surface for privacy risks. So what does this mean? Our key findings in the study showed that teen-directed apps we looked at requested a high level of permissions. So 11 median and permissions were requested, and six median dangerous permissions were requested, which I'll explain more later. And they also had a high level of trackers integrated, so there was a median of 10 trackers integrated into teen-directed apps. Next slide, please.

For more direct comparison, you can see that our key findings also showed that a majority of teen-directed apps were supported by ads-- so a really high amount, 82%, versus less than half of our general apps that we looked at. Next slide, please.

So before I get into a little more detail about our findings, I'll talk about our methodology. So as the title indicates, we pulled apps from the Google Play store. This is because Apple has tighter controls around their app store and iOS apps, so it's more difficult to analyze them. And we also limited our study to free apps.

So we have two data sets in the study, as you can see on the screen. We had our General Dataset, which we used as a point of comparison to be representative of the whole app store. And we got this by scraping the top 200 apps from each genre in the Play Store and then spidered that out to get all of these similar suggested apps from those top 200 apps in each genre, which led us to get a dataset of almost 54,000 apps.

And then out of those 54,000 apps, we narrowed it down to get our teen-directed dataset, and we did this using a couple of different methods. First, we pulled the apps that had 20 million or more installs. And then we applied a multifactor framework that we created to figure out which apps were likely teen directed out of those. And to figure out that multifactor framework, we first adopted the FTC factors for determining child-directed services under COPPA to teenagers-- so looking at subject matters and celebrities that might appeal to teenagers, like high-school pop culture, things like that.

Then we looked at industry standards that were out there, like the MPAA ratings for movies and the ESRP ratings for video games to see what they rated as appropriate content for teenagers. And then finally, we looked at our own general knowledge of what teens are using. So apps like TikTok are included in this dataset, obviously. And then finally our team-- we finally got a data set of a little bit over 1,100 apps that were likely teen-directed, which our team then used static analysis on to generate our findings. Next slide, please.

So first, I'm going to talk about the monetization methods that we looked at in our study, the first of which being advertising. For some background, there's two types of advertising. There's contextual advertising, which is pretty simple. It doesn't rely on data collection to serve ads. It simply looks at the content that a teen user is looking at to infer their interest and then [INAUDIBLE] limit ad based on that content.

The second type of advertising is a little bit more invasive. It's called interest-based advertising. It's also called "targeted" or "behavioral advertising" because it relies on data collected about the teen-user's behavior to serve them an ad. So as you can see from the flows on the slide, it shows how that data is collected from teen users, and then it's handed over to various third parties who then repackage that data as targeted advertising sent back to teen users.

Targeted ads might not sound like a bad thing, but again, keep in mind teens are already exposed to 30% advertising across the board than general audiences. And in addition, what often happens is that profiles are built around teen users using this data, which reveals a full picture their life and interests. Next slide, please.

The second type of monetization we looked at is in-app purchases. Again, these are ways that apps generate revenue by letting users spend real money on things like power-ups, or upgrades, or even extra content in an app. Our study found that in addition to seeing more ads, teens are getting bombarded with in-app purchases at a much higher volume than general-audience apps. And so as you can see, there was about four times as many general-audience apps with in-app purchases as those without. But in the teen dataset, that skyrocketed to about 13 times as many apps that are offered in-app purchases.

Now, these become especially problematic because in-app purchases often use dark patterns, which lead users tc spend more money and more time in the app and they wouldn't ordinarily. Dark patterns are designed to go undetected, and since adults often fall victim to them, it stands to reason teens are not likely to notice this type of subtle manipulation.

In-app purchases and teen-directed apps also capitalize on peer pressure to drive spending. So as I mentioned earlier, the majority of teens use social media several times a day, and they're very tuned in to what their peers are doing. Paid content and in-app purchases can contribute to FOMO, or Fear Of Missing Out. And you can see this being leveraged, particularly in social-media apps.

I have a couple of examples. Twitter Blue is a new subscription service where Twitter users can put their own tweets behind a paywall, and you have to pay, I think, about $3 a month to see those tweets. On TikTok, it's a little bit different. They're testing a feature where users can pay to promote their content on the main-app algorithm called the For You page so more users can see it, potentially.

And so if a teen isn't paying for these kinds of features, it can create the impression that they're missing out on popular or exclusive content. That can drive them to make that purchase when they might not have ordinarily. Next slide, please.

So now I'm going to talk about trackers. They've been mentioned a few times throughout today in other presentations. But you can think of them as the vehicles that collect information from the teen users and transport it to those third-party companies, like advertisers.

They're bits of software code that are integrated into mobile apps, and they can be first-party, which is owned by the app developer, or what's more common is third-party trackers, which are owned by other companies. They're also usually set up to collect a specific type of information. So for example, some trackers collect device identifiers, which are unique alphanumeric identifiers that can be used to track the same teen user across different services. So that is also called "cross-app" or "cross-device tracking."

Trackers are used to facilitate interest-based advertising most often. And again, they're used to build those user profiles by combining information from those different sources. The more trackers that are used, the more comprehensive a profile about a single teen user can be.

So our study also found that seven out of the 10 most common trackers and teen apps were owned by Facebook or Google who, in addition, each had more than one tracker each in an app. So this raised the question for us of why these companies want to collect so much information about teen users. In Google's case, as a search engine, they could use this information to personalize content and control what the teen user is seeing online and the type of information that they even have access to.

Because they're integrated into so many apps, as well, Facebook and Google, essentially, have the power to shape how this future generation is thinking by manipulating their online experience. And they can completely block or push other viewpoints based on that teen user's behavior and targeting content to them. Next slide, please.

So the last finding I'll talk about is dangerous permissions. And dangerous permissions are defined by Google as those that involve privacy or could affect the user's stored data. So these are provisions like location, microphone, camera, or accessing the contacts on a user's device.

So in the larger context, going with that same technology as trackers as the vehicles of information, you can think of permissions as the key that starts the car. And what I mean by that is permission requests get the app and any third parties that are integrated into it the ability to access information outside of it that they wouldn't normally be able to. So again, without that permission being granted, the tracker would not be able to collect the information about the user.

So permission requests could create potential misconceptions, especially with teen users who might not be thinking about the long-term ramifications of the data collection that occurs when they grant the permission outside of that instant that their granting it. So some of those misconceptions could be that if an app is asking for it, then it needs that information to function, it's necessary, or that asking for the permission-- it means that the app is being transparent about what it collects.

This isn't necessarily the case, unfortunately. Apps often asked for more information than is necessary to function. Going back to that key finding that there were 11 median permission requests in teen-directed apps, which is a very high number. It's unlikely that all of those apps needed that much information to just function and provide their services. This goes against the principle of privacy by design, which is only collecting information that you need. Next slide, please.

So why does this all matter? So our studies show that there is an unchecked ecosystem of data collection for a uniquely vulnerable teen audience. They see more ads than general audiences, they're exposed to more data collection via targeted ads and trackers, and there's a graded potential for them to be manipulated via in-app purchases or permission requests.

Neither existing or proposed policies adequately address the unique risks that teens are facing online as a result of their heavy engagement with digital platforms. We don't treat 16-year-olds the same way that we treat six-year-olds offline, so why are we treating them the same online right now? We need to approach conversations about teen privacy with more nuance. Next slide, please.

So thank you for your time. You can read the study at bbbprograms.org/risky-business-teen-privacy. There's a hyphen in-between each of those words. And if you have any questions, you can email us directly at tapp@bbbnp.org. Thank you.

MANMEET DHINDSA: Thanks, Cameryn. That was great. I want to ask you a few questions about some items that you mentioned during your presentation. So during your presentation, you raised a number of concerning practices that are more likely to occur in teen-directed apps. For example, you mentioned that these apps may include more requests for permissions to data, including those that you discussed as quote, unquote "dangerous permissions." Why do you think these apps and app developers are more likely to include these practices in teen-directed apps as opposed to general-audience apps.

CAMERYN GONNELLA: That's a great question. And actually, so our study wasn't set up to answer that question, but it is a great topic for future research. That being said, I do have a couple of ideas why this might be happening. The first is that the data ecosystem might have evolved to reflect the reality that teens are online more and that because they've spent much of their lives online, they're less likely to reject apps that have a lot of ads or things that have these types of practices. They're less likely to stop using an app just because it runs ads, for example.

And then also, we didn't look at the content of the ads themselves, but then there could be something about the content of ads or the way that they're being shown to teen users that could shed a little bit more insight as to why they're seeing more of them. For example, a lot of ads today are designed to look like content. So on Instagram or Twitter, or they just look like regular posts from other users, so it's less disruptive to teens, potentially.

MANMEET DHINDSA: That's really helpful to get a better understanding of that. And one other things I wanted to ask you-- and it's something that you and I had previously discussed-- was that, you know, I think it's your position that self-regulation may be the most effective way to handle online privacy routine. And I'm hoping to provide a little bit more detail about why you think self-regulation is the way to go here over, for example, to implement policy, especially when, as you just mentioned, some in the industry may arguably be benefiting from the lack of protection around teens right now.

CAMERYN GONNELLA: Yeah, absolutely. And that's a great question. And I did mention that this paper was done as kind of a first step towards a potentially self-regulatory program, the Teenage Privacy Program that we're developing. And this is because at the national programs [? of note, ?] we've known that many industry members want to do the right thing, but the current regulation and even self-regulatory systems, focus on younger children, so for example, the COPPA Safe Harbors, like Mohammed mentioned.

There's also been a lack of focus and consensus on how and even whether to act regarding teen privacy because it's such a nebulous space right now. And, again, the complexity of the digital ecosystem might make it really difficult for all the different players-- so all those different companies that own the trackers, the different app developers, et cetera. It's going to be hard to create a comprehensive and transparent system that they all abide by.

And then we also have the opinion that the type of industry self-regulation that would best serve consumers in the market would just make sure that there is transparency regarding whatever principles and policies are in place. And then ideally, there's an accountability element through a reliable third party to make sure that that transparency and accountability for compliance is abided by.

And again, this is why National Programs is looking at the teen space through our teenage-privacy program. The study was our first effort, and we're committed to continuing to look at issues surrounding teen privacy and working with responsible companies to find solutions to this.

Thank you so much, Cameryn. As we're nearing the end of the panel, I want to bring you both together for a joint discussion. So you both have raised really important issues related to children's and teen privacy. And I love to hear from both on whether you have any suggestions on what policymakers can do to address some of the issues that you raised by your research.

Mohammad, I know that you mentioned in your presentation that you believe that-- I think the term used was "stringent regulations were necessary." So would you like to start off there?

MOHAMMAD MANNAN: Yes, thanks, Manmeet. So in our research group, actually, we check in on these kind of security products, not only for children but for other stuff, as well. I mean, some companies are quite proactive. So if we report something, they will be quickly-- they will go after it quickly and fix it. But most of them, they don't bother to even respond. It's really difficult to reach them.

So I think this is happening because there is really no regulatory pressure on them. It's like if they fix something, it is voluntary. If they want to do it, it may look good on them, and that's why they might do it. So I think there must be regulations for accountability.

And this is somewhat happening with GDPR now with data breaches, but I would suggest instead of waiting for the data breach to happen, there should be regulations in the data-collection phase so that only a minimum amount of data is collected from users. And I was quite happy to hear that this approach was highlighted in the opening remarks by the FDC commissioner this morning.

Now these companies, in my experience, I see that they can collect whatever they want. And most of the time, it is way more than what is necessary for their functionality. And they don't even bother to protect the sensitive data that they collect.

So as researchers, we see-- we can highlight some of these problems. We can design this framework to detect them, which may help regulators. But at the end of the day, there is no strict regulations. Privacy can go on for 10 more years, then we'll be discussing very similar issues.

CAMERYN GONNELLA: Yeah. I agree with a lot of points that you just made, Muhammad. And, obviously there are some policymakers that have taken measures in certain states on teen pregnancy specifically. So for example, the CCPA, the California Consumer Privacy Act, requires that 13 to 16-year-olds opt in to the sale of their personal information.

We just don't think it's super likely that a 13, or 16-year-old is going to understand what it means when their personal information is sold versus not consenting to it being sold, especially when businesses and policy makers are still largely debating when a sale occurs online or what even constitutes a digital sale. But to your point, it would be great to see some sort of educational component, as well, to ensure that teens know what the consequences and the impact of consenting versus not consenting would be online.

And while I agree with you, this is why we advocate for self-regulation as a potential solution to this, one is a matter of practicality because none of the bills that I mentioned in my presentation have passed yet. And so self-regulation can be a more expedient way to effect change and implement new standards. But one way that self-regulation could also be particularly useful is establishing best practices for the industry to follow. And that's where accountability becomes so important. Like you mentioned, there has to be some kind of check on the industry to make sure they're actually following what they say that they're doing.

And then by establishing best practices, policy makers could then potentially include those best practices in future policy to further encourage adoption by industry. This is the sort of model that developed around the COPPA Safe Harbor. And I know you mentioned that in your paper, as well. We also operate a safe harbor, and we've been really proud to work hand-in-hand with the FTC in the children's space to help encourage businesses to adopt the well-established privacy practices for younger children. And so we think that's worked fairly well.

MANMEET DHINDSA: Thank you to you both. That's really helpful. And we only have a few minutes left, but I found both of your papers to be really interesting and exciting work. And I would love to hear more about any other research that you have planned for the future. Cameryn, do want to start with that one

CAMERYN GONNELLA: Yeah, so as I said, this was kind of done as a kickstart to our Teenage Privacy Program. So obviously, a lot of the questions that you asked today will probably be used as a jumping-off point for future research to see exactly what best practices would be viable and effective in the teen space. And regardless of what those practices [? might ?] be, it's going to be really important to balance parental authority and then the autonomy of teenagers, as well, when determining best practices for privacy online.

MANMEET DHINDSA: Mohammad, [INAUDIBLE]?

MOHAMMAD MANNAN: Yeah. I just mentioned earlier that, generally, were are interested in-- in our group, we go after the products that claim to provide additional security or privacy. So at least this product should come with some security and privacy in themselves because they claim to do-- to provide these kind of features to users. These are the kind of products that we target, whether they are for enterprise use or for a vulnerable-user group, like children.

So in the future, we want to continue in this direction, maybe target other vulnerable-user groups or some fancy enterprise-targeting products that may claim, OK, we can solve all the problems, like solar wind, or similar issues like that, and ransomware, and all these new attacks that we are seeing every day. And there are actually products out there that will claim, like, look, we are using AI and all sorts of fancy things in the background. We can really make you secure. So we try to target these types of products in our analysis.

MANMEET DHINDSA: Great. Well, thank you both very much for a very interesting panel. I know I personally learned a lot. So thank you both for your presentation and your work in this field, and we look forward to seeing your work from both of you. And with that, I will turn it over to our next panel.

[? MOHAMMAD MANNAN: ?] OK.

[MUSIC PLAYING]

MODERATOR: All right, Christina, are you starting the panel?

CHRISTINA YEUNG: Sorry about that. Hi, everyone. Thanks for joining us today for the last panel of the day, Privacy in the Pandemic. The pandemic has had far-reaching effects. It's one of the reasons we're doing PrivacyCon virtually again this year. This panel will explore how the epidemic has changed people's online experience, from phishing attacks, to learning more about what people think of the individual social-media platforms designed to prevent the spread of misinformation.

One quick note-- we'll be doing the Q&A session at the end after both panelists have presented. Please send in your questions to privacycon@ftc.gov or by using the hashtag #privacycon21 on Twitter. I'm looking forward to a fun discussion.

With that said, I'm happy to introduce our first presenter, Marzieh Bitaab. Marzieh is a second-year PhD student at the Arizona State University, and she's here to talk about the [INAUDIBLE] titled, "Scam Pandemic-- How Attackers Exploit Public Fear Through Phishing." Marzieh, take it away.

MARZIEH BITAAB: Hello, everyone. My name is Marzieh Bitaab, and I'm going to talk about how phishing trends changed during the pandemic and how attackers exploit the situation. Next slide, please.

Let's just talk a little bit of background. A phishing attach is one of the most prevalent web-based trend and have caused substantial damage to victims. If a victim is successfully fooled by attackers, and he or she will open a website and submit their sensitive information, such as credit-card numbers or account credentials.

And browser-based phishing detection plays an important role because of their scale and the fact that they're embedded in their browsers. So people heavily rely on these technical mitigations. However, several research works revealed the limitations of these black list anti-phishing defenses, and they all offer a significantly faster approach to protect users effectively. These works all imply that as long as the standard anti-phishing defense is operated in a reactive manner, phishing will remain a significant threat to users. Next slide, please.

The pandemic has changed daily life across the globe from different aspects. First of all, the widespread lockdowns, travel restrictions, and work-from-home arrangements have increased users' reliance on online services. And second, the continuous updates from different sources, such as social media, have caused panic among people. And third, people desire to help each other, especially who are affected, during the global disasters. Unfortunately, this increased usage of the internet with the unstable, emotional [INAUDIBLE] user has left them vulnerable to social-engineering attacks, such as phishing and scams, more than ever. Next slide, please.

And so let's see some of the examples of this phishing that we are talking about and people are encountering. [INAUDIBLE] people [INAUDIBLE] a huge variety of emails that impersonate authorities, such as WHO or CDC, and ask them to donate to bogus causes. And next, please.

For example, this one is an email that appeared to be sent from CDC and ask people to donate Bitcoin to fund COVID-19 vaccine research. The other example looks like Microsoft Outlook interface. It asks them for a username and password to show information about new cases of infection around your city. Next is the example that we see different spam campaigns using face masks or gloves. These are just a few examples, and there are much more out there. Next slide, please.

So to identify these trends and to understand how these phishing trends change, we collected a different variety of datasets. We collected news articles and government announcements about phishing and scams related to the pandemic. We also collected corona-related discussions from two large underground forums to understand how cybercriminals' activity changed during the pandemic.

Then we gathered DNS records, issued TLS certificates, and reported phishing websites to see how the pandemic affected internet infrastructure. And then to understand what kind of content has been used, we crawled the source code of these malicious websites. Finally, we collaborated with a major financial-services organization and used the specialized network monitor to analyze the victims traffic to phishing websites and volume of reported phishing emails.

Unfortunately, due to time limitation, I'm not going to talk about the first two datasets in this presentation. And our observation period from this study is the first four months of 2020. Next slide, please.

In the first stage of this study, we observed that the number of news related to COVID-19 had rapid growth. So it motivated us to further investigate this matter. And we looked at the registration patterns.

We collected DNS records from different sources-- Domaintools, RiskIQ, and Whois domain search-- to find corona-related domain names. Then to understand the activity, like to see if two of them are actually being used or they've been registered to be used for other purposes or even that's used at all, we looked at the certificates. We collected [? over ?] 44 million TLS certificates using Google Rocketeer CT log. Next, please.

Here, the figure on the top shows the number of newly registered corona-related domain names started increasing from March. And in January 13 is the time that WHO declared a global-health emergency. And March 11 is the time that WHO declared the outbreak of COVID-19 a pandemic. As the number of newly registered corona-related names increased, it'd be harder for a users to distinguish between realted [INAUDIBLE] websites and malicious ones. The figure on the bottom of the number of certificates issued to corona-related domain name-- increased from February and peaks in March. Next, please.

OK, so far we've been-- we saw the registration pattern and certificate pattern. Then, for the next step, we'll look at the websites, phishing websites, that are being reported. the number of host names that were reported to two major clearinghouses of URLs-- APWG and OPENFUSE-- did not increase significantly during COVID-19. However, the count of host names itself failed to accurately reflect the damage caused by COVID-19, as some certain high-impact websites received more traffic because of their increased spamming activity or their ability to evade defenses.

So to deepen our insight into phishing trends, we used a recently proposed network-monitoring approach. I'm going to talk about this network-monitoring approach in high level. Phishing of websites often embed a tracking code or images that are hosted on external servers. For example, a phishing website that impersonates Microsoft-- if you look at the request, we can see that, actually, three of them goes through the server of the organization that's being impersonated.

If we're able to track these web events, we can track the traffic of victims to phishing websites. However, to have such data, one needs to be the organization that's actually targeted by attackers. So we collaborated with such organizations and analyzed two additional datasets-- traffic to websites by targeted victims and phishing emails reported to the organization. Next slide, please.

As shown in the figure, the network monitor recorded a sudden increase in the number of victims in March. And this number stayed elevated in April. Overall, in March and April, we can see 2.1 and 1.6 times more victims than at January. And January is the time that the organizations usually see elevated activity because of the holiday shopping season. And also, the number of emails reported to the organization validate this increased activity. Next slide, please.

So we see that the number of phishing websites reported to blacklists did not increase significantly. On the other hand, we have the record-breaking number of victims. The question that arises here-- what was the content that attracted so many victims? To answer this question, we crawled the source code of scamming and phishing websites and grouped them to four main categories.

The first category is donation-themed, where the victim would think they're making a small donation. However, instead, the attacker would steal their information. The other category is PPE sales. Personal Protective Equipment was in high demand in the early months. And it was in short supply in major retailers. So attackers exploited this high demand and created fake websites that sell fake PPE. Next slide, please.

Attackers also exploit events that are related to corona. For example, to help address the hardship, financial hardship, the US government offered stimulus payments. However, a different group of people received this payment at a different time. When those who received their check shared about it on social media, others started to worry if and when they were going to receive their payment. So phishers were quick to disguise themselves as IRS and steal people's personal information.

And the final category would be shopping websites. Fraudulent shopping websites tried to keep up with the look and feel of a legitimate website. As more and more legitimate organizations include COVID-19-related information in their website, such as statistics, or new policies, fraudulent websites also include such information. Next slide, please.

I'm going to wrap this presentation with our key takeaways. Even though, in our dataset, the number of phishing attack leveraging the pandemic seems to be negligible, there were still a record-breaking number of victims. We believe this is because the attackers exploited pandemic-- people's pandemic-related wants and needs with high-quality websites.

Also, we observed that the number of news related to the pandemic increased in March. However, there were still many victims. This implies that entire phishing systems [INAUDIBLE] [? mitigation ?] is not enough to protect users. Next, please.

We collected 467,000 domain names related to COVID and 17,000 certificates issued to correlated domain names. And less than 1% of these domains were reported to blacklists. And once you write that list, [INAUDIBLE] less then thousands of them are benign. On the other hand, we have the FTC report that shows people lost $40 million to COVID-19-related fraud from January to May. This implies that phishing is just one type of corona-related attack.

And there are other types, other [? scamming ?] sites that cybercriminals are [? then ?] exploiting the ecosytem's lack of defense against other scam types. Next, please. Thank you for listening to my talk.

CHRISTINA YEUNG: Thank you very much for your interesting presentation, Marzieh. Next, we have Christine Geeng. Christine is a PhD candidate at the University of Washington. They are here to talk about their paper titled, "Social Media COVID-19 Misinformation Interventions Viewed Positively, But Have Limited Impact." Christine, take it away.

CHRISTINE GEENG: Thanks for the intro. So hi, everyone. I'm Christine. And I'll be presenting work that I've done with Tiona Francisco, Jevin West, and Franziska Roesner. Next slide.

And just to give an overview of the project that I'll be presenting, we conducted a mixed-method survey on social-media users' attitudes towards false-information labels. And this was conducted in March, 2020-- so over a year ago, during the beginning of the pandemic. Next slide.

So it's probably not news to anyone here that coronavirus-related misinformation has become a major problem on social media. To give just one example, when the Plandemic COVID-conspiracy video first popped up on Facebook, according to this graph, in only a couple of weeks after that first posting, it amassed millions more interactions than other popular videos that were uploaded to Facebook around the similar time. Next slide.

So given the severity of this issue, platforms have taken steps to ramp up their misinformation action, for example, increasing fact checking and labeling posts. So in this example, this is a screenshot of Facebook's false information label, which they overlay in certain posts. Next slide.

And platforms have also started partnering with health organizations, like the CDC and the WHO. So in this example, if you go to Instagram and you search for anything related to the coronavirus, you'll see this generic banner, which basically says that users can go to the CDC for trusted health information. Next slide.

Of course, with all these new labels and interventions that platforms have, this raises the question of how effective they are in combating this information. So prior research done on Facebook has shown that having a related stories fact checker label, as you can see in this image, can significantly correct misinformation. Of course, these are the labels that existed on Facebook back in 2017, and they've changed the kinds of interventions they've used since then. Next slide.

Of course, it's not also-- it's not only the platforms that are collecting this information online. Prior research on Twitter is also showing that users sometimes take steps to correct rumors. So this diagram presents the mental model of a Twitter user deciding whether or not they're going to correct something they've previously tweeted that turned out to be false. Next slide.

Well, clearly, the circumstances of COVID-19 and COVID-19 misinformation is novel, so we felt compelled to study these interventions within the specific context. So that led to our research questions-- and number 1-- what are peoples' attitudes towards social-media interventions for COVID-19 misinformation, including generic banners linking to authoritative sources and specific post labels? Next slide.

And 2, how do people discover that COVID-19 misinformation is false? Specifically, what was the role of platform interventions in this discovery compared to other methods? Next slide.

In order to answer these questions, we conducted a paid mixed-methods online survey in March, 2020-- so during the beginning of the pandemic. And we also felt compelled to collect data quickly because this was such a novel scenario. We recruited through our personal networks, [INAUDIBLE] and our study was deemed exempt by the UW IRB. Next slide.

So the first part of our survey consisted of open-ended questions around what participants thought about Facebook, Twitter and Instagram misinformation interventions. We also asked them to rate these on a five-point scale for how helpful it was. Next slide.

So to show what these generic web banners look like, here are the banners from Instagram, Twitter, and Facebook, which all just say, you can find trusted health information if you go to the CDC or to an external site. Next slide. And these were the post-specific interventions that we asked participants about. So here, we have one from Twitter, which has the label of manipulated media, as well as the Facebook one that I showed earlier. Next slide.

In the second part of the survey, we asked participants for their prior experiences of seeing COVID-19 misinformation, and we asked them how they discovered it was false and what they did when they realized this. Next slide. So here, I'll present the results from our open-ended question of how people felt about these banners. And this chart shows how we qualitatively coded these results.

As you can see, most of the respondents had positive reactions, while many stated a neutral response. And many also stated they felt these banners were necessary because they were already informed about a certain health information. On the other end of the spectrum, some participants expressed anger, as well as worry, that having these interventions could be abused to censor information, in their words. Next slide.

We also found that participants rated post-specific interventions more helpful than generic ones significantly for Facebook and not significantly for Twitter. Next slide. We also asked participants the question of, what did you do when you realized COVID-19 information someone else shared was false? 55% said they did nothing, 18% they corrected the person publicly, and 17% said they corrected the person privately. Next slide.

For the very similar but slightly different question of, what did you do when you realized COVID-19 information that you believed was false-- 57% of participants said they did nothing, 27% shared the correction, and 2% said that they unshared it if they had previously shared the post. Next slide.

So what can we take away from these results? Well, first, social-media platforms should increase specific misinformation-labeling efforts. Next slide. This is because we found that we had gained mostly positive responses from participants about these interventions. And participants also found the specific interventions to be more helpful than the generic banners. Next slide.

We also want to point out that participants, not just social-media platforms, correct misinformation in various ways. Next slide. In their open-ended responses, participants discussed correcting each other in a group chat, or adding comments with corrections to posts, or liking existing corrections. And some participants mentioned reporting the misinformation posts or filtering unwanted content through their feed. Next slide.

As with any study, our survey came with certain limitations. Our participant pool was a convenience sample and not representative of the US population. Second-- platform interventions and COVID-related news is rapidly changing. And finally, these interventions that we survey people on, and their responses, came from March, 2026-- 2020. So it's been over a year since then, and a lot has changed on both platforms in terms of the kinds of COVID conspiracies that are going around and also, in general, people's relationships with COVID and COVID information-- next slide-- which leads us to our feature-work questions.

So first, how do users respond to new platform interventions. Next slide. And just to give you an example, in October, 2020, Twitter rolled out new misinformation labels during the US presidential election. So in this new intervention, they label the tweet as containing misleading information but state that it's in the public's interest for the tweet to remain accessible. So users can still click View and will still be able to access the information. Next slide.

Facebook has also introduced a new generic banner, which basically takes the user-- the user's to more information on vaccine infos. They've basically labeled every single news article that has the word "vaccine" in it with this generic banner. Next slide.

However, it's still unclear how effective having these vaccine-information labels are. As one news article pointed out, Facebook knows that adding labels to Trump's false claims does little to stop their spread. And basically, what the article talked about was that even if Facebook labeled Trump's false claims, many people still reshared that information, which meant that COVID misinformation was still spreading on the platform.

And this raises the interesting question of when it is more effective just to just outright delete COVID misinformation and conspiracy posts versus labeling them. Next slide. And second-- or the last question for future work is, how can platforms make it easier for users to share misinformation corrections? Next slide. Research on that question can build on the prior work of [INAUDIBLE], who found best practices for correction on social media, which includes correcting early, repeating corrections, offer an alternative explanation, and include a credible source. Next slide.

So just to wrap things up, I'll repeat the takeaways of this presentation. So first, post-specific interventions are more effective than generic banners. Second, when participants came across misinformation and realized it was false about one third-- realized it was false, about one third made a correction.

And finally, as I've shown, platform designs, COVID-19 misinformation, and cultural contexts change really fast, which means that this kind of research is probably needed over a sustained amount of time, which is often very difficult, for-- particularly for external researchers who don't have the same sort of access to data as people who work at Twitter or Facebook necessarily do. And in order to better hold these platforms, these social-media platforms, accountable, nonaffiliated researchers should be able to audit their data with open-access academic research.

Thanks. And this work was possible through the Paul G. Allen School of Security and Privacy Lab and through the Center for an Informed Public. Thanks for listening.

CHRISTINA YEUNG: Thank you both for some really interesting presentations today. And I'll just open it up to the Q&A starting with Marzieh. Marzieh, when you were collecting these newly registered domains, were you able to observe where the domains were being registered? Or were most of them registered in the US?

MARZIEH BITAAB: Yeah. If you look at the country that the domains were registered at, we can see that, actually, 44% of them were registered in the US, followed by Germany that has less than, like, around 4% of them. So almost half of them were registered in the US, which is alarming.

CHRISTINA YEUNG: For sure. You mentioned that for some of these domains, some of them were-- or most of them were actually not directly related to phishing but rather, potentially, fraudulent shopping websites. Are the ways for consumers to know if they're on these types of sites? And what types of advice would you give to people who want to make sure they're actually on legitimate sites?

MARZIEH BITAAB: So, actually, of course, there are some things that users can do. Education is really important about detecting both phishing and scam websites. It's not like if we have a good defense mechanism, we don't need people to know about these websites.

So one of them, the easiest and most important thing that anyone can do is to, like, before they want to do any purchase or enter any sensitive information, they can look up their presence, the online presence of the website, and look for other people's experience or reviews. And I think this is the most basic [INAUDIBLE] step that anyone should do.

CHRISTINA YEUNG: And turning to Christine, are there a clear ways social medias can improve labeling false information? During your presentation, you said that some of these social-media platforms acknowledge that the labels aren't effective. Have you thought of better ways to prevent misinformation from being spread?

CHRISTINE GEENG: That's a really good question and, also, probably a hard one to answer just because, even amongst our participants, it seemed like there were varied responses on how people reacted, mostly positive, of course, which is good. But there were some people who mentioned that they didn't necessarily trust the platform itself, like Facebook. So therefore, they didn't trust whatever they have to say about the coronavirus.

My sort of impression, though-- that it's better that they're doing something rather than nothing. But I think, any sort of future work in this area should look at is the efficacy of like outright deleting coronavirus misinformation versus just labeling it but still allowing it to be reshared, retweeted, liked, the gamut, yeah.

CHRISTINA YEUNG: Kind of building on that, taking a step back from removing posts, let's say that I, as a regular user, encounter-- or I see a post that, I think, contains false information. Do any of these platforms allow to flag-- allow users to flag posts containing false information? Or does it just fall under general abusive flags that the platforms already have?

CHRISTINE GEENG: Yeah. So Twitter definitely has a specific misinformation, like, reporting label. I think Facebook does as well, but it might be a little more broad. Like, Twitter allows you to, like, categorize it by, like, politics, et cetera. I would, like, double-check. But they do allow reporting, even if it's unclear, like, when or how that gets used to the user.

CHRISTINA YEUNG: That makes sense. Turning it back to Marzieh-- so if you have to focus on phishing prevention and preventing people from actually getting to these websites, what do you think you would prioritize?

MARZIEH BITAAB: So it is important to remember that attackers always remain several steps ahead of modern anti-phishing defenses, and they take advantage of the global disasters, like, to harm users. So it's important to solve this problem collaboratively, to collaboratively deploy anti-phishing systems that can change according to the changes in the ecosystem and maybe go further and have a proactive defense mechanism instead of reactive, like this is the most important thing. I cannot emphasize this enough, that having a proactive defense mechanism would help to prevent phishing and protect users more effectively.

CHRISTINA YEUNG: And just building on that, do you think [? CAs ?] should be more proactive, as you were talking about, in revoking the certifications after discovering websites are used for phishing?

MARZIEH BITAAB: So the thing is, as I mentioned, it's something that can be done in different steps. That's why I mentioned the [INAUDIBLE]. They can prevent it both at the time of registration if the registrars could be more conservative or cautious during the registration time and also, something that the [? CAs ?] can focus on, even before and after they've been reported. So yeah, I think that's something that should be done collaboratively with-- in different steps, to have an [? acceptable ?] result.

CHRISTINA YEUNG: All right. And just as an overall wrap-up question for the both of you, do you have plans for doing this work? And where do you see this going in the future? Christine, I'll start with you.

CHRISTINE GEENG: Yeah. That's a good question. What I'm curious about is there was like a recent Pew report that talked about how you can categorize different information-- consumers into different categories, like people who are skeptical versus people who are not. And I'd be really interested in seeing how that relates to the efficacy of seeing a platform label something as false. Yeah.

CHRISTINA YEUNG: That would be really interesting. I'd be curious to find out more, as well. Marzieh, I know you were talking a little bit about looking more into these fraudulent sites. Is that the direction you think future work will go? Or do you have other plans?

MARZIEH BITAAB: So for the future work, we are thinking about having a new ecosystem defense. We know that one of the important steps-- next step would be to identify other types of scam websites, not only focus on phishing websites, and try to detect them and protect users from [INAUDIBLE] scammer site. That will be the future direction for us.

CHRISTINA YEUNG: That also sounds really fascinating. All right, just to wrap up, I'd like to say thank you very much, both of you, for presenting at PrivacyCon today. Great job, everyone, and thank you very much. I'll pass this over to Lerone for the closing comments for PrivacyCon.

[MUSIC PLAYING]

LERONE BANKS: Today, we've heard from researchers about privacy and fairness risks, and machine learning, and practical methods that organizations can use to audit their algorithms. Researchers also presented their work on analyzing alternatives to communicating policy policies and quantifying consumer tracking in ads. This year's event included panels on timely privacy issues within the context of IoT, teens, and the pandemic. It has been a day full of research that can support data-driven decision-making and executing the FTCs mission.

I would like to express gratitude to Jamie Hine for his relentless drive to bring this year's event to life for the sixth year and reiterate his thanks from this morning to the supporting team, all of the moderators, and the researchers for their informative and compelling work. I look forward to continuing to expand the role of research in the agency's evolving efforts, as laid out by Commissioner Slaughter and [INAUDIBLE]. Thank you for attending PrivacyCon 2021, and see you next year.