FTC Cross-Device Tracking Workshop
November 16, 2015
Segment 1
Transcript

MEGAN COX: Good morning. I think we'll go ahead and get started now. Thanks for your patience this morning. On behalf of my colleagues here at the Federal Trade Commission, I'm happy to welcome you to our workshop on cross device tracking.

My name is Megan Cox, and I'm a staff attorney in the Division of Privacy and Identity Protection here at the commission. Before we get started here at the program this morning, I need to review just a few administrative details.

Please silence any mobile devices you have. If you must use them during the workshop, please be considerate of your fellow audience members and the panelists. Please be aware that if you leave the Constitution Center building for any reason, you'll have to go through security to get back through. So please bear this in mind, especially if you're on a panel this morning, so we can do our best to remain on schedule.

The restrooms are located just down the hall outside the auditorium, and the Plaza East cafeteria is located inside the building, so you can go to it without having to go through security again. It is open now until 11:00 AM and then from 11:30 to 2:30.

And most of you received lanyards this morning with FTC event, and we do our best to reuse these from event to event, so if you could please turn it into staff on your way out, we would really appreciate it.

If an emergency occurs this morning that requires we leave the conference center, we ask that you follow the directions that you will hear over the PA system. And if an emergency occurs that requires us to evacuate this morning, the building-- an alarm will sound.

And we ask that you evacuate through the 7th Street exit. After leaving the building, you'll turn left and onto 7th street and across East Street to the FTC emergency assembly area where we can wait until we are cleared to come back to the building. And if you notice any suspicious activity, please alert security.

Please be advised that this event this morning is being webcast, and you might be photographed or recorded. So by participating in this event, you're agreeing that your image or anything you say or submit may be posted indefinitely at ftc.gov or one of the commission's social media sites.

And we are happy to welcome those watching via webcast. We'll make the webcast available along with all of the agenda and bios and other materials on our website to create a lasting record of the event.

For those of you on Twitter, FTC staff will be live tweeting today's workshop with #tag ftcxdt. And we've comment cards available here in the conference room, so audience members will be able to submit questions to the panelists. And you can turn them into our aids here, Carry Davis and Kristen Birch over here, if you have questions you want to submit to the moderators.

And as a reminder, the public comment period for the workshop is open through December 16, 2015, so we urge all those here in attendance today and thinking about these issues to submit public comments by then to our website. And lastly, I want to thank our panelists for taking part today. We are grateful for your time and careful consideration of these evolving technologies and consumer issues

And aside from the folks that you'll see on stage here today, this program would not be possible without the great work done by Peter McGee, Jared Ho, Aaron Alva, Tina Young, Phoebe Ruge, Matt Smith, Carry Davis, Cheryl Thomas, Carry Kahlua, Fawn Buchard, Crystal Peters, and Bruce Jennings, alongside our paralegal support today from Bianca, Spencer, Shea, and Carly.

And without further ado, it's my pleasure to welcome our chairwoman, Edith Ramirez.

EDITH RAMIREZ: Thank you, Megan. And good morning, everyone. Welcome to the FTC's workshop on cross device tracking. In the early 2000s, online tracking was used primarily to determine how many individuals saw or clicked on ads. But things have changed a lot since then.

Today, tracking is a much more sophisticated and complex enterprise, involving a multiplicity of players as companies seek to link consumer behavior across a wide range of connected devices. Now, to be sure, cross device tracking provides significant benefits to consumers.

I can start reading a book on my tablet at home and then pull it up on my smartphone on the exact page I had left off if I find myself stuck in a long line somewhere. Or I can search for a pair of shoes on my notebook computer at home and then take advantage of a promotion using my smartphone when I'm at the store.

Cross device tracking can also help companies implement fraud prevention programs as they learn which devices typically access consumer accounts. But some consumers are simply not comfortable with their browsing behavior on one device informing the ads that they see on another device and can quickly see how our privacy might be invaded as the lines between work and home and other formally distinct parts of our lives increasingly become blurred.

Today, for example, someone who searches online about a medical condition in the privacy of her home could very well see advertisements the next day at work related to that condition or the next evening on the families smart TV.

Many of the questions raised by cross device tracking techniques are familiar. Are the new methods operating in a way that is transparent to consumers? Can consumers be given effective opt-out choices? And if not, what can be done to provide consumers with more control?

But certain attributes of cross device tracking raise additional questions and challenges. For instance, some of the techniques that company's employee to collect data passively across devices mean that online tracking is even more hidden from the typical consumer.

And as data about consumers is compiled and shared by an increasing number of companies in the tracking ecosystem, advertising firms, exchanges, and onboarding companies, to name just a few, the number of entities who have access to online information collected about consumers continues to grow.

These are some of the issues that are speakers and panelists will be addressing. Our aim is to learn more about the techniques that are being used today and those that we might see tomorrow to track consumers and also to discuss how we and the various players in the tracking ecosystem can address these issues.

Now, to set the stage for the discussion, I'd like to begin with a little bit of background on the evolution of tracking. And then I'd like to highlight a few more of the questions about the privacy implications of the new cross tracking-- cross device tracking techniques that companies are now using or are likely to be using soon.

Let me start by taking you back to 2009 when the FTC issued its report on online behavioral advertising. At that time, the FTC was mainly concerned with tracking across websites through cookies on a single computer. We later saw new forms of tracking emerge, first flash cookies and then browser history sniffing and device fingerprinting, among others.

Last year, in our data broker report, we talked about the practice of onboarding, where companies combine offline and online data to create even richer consumer profiles. And now, we're no longer talking about tracking across a single browser on a single computer.

Companies aim to follow consumers across all of their connected devices, smartphones, tablets, desktop, and notebook computers, connected TVs, and even smart watches, and other wearables.

The web of linked devices also referred to as a device graph enables companies to know the multiple devices are connected to the same person. With this information, advertising can be targeted to a specific consumer across his or her devices and consumers actions or purchases can be attributed across their interactions.

As I noted earlier, I may now use my smartphone to take advantage of a shoe discount based on an ad that I saw at home. Cross device tracking and linking tells advertisers that I'm the same person who saw the ad and made the purchase.

As the number of devices we use to grows, so does the degree of linking that occurs. To do this, companies rely on a mix of what are known as deterministic and probabilistic techniques.

Deterministic linking is based on information a consumer provides to a website or service, such as when they log on to a social network or email account. Probabilistic models work more passively by making inferences based on information the user has no control over, such as shared IP addresses or location information, when two devices are consistently used together in the same household.

This approach is an example of some of the big data techniques we highlighted last year in our big data workshop. Both methods have proven to be effective tools in tracking everything from the types of articles you read to the types of products you buy. The experts that we've assembled today will provide additional details about the techniques company's employee to track consumers across devices.

So what privacy challenges are raised by these new and more sophisticated tracking techniques? First and foremost, they raise important questions about transparency While tracking itself is not new, the ways in which data is collected, compiled, stored, and analyzed certainly is.

We know, for instance, that behavioral techniques based on passively collected data are used to infer whether an individual might be in a particular target demographic or be interested in particular products.

Now, many of these same techniques are also being used to try to infer the actual identity of that individual and what other devices she owns. This more extensive tracking allows companies to connect more and more of consumers offline activities with their online behavior. This results in more detailed and more personalized consumer profiles that are assembled, traded, and shared by a growing number of entities in the data ecosystem.

They do this under the veil of anonymous identifiers and hashed PII. But these identifiers are still persistent and can provide a strong link to the same individual online and offline. Not only can these profiles be used across sensitive inferences about consumers, there's also a risk of unexpected and unwelcome use of data generated from cross device tracking.

Certain data could be misused by unauthorized third parties in a way that affects consumers access to loans, jobs, or educational opportunities. In addition, information is now stored in large volumes for longer periods of time thereby increasing the risk that it could be used for unexpected purposes or left vulnerable to security breaches.

Second, these concerns are exacerbated by consumers lack of awareness of and choices about tracking. As it currently stands, there are almost no tools that allow individuals to know what devices are linked together by tracking companies or specifically linked to them.

Furthermore, while certain tools to opt-out exist, most controls do not allow opting out of the underlying data collection and linking of identifiers. They only allow opt-outs of targeted advertising.

How can companies engaged in cross device tracking provide consumers with choices that will be honored? How will we assure that opt-outs are effective and do not conflict with other existing controls, like browser controls? And will industry provide an effective mechanism that allows consumers to exercise one opt-out for all devices?

Understanding how different technologies work and their limitations is necessary to have a meaningful discussion about the notice and choice options consumers can and should have available to them in this arena. Finally, I should note that, in addition to notice and choice, we need more discussion about the role that data minimization as well as security and accountability play in addressing privacy concerns raised by marketing through cross device tracking.

For example, a device graph that probabilistically links devices as being likely related to tracking may quickly grow stale and may not be necessary to retain for long time periods. We should consider how these important principles can be applied in the context of cross device tracking to mitigate privacy concerns.

For today's workshop, we've gathered industry representatives, academics, technologists, and consumer advocates to discuss these and other important implications of cross device tracking as well as potential solutions. As we consider what additional steps might be necessary, the FTC will continue to monitor the marketplace and take action as needed to protect consumers.

For example, in our 2011 case against online advertiser ScanScout, we alleged the company had deceptively claimed consumers could opt-out of tracking by changing their computers web browser settings to block cookies. We alleged that the company actually tracked consumers through flash cookies, which consumers could not control using browser settings.

In another case, we alleged that Epic Marketplace make promises to consumers about the limited nature of its tracking when, in fact, it used history sniffing technology to track consumers across the internet. And in our case against a national company that franchised Aaron's rent to own stores, we alleged that it engaged in an unfair practice when it didn't tell consumers it could tract their activities through web cams attached to their leased computers.

No matter what the technology, we are committed to ensuring that companies are truthful and refrain from engaging in deceptive or unfair conduct when it comes to tracking consumers. We've also encouraged industry to adopt best practices and are pleased to see that many companies are working to differentiate themselves in the marketplace by being more privacy protective.

Products touting privacy features are on the rise, and tools empowering consumers to protect their privacy, such as ad blockers, are growing in prevalence and popularity. The Digital Advertising Alliance and Network Advertising Initiative have also taken steps to enhance privacy protections in the online advertising space. These are organizations' self-regulatory principles encourage members to provide increased transparency and offer consumers control over data collection for certain practices among other things.

DAA and NAI have also developed useful opt-out tools for online data collection covered by their self-regulatory codes. For instance, NAI has also issued guidance relating to the use of non-cookie technologies, emphasizing that members should honor user opt-outs regardless of the technology that's used. NAI as also currently developing and testing a new centralized opt-out tool that will inform consumers when NAI members to use non-cookie technologies for interest-based advertising.

DAA, for its part, recently announced that it's beginning enforcement of its principles in the mobile environment. It also just launched an updated version of its mobile app for Spanish speakers. The app aims to provide an easy to use interface for consumers to set their preferences for data collection and use across apps for certain advertising and uses.

These are all steps in a positive direction. As tracking becomes more sophisticated, it's imperative that companies throughout the tracking ecosystem rise to the challenge of fostering technological solutions to inform consumers, offer choices, and honor those choices.

We recognize that we won't be able to identify all of the potential challenges and solutions today. But this workshop is an important step forward in helping us understand how evolving tracking technologies are working so that we can move toward better safeguards and effective choices for consumers.

Working through these novel and complex issues together today and in continuing conversations will help ensure that consumers' privacy interests are protected while allowing for continued innovation in the digital marketplace. ,

Now, before I close, I want to just take this opportunity to thank Megan Cox and the rest of the team of organizers from both our Division of Privacy and Identity Protection and our Office of Technology Research and Investigation for putting together today's workshop.

So let me now hand the floor back to Megan. Thank you very much.

MEGAN COX: Thank you Chairwoman Ramirez. It is now my pleasure to introduce our first presenter of the day for an overview on cross device tracking, the policy director in our Office of Technology, Research, and Investigation here at the FTC, Justin Brookman.

JUSTIN BROOKMAN: Thanks very much, Megan. Good morning, everybody. So as Megan said, part of my job here today is to give a little more background on what cross device tracking is.

This is what I spent a lot of the-- my first couple months here at the Federal Trade Commission doing, talking to companies that are engaged in cross device tracking, trying to get a sense of how the technologies and business models work, also doing a lot of tests in our tech lab to get a sense of what about cross device tracking is observable to end users.

It works. So just what I'm going to cover today, background on behavioral tracking, generally, and some the motivations for expanding into cross device, a general summary of some of the methods that companies use to track-- we will get into a little more detail on what probabilistic and deterministic matching models mean-- talk a little bit about some of the research we've done in our tech lab about what we've seen so far about cross device tracking.

And then I'm going to frame some of the open questions that are out there about cross device that Ashkan's and Megan's panel are going to dive into a little more deeply. So this shouldn't be surprising to anyone, but just a little bit of level setting.

This is the most basic 15-year-old model of behavioral advertising. Lots of companies out there operate as third party service providers to publishers, like ESPN and The Guardian, and New York Times here.

And if you serve an add for the first time on ESPN, you have the ability to set a unique cookie. Here's it's 4GASR, et cetera. And then, later, when that company serves ads on-- or analytics or another otherwise embedded on The New York Times or The Guardian that can recognize that cookie.

And so, over time, they can generate a pretty long log of all places that they've seen me around the web or in that browser and can use that information to potentially develop a behavioral profile about that user. OK

So cookies are pretty powerful for keeping state on user over time. They do have some pretty significant limitations. So first, they're very fragile. They're easily deleted or blocked by consumers and often are. Sometimes, security or anti-virus software might be configured to automatically flush out tracking cookies over time.

And they're also-- relevant for this purpose is their browser specific. They're not really designed to enable cross device tracking. They're not even useful for cross browser tracking in the same device, right? If an ad company sees me on Chrome and The New York Times and then later sees me in Firefox or Explorer, another site, it can't necessarily correlate me using cookies. They're browser specific.

They can maybe make a guess based on my IP address or information about my device or the timing of when they saw me, and those are some of the probabilistic methods we're going to talk about in a little bit. But they can't use necessarily cookies for it.

And so this is actually one of the variance we've seen about cross device tracking. It's really just cross browser tracking on one device. Relatedly, browser cookies are not traditionally tied to personally identifiable information for a few reasons.

First, a lot of these companies traditionally engaged in cross device tracking didn't necessarily have a consumer facing side to have a consumer relationship where they get an opportunity to ask, hey, what's your name. But they also-- there has also been longstanding privacy concerns about this. Web browsing has always felt and been described to us as anonymous.

It's the old New Yorker cartoon. On the internet, no one knows you're a dog. And so this came up pretty early in the FTC's work on this issue with the DoubleClick-Abacus merger in 2001. DoubleClick was a behavioral advertising company that was buying Abacus, which had real name information.

In the end, they decided not to merge databases. Abacus was later sold off, but this conceptual line between email and name on one hand, PII, and browsing activity has persisted. And a lot of privacy policies that we've looked at over the past couple months, in preparation for this workshop, still very clearly make that distinction. And there's usually a lot more limitations on the sharing of real name information.

But the ecosystem's changing, and now there is probably a lot more both supply and demand for cross device tracking. So on the demand side, maybe single browser tracking maybe less effective than it used to be because the browsing experiences is lot more fragmented.

We all have more devices. We have a home computer. We have a work computer. We have a home phone and work. We have tablets. And so it's harder to get a consistent viewpoint into what people are doing online. And there's also a lot more supplied information in those same increase in smart devices, but also devices that we've always had are getting smarter and have the ability to generate information, share information about us.

So, first, it was phones. And now they are very smart and have a lot of information about us. Increasingly, gaming consoles are also computers. My Atari 2600 growing up wasn't all that smart, but now Xbox, PlayStations that I have are basically computers.

And TVs are increasingly smart and can able cross device tracking in a lot of different ways. First, cable providers are increasingly able to target individual homes with personalized ads.

It used to be the case everyone in the same area saw the same ad. Now, they can leverage data to deliver personalized ads. And a lot of the cable companies have dedicated commercial slots to each hour for what they call addressable television.

Also, the TVs themselves increasingly have the ability to monitor what you're viewing. And so there are new stories last week about how Vizio updated their privacy policy to say they now have the ability to monitor and share information about what you're viewing with third parties.

Then, going forward, I think other devices are going to keep getting smarter, connected cars and then the Internet of Things. That's symbolized on my little toaster there, which is probably not all that smart.

So the reasons why you might want to engage in cross device tracking. I think targeting is probably the most obvious. So if I'm shopping for shoes at work, don't end up pulling the trigger, maybe they want to show me an ad on my iPad at home when I'm surfing in the evening.

Also useful for security, especially in the first party context. If I suddenly login from a new device, a company might want to check with me to make sure I wasn't being hacked. It could also be useful in the third party context of potentially useful in detecting devices or user agents said to be doing fraudulent page views.

Analytics, companies just might be curious about how people browse their sites or engage with ads on different sorts of devices. Measurement, companies like Nielsen and comScore publish reports about how many people visit which sorts of sites. Having visibility in how users visit across different devices can give them a more rich perspective.

And then attribution, this is people getting credit for ads they shown previously on different devices. And talking to a lot of companies engaged in this space, this seems to be one of the primary early use cases for cross device. And so, here, I see an ad for Wonder Bread on my phone and then later make a purchase on my home computer. Or I press the Amazon Dash button on my smart toaster. Or then maybe I go to Safe Play and buy it.

It's not exactly cross device tracking, but it may leverage some of the same techniques and the same deterministic techniques we'll be talking about a little bit later.

So summarizing the technology. So as chairwoman said, probabilistic is based on inferences about likely connections between devices or users. Deterministic is tying multiple devices to a persistent identifier like an email address, a login credential, or hash personally identifiable information.

So let's start with a probabilistic matching. Here I have three different devices. I've got unique identifiers for all of them, but I don't necessarily know if they're connected right now. But I do have some information where I can start making some guesses.

And so, let's say, this first computer on the left, we see that it tends to be logged in during the day. It usually has the same email IP address. And by the way, it seems to share a local network during the day with a phone, which is during 9:00 to 6:00 on weekdays. Then, the cellphone goes around and does cellphone things. But then in the evenings, in the mornings, and sometimes on the weekends, it's persistently connected to the same home network-- same network, which, by the way, that seems to share with another computer, which tends to be only on in the mornings and evenings and maybe on the weekends.

And so maybe based on that, we can maybe take a guess. That first computer sounds like it might be a work computer, and the computer on the right sounds like it might be a home computer. And maybe the cell phone is the phone that goes back and forth each day. And so maybe you can say with 80% possibility-- number of time completely making up-- that they are probabilistically linked.

And then you have other information too. Maybe, you get location information. And you can say-- you can notice from geolocation information that the work computer and the cell phone tend to be in the same spot a lot. In the evenings, the cell phone tends to be in close proximity to the home computer, and so maybe that gets you up to around 95%.

And then if you're an ad company, maybe you have viewed the browsing habits too. You can see that on one computer, someone reads a lot about technology news and University of Virginia sports and visits Capitol Hill blogs and reads about Arsenal Football. And that's pretty uniquely identifying, right?

I mean that describes me. It probably does not describe anyone else in this room. And that's assuming you all read a lot of tech news. And so based on all that, maybe they can bump it up to a 98% possibility that these three devices are, in fact, linked.

And then we have, like the chairwoman called it, a device graph. You can say that we think these three devices are related. And you can do some of the things we talked about. You can serve re-targeted ads. You can measure for attribution. And so that's probabilistic, again, not tied to personally identifiable information.

Deterministic, there's a few different ways you can do deterministic linking. Obviously, the easiest way is to get people to log in directly from-- into their services. There are lots of services that I log on to on different devices, Gmail, Facebook, Twitter, other mail accounts. Some message board I log into across a lot of different devices.

But some of those devices that I log in to also provide third party functionalities, like advertising or analytics on other websites. And in doing so, they can see what I do off of their services.

So when I said before that, traditionally, a lot of the cross site companies didn't have that consumer facing side, didn't have access to personally identifiable information, increasingly is not really true anymore.

So here it is in the first party context, and this seems pretty straightforward. This is Ashkan's Google account, and they're telling him that you're logged into a Mac. And you're also logged in on your phone.

And I think we're probably glad that Google is doing this. If someone-- suddenly someone logs in from Australia, it may be a fraudulent attempt to access the account. And generally, I think we want and expect first parties to remember what we do across different devices. So if you open an email on one device, it's marked as read when you open it on another.

But a lot of these logged in sites now may appear in other contexts as well. So here I am going to The Washington Post. I'm not logged in. I'm reading a story about why I should be drinking rose wine in the fall. But it's not just Washington Post here. There are a lot of other companies on this page as well.

So on the upper left, you see there are widgets for sharing the stories I want to on Facebook and Twitter and Pinterest. And a lot of times, these widgets are served directly by these companies so they can automatically log the fact that I viewed the story even if I didn't decide to share it.

There's also circled some third party advertisements. They were served by companies that I'm currently logged in to or maybe just given personally identifiable information. They may be able to link that into a cross device profile about me.

And then in the upper right hand corner, which you may or may not be able to see this little ghost that's circled. That's the ghostery browser extension, and it's an extension you can put in to get a sense of how many different third parties are embedded in a particular page.

I can't read that. I think it says 29 or 39, which I think is not inconsistent with a lot of other sites. Certainly, the numbers-- we've seen numbers that go a lot higher than that. They can be serving ads. They could be serving social widges.

They could be embedded just as a blank pixel that I can't even see. But if I had logged in to them, it is possible they could be tracking me across the devices by that personally identifiable credential that I gave them. It doesn't necessarily mean that they are. It really depends on how they've configured their systems. But it's increasingly possible.

So this is the first party deterministic matching, three different devices. I go out of my way to say, hey, I'm Justin on all those devices. And I understand that. But if those services are also embedded in third parties, they have the ability to see what I do off of those services, on those three different devices, and can build up a pretty rich cross device profile about me.

So what happens if I don't have access to login data. One thing I can do is I-- oops! Sorry. Let me get back-- is I can try to find someone who does, look for a partner to get identifying information.

So if you go to a site that I log in to, they could share the fact that, hey, this is justin@email.com. Like I said, there is a traditional reluctance to share personally identifiable information, so maybe, instead, they would share just a hashed version of that identifying information.

And a hash is a function that generates the same exact output when it receives unique input, but it's designed to be very difficult to reverse engineer just by looking at the output. So here, I made up an email address, justin@domain.com. It doesn't work. Don't try to email me there.

And when I run that through the MD5 hashing algorithm, it generates this random-- pseudo random number. Every time I run it through that, it should generate that number.

And so an advertising company could have relationships with various log in sites. If I log on to those sites on different computers, it could send the advertising network that same hash each time to correlate me as a unique user.

So what does that look like? I'm an advertising network, and advertising networks probably an imprecise term these days. Companies tend to be a lot more specialized. But I have visibility to these three different devices. I have a cookie on the desktop computer. I have an advertising identifier on the iPad. And the TV just creates a device ID about me.

At first, I have no way to link them. Well, let's say, instead, that I log on to a new site as justin@domain.com on that desktop. And then I log on to my iPad, could log on to the same news site. Maybe I log into a shopping site instead with justing@domain.com. And I log into a video service from my smart TV to watch movies.

In each of these cases, the company could have a relationship with a third party, and they could send back the advertising network. It could just send back my email address. It could also just send back a hash of that email address, and now the advertising network can see this is the same hash identifier across three different devices and can have that device graph about me. And I can add to it over time. They see other devices come in that have a hash of that same email address, they can add that to the device graph as well.

You can also use email to enable cross device tracking, not email address, but actually email, the service. So here, if I purchase an item on a shopping site as justin@domain.com and I get added to their marketing list, later, if I click on an ad in that email or maybe just even open it, that company may have the ability to embed a unique URL into the email. And then they can know that the browser that opened this is associated with justin@domain.com.

And again, it maybe just-- depending on my mail client when I open it, maybe it just is when I had to actually affirmatively click on something. And, at the same time, if they wanted to then share that information with a third party advertising company, they could send a hash to them.

And then by clicking on the three separate emails or two separate emails and making the shopping experience, that third party company then could maybe add me to a device graph.

And so is hashed personally identifiably information still personally identify information? It provides a layer protection. It doesn't necessarily protect against all threats.

Ed Felten, the chief technologist here in 2012, wrote a blog post about does hashing make data anonymous. If you think it has limitations, then query the alternative. Would you prefer being tracked by unhashed personally identifiable information?

Alan Chapel recently wrote an article op-ed in Ad Age arguing that maybe we should be using personally identifiable information to track people across devices, and it might let smaller players compete with companies that do have access to the login data directly.

And then there are a lot of variations on these different methods. They're blended models, a deterministic company that links people by-- let's say-- a hashed email address could extend their reach to include probabilistic data.

They say, OK, we see these three companies are linked by a unique identifier. But we'll see this fourth device. And it seems like it shares the same IP address a lot, and it's probably linked. They can do that.

Probabilistic companies may also partner with deterministic companies to verify if their algorithm is right. They may say, hey, look we think these companies are related-- or these three devices are related. Are they? And they can get some visibility into how accurate their models are.

And then companies can share device graphs. They can lease device graphs with each other. The company could arrange to do a cookie sync over some period of time where, every time they see a device, they could phone home to a cross device company saying, hey, this is cookie 1234. That's what I call it. And then later they can get a report back from that company saying, OK, we saw all your cookies. We think these bunches are related.

And it's just-- these are some models. Obviously, a lot of these models are continuing to evolve. There are advertising companies that listen to-- that use Bluetooth or microphones to listen to physical beacons or TV advertisements. I think it's fair to say this area is evolving really rapidly and maybe potentially challenging traditional consumer expectations about their privacy and about their consumer-- their computers and smart devices.

So we decided to try to take a look at a bunch of sites and get a sense if we can determine when cross device tracking is going on. We looked at the top 20 sites in a list of categories for news, sports, shopping, games, and reference.

And we spent days trying to get a sense of what's going on. It's really hard to determine objectively, from the end user point of view, when cross device tracking is going on. We certainly detected that a lot of sites that embed companies that do engage in cross device tracking-- probably the highest number of these sites you log in to, like social services. But also on at least considerable minority sites, companies that collect log in data from partner sites or that engage in purely probabilistic cross device tracking.

Again, not clear how these companies are collecting and storing the information. They may be segmenting it. Or they could be collecting personal information just to onboard off site data to upload demographics but not necessarily link across devices.

The considerable majority of sites that we looked at did have the ability to collect personal information, like email addresses from people. Usually, this is part of a registration process. We do see a lot of instances and companies transmitting user names or email addresses to third parties or fairly standard hashes of those email addresses, like MD5 I talked about previously.

Again, not entirely clear why. It could be for other purposes. It could be accidentally. It's just part of the sign up process.

And then we tried to look in the privacy policy of these 100 sites to get a sense of what's disclosed, and there's not a lot of information in there about when information would be shared to facilitate cross device tracking. Often, a discussion of behavioral tracking in general. Oftentimes, a link to one or more of the self-regulatory regimes, trade associations like NAI and DAA.

But it was difficult to get a sense of the scope of cross device just from looking at these policies. A lot did show-- state they would not share personally identifiable information. There are often exceptions for service providers, for business partners.

I'm not clear what that means in this situation. Also, very few discussed what the rules were for hashed personal information. I think only really a handful of those 100 sites talked about hashing PII.

And so how much transparent-- that raises the question, how much transparency should there be? What do consumers expect? Do they want to be overloaded with information? If cross device tracking is going on, what should consumers be told and how?

I know a lot of the self-regulatory regimes do have rules around transparency of tracking mechanisms. I look forward to hearing from the folks on the panels about what their views are on what consumers should be told.

Also, whose responsibility should it be? Should it be the first party that has the direct relationship with the consumer? Should it be a third party that probably have more expertise in cross device tracking? Is sufficient information conveyed by the ad choices icon? And if I click on it, am I told-- what should I be told?

And then what do we do for Internet of Things devices, right? When I hit the Amazon Dash button on my smart toaster to order Wonder Bread, there's no screen for me to get any information. What should I expect with-- how can I learn more information about that?

And then how much control should I have? So this is-- let's say this is my device graph here. A company has visibility into me on all these different devices. If I make the decision to opt-out on my iPad, say, by deleting cookies or by setting some sort of persistent setting or downloading and installing one of the self-regulatory models, what happens for all those other devices?

Should I be opted out for all of them? Is that technologically possible? What happens then if the opt-out flag or opt-out cookie goes away one of those devices? Should I be opted in again?

If this information is tied to a unique identifier, like an email address, should they retain that email address? Use it like a Do Not Call list, so if a new device gets-- tries to get added to the graph, they wouldn't do that going forward.

And then how can this scalablely be done across so many devices in so many different services, right? That's challenging enough on the browser. We've been grappling with it for years.

And how do you actuate your privacy preferences against so many companies on so much of your stuff? And then what should you be able to control? This is something that the chairwoman mentioned. A lot of the self-regulatory models, are you able to restrict targeting?

Should you be able to restrict collection and sharing? Is that possible? Is that plausible?

And then I think the elephant in the room for a lot of these discussions is the rise of ad blocking technologies. It may be done for privacy reasons. It may be done for security reasons. It may be done for performance or bandwidth reasons. It may be done because someone just doesn't like ads, right?

But it has definitely been a palpable push back on user autonomy in recent years. What does this mean for the future of cross device tracking? Is it going to lead to maybe greater restraint? Will it lead to greater capacity? Will it lead to consolidation of market power?

I'm not sure. But I look forward to hearing what the panelists' perspectives are on these and other issues. And with that, I am done.

Thank you very much for your time, and I'll invite Megan back to the stage. Thanks.

MEGAN COX: Thank you, Justin. That was wonderful. It's now my pleasure to invite Ashkan Soltani, our chief technologist here at the FTC and the first panelist for a technological perspective on cross device.

ASHKAN SOLTANI: Thanks, everyone. We're running about five minutes late, so we'll probably go to about 11:50-- Sorry, 10:50. And then we have a break where if folks want coffee or refreshments, there's a little food court down the hall that goes till 11:00, so you just rush over there.

Anyway, thanks everyone for organizing. Thanks to Megan for setting up this workshop. And thanks for our panelists. We will do brief introductions. Joe Hall from the Center of-- I'll let you guys go ahead and introduce yourselves. Joe?

JOSEPH LORENZO HALL: OK. Hi, my name is Joe Hall. I'm the chief technologist at the Center for Democracy and Technology.

ASHKAN SOLTANI: I think it's on. I don't think we need to push.

JOSEPH LORENZO HALL: Don't worry.

JONATHAN MAYER: I'm Jonathan Mayer from Stanford. I'm finishing up a PhD, and I'm also a lawyer.

ANDREW SUDBURY: Hi, I'm Andrew-- Hi, I'm Andrew Sudbury. I'm the founder and CTO of Abine, which is a company that makes privacy blockers and other kind of privacy protecting software for consumers.

JURGEN J. VAN STADEN: And I'm Jurgen Van Staden with the Network Advertising Initiative.

ASHKAN SOLTANI: Great. Thanks. So we heard Justin's great presentation and the chairwoman's remarks about cross device tracking. One thing I'd ask folks to weigh in on is what's different about cross device tracking. Is it just an evolution of OBA and other techniques? Or is there something fundamentally different about the technology that raises new technical challenges and concerns?

JOSEPH LORENZO HALL: So I always talk first. I'll do it again here. So I do think it is an evolution. I do think there are aspects of it that are fundamentally different.

Even deterministic tracking is something that continues the conversation of all the ways in which that could be improved, I think from our perspective, in terms of transparency and control. The cookie is looking mighty wonderful these days compared to some of the stuff.

You may have seen some of the dust up last week around audio beaconing technologies or ultrasonic embedded identifiers that another device you may own is listening for and can then re-associate that device with the device graph as Justin Brookman presented earlier.

There's very little transparency about how those operate. I know there's some changes in how folks associated with the DAA may be reporting their use of that kind of stuff. There's very few user controls. So, for example-- and I'll show up in a second-- we'd have to essentially find a way to make browsers essentially not emit sounds that humans can't hear in order to be able to do some of this stuff.

And it could get even weirder when-- I haven't heard anyone propose this yet, but I'll do it right now-- visual beaconing, so using infrared paint or infrared diodes to cycle something that your phone may be designed to capture using the camera module with apps that do that.

I think apps should have to actually state that they're using that kind of technology.

ASHKAN SOLTANI: Sweet. Anyone else?

ANDREW SUDBURY: I guess I'll jump in. I'll say I think there's a difference because of the way we actually use our devices. I mean we didn't have smartphones that were letting people track our browsing habits, know where we were, nearly 24 hours a day. And so, in addition, we have very little control over what happens on our smartphones for the most part.

We are pretty limited in our ability to even apply the kinds of security and privacy protections that we might have on a traditional desktop machine. I think so the way in which we've used the web as well as the preponderance of the devices that nearly everyone here has one is a substantial difference.

JURGEN J. VAN STADEN: I think I would agree with that. There's-- I think it's more of an evolution. I think I would disagree that this is a collection of data points that we weren't collecting before, that companies weren't collecting before.

Companies are collecting data on mobile devices right now. It's just that we're able to-- or the companies that engage in this practice are able to use it across different channels in a different way. So I think it's more of an evolution, and I think it does raise some concerns. But, in my opinion, it's more of an evolution.

JONATHAN MAYER: So I certainly agree with the evolutionary perspective. So I guess we got everyone on a panel in DC to agree evolution is real. The point I wanted to add was that one particular mode users sometimes take advantage of to try to control their privacy is separating devices or separating browsers.

There's the browser they use for their banks, and then there's the browser they use for all their other stuff, that sort of siloing. And cross device tracking breaks down those silos. So one way in which consumers might try to protect their privacy becomes less viable.

ASHKAN SOLTANI: So what are the primary advantages or benefits to cross device? Chairwoman highlighted being able to keep state on using your eBook reader. Some folks will login to browsers to sync their bookmarks.

Some mobile apps require that you-- or mobile device platforms require that you sign in in order to receive security updates. So what are-- from the purpose of cross device tracking, do you define those as customize tracking? Are the benefits to the user being able to change state?

And then for the benefits to the OBA or ad industry side, is it just for behavioral profiling? Is it for conversion tracking? Is it primarily for attribution, et cetera?

JURGEN J. VAN STADEN: A lot of--

ASHKAN SOLTANI: Yeah. Yeah. So just talk about the benefits, I guess.

JURGEN J. VAN STADEN: Yeah. So a lot of benefits. Certainly, in the OBA, IBA space, I think, one of the key benefits is the ability for marketers to provide consumers with a much better experience.

I think a lot of consumers are frustrated that one particular ad might follow them around 50 times or 100 times. This puts them in a position where a marketer can much more clearly manage their brand and the reputation of their brand.

I also think that there's some serious competitive benefits too-- or benefits competition where when you are talking about cross device, the ability to leverage that technology to stay competitive potentially against some of the first parties that have very large user bases. So there's a competitive standpoint of it as well.

And then the security side of it as well, from an ad prevent-- ad pro prevention detecting malicious technologies. Maybe you've got a bot farm. Maybe you've got a group of IP addresses and types of devices that are affecting your ad serving capabilities.

I think there are certain ad delivery and reporting benefits for this technology as well, not just the delivery of the ad.

ASHKAN SOLTANI: Yeah. If you had to rank them, about top three ranked order?

JURGEN J. VAN STADEN: I think it depends on the company, and I think it depends on who's using it. But I think-- in terms of the advertising industry, I think it's potentially the ability to provide marketers and advertisers the ability to better understand their audience and give publishers the ability to monetize their content in a new way, so I'd say those are probably the two biggest benefits.

ASHKAN SOLTANI: Mhm. Great. Any other benefits?

JOSEPH LORENZO HALL: I was going to say some of these probabilistic methods get pretty close to deterministic in terms of attribution. And so to the extent we had shaky attribution, I think we're going to have less shaky attribution, so that could be certainly a benefit.

ASHKAN SOLTANI: Is this for like showing how many people saw an ad or whether the number of uniques are this--

JOSEPH LORENZO HALL: Yeah. And I should say at CDT, we're a big fan of doing privacy preserving return on investment calculation. And so, to the extent that some of these could be used in a less-- I don't know the right word-- right adjective to use-- in a less privacy impactful way to do some of that kind of stuff, that would be awesome. I don't see that being done right now. It seems to be on the early stages of that.

ANDREW SUDBURY: I would just like to make the point that I think it's clear that first party cross device tracking could have benefits to consumers. But the rest of the benefits we're discussing are not direct benefits to the actual users of the devices.

We were basically saying that, in order to of improve the efficiency and effectiveness of ADR and to prevent click fraud, we're asking consumers to essentially give up information on every site they go to, every place they been, and every purchase they make. And I'm not-- I think you have to look at the entire system to decide where the cost benefits really lie and if this is truly a net benefit to the entire system.

[INTERPOSING VOICES]

JOSEPH LORENZO HALL: ADR? Check. ADR?

ASHKAN SOLTANI: Ad delivery recording.

ANDREW SUDBURY: Yeah.

JURGEN J. VAN STADEN: I'm sorry. You're saying for the ad delivery and reporting part of it.

ANDREW SUDBURY: For all of it. Whether it's the more effective ads leads to better use of advertising dollars and all these other things or we're preventing click fraud, which reduces company's costs, all these are benefits that are not directly to consumers.

You can make some trickle down derivative argument that it's useful to consumers. But I think the trade off of what they're giving up could be-- a substantial.

JURGEN J. VAN STADEN: Well, I do not think that enjoying the content and all the services that we enjoy on the internet for free is a trivial argument. I think it's very clear that consumers derive a tremendous benefit from this technology and from the use of advertising online.

I don't think we should trivialize the fact that consumers enjoy an enormous amount of content for free and services for free that's funded by advertising. I think that's a very real benefit, and for marketers and for advertisers who have a very real business need to manage and protect their brand. This is one technology, in a stack of technologies, that allows them to do that. And I think you'll see the consumer experience get better over time.

ANDREW SUDBURY: I didn't mean to turn it into a debate. For free is debatable. Anyways--

ASHKAN SOLTANI: Any other comments? What about the technical risks, right? So we talked about-- you mentioned some-- can you elaborate what are the primary risks, like collapse of-- Jonathan, you mentioned one, collapse of different devices that you might use for different context from a security posture or-- do you guys want to elaborate?

JOSEPH LORENZO HALL: Yeah. So information leakage is, unfortunately, what I spend-- I'll make up a statistic-- 80% of my time thinking about. And so when Johnathan, for example, showed us the super cookie activities that were happening with, I believe, Verizon and AT&T in terms of injecting into HTTP, so web browsing streams, a unique identifier, that was the kind of thing, as a network guy, I'm like, oh my gosh. This is highly unfortunate.

Now, I have to teach everyone how to use a VPN, which some people should know how to use. But you're having to really get sophisticated in the technical countermeasures you employ so that that kind of activity can't happen. With some of the stuff, like the audio beaconing or the thing I made up earlier, the visual beaconing, I'm not sure we know of ways to protect against that stuff.

We essentially have to start everything with the volume very, very low, which is maybe not a bad idea anyway. But, increasingly, become essentially all of these techniques are surreptitious by design, they don't--

ASHKAN SOLTANI: The deterministic?

JOSEPH LORENZO HALL: The probabilistic--

ASHKAN SOLTANI: Probabilistic.

JOSEPH LORENZO HALL: --types are, for the most part, surreptitious by design, they don't adhere to a lot of the things that we care about in the web framework with the single origin policy and the web security model. That's why the Worldwide Web Consortium Technical Architecture group came out with a statement saying-- it was about unsanctioned tracking, basically saying use tracking mechanisms that respect the security model of the web.

All these unsanctioned forms of tracking that people don't agree to and they don't have good ways to control are fundamentally undermining trust. I mean, eventually-- for some of us, it's turning the web into this very adversarial environment that we typically associate with dissident communities, not consumers for example. But I'll just stop there.

JONATHAN MAYER: So in the same way that the scope of information that's collected is evolutionary, I guess my view would be the risk associated is also-- the change is evolutionary. The risk associated with traditional collection of a users online activities, whether via cookie or some super cookie or any other mechanism, are just more so when those activities are collected across a range of devices.

And those risks, my favorite way of thinking about them is to game out a range of hypotheticals. So one possible outcome is there's a data breach. One possible outcome is there's some bad economic decision in a way-- made in a way that's disadvantageous. Another way is there is some sort of rogue insider use.

You can game out who does what bad thing, under what circumstances, and what's the impact, and then across the range of possibilities, figure out what do you think the probable harm is. And I think everyone can agree that the odds of any particular bad thing happening are probably relatively slim.

The odds that a particular consumer has a particular bad economic decision made about them, for instance, pretty low. The odds that a particular advertising company has a particular sort of data breach, pretty low. But, in the aggregate, I think it's reasonable to think that the scale of risk associated with collecting and storing this activity over a long period of time is starting to get pretty large.

ASHKAN SOLTANI: Any other comments?

JURGEN J. VAN STADEN: I fully agree that technologies that companies use should have appropriate transparency and control mechanisms. And where that's not present, certainly, they would fall afoul of the NAI code and the DAA code. So I think we've got to be careful to not characterize behavior that consumers are clearly informed about and choose to participate in and where effective choice mechanisms are offered.

You can opt out of some of these-- all of these technologies. Certainly for the responsible parties that engage in responsible data collection and advertising practices, they offer choice mechanisms. So while I agree that the technologies might, on the face of it, seem as to your word surreptitious, I think the disclosures have been made.

And certainly, the requirement from industry-- from the self-regulatory side is that the disclosures are made, and that notice is provided, and that consumers are given the choice.

ASHKAN SOLTANI: We'll talk about some of the trace and control mechanisms as well as the new announcement today. Going back to the technology, just for a second, what are the most reliable signals for cross device tracking? So we know in the deterministic side, often, they're authenticated, so you have whatever your user auth is.

And then for the deterministic side, what seem to be the technologies used? And how much transparency or how much control do users have of whether to reveal that on their own?

JURGEN J. VAN STADEN: Are you talking about deterministic?

ASHKAN SOLTANI: So probabilistic.

JURGEN J. VAN STADEN: Sorry, probabilistic? So from a probabilistic standpoint, I think it's all the data that is generally provided into the ecosystem through a typical HTTP request. So certainly, IP address is important although consumers move around from place to place quite a bit. It's my understanding, and so it's less reliable than, say, your cookie ID.

So I still think that that identifier is probably the biggest part of it.

ASHKAN SOLTANI: IP address.

JURGEN J. VAN STADEN: Location. IP address, right. So the location is a big part of it, potentially, as well as some other data points that are within the-- like I said, within the regular HTTP request. It's hard because companies-- this is a proprietary area for companies. And so it's hard to know exactly what their makeup is.

ASHKAN SOLTANI: Cool. Any other comments as to--

JONATHAN MAYER: I guess I-- just to clarify the IP address value in associating devices, just to the networking perspective. So specifically, usually the way home networks are configured, there's one public IP address for the entire home. And then devices on the network have different IP addresses.

And so the consequence is, when you're at home on your wireless network with your laptop and your computer, to the public world, they all have the same IP address. And so that's why, when you're on a home network in particular, there are relatively few devices. And so it's a very strong signal when two devices are consistently showing up with the same public IP address, especially nighttime.

ASHKAN SOLTANI: Right. Even though it changes, right? It's the data point that, at this point in time, they're here. They're co-located. It doesn't matter where they're co-located. It's that they are co-located. Is that right?

JONATHAN MAYER: Yeah. So certainly, the temporal element is super important, not just that they have the same public IP at some time, but there's a pattern at the same time. And that really can divide the world down to smaller and smaller segments. And that's the key insight of probabilistic matching that if you can get the pool of candidates down enough, you can get your match rate very, very high.

As for public IP's rotating, internet service providers differ on their practice there.

JOSEPH LORENZO HALL: And sometimes, you can't intuit quite a bit of information just from that. So, for example, a given block of IP addresses may be unequivocally associated with Comcast's, just as an example, home internet access product versus a business product or something like that.

And maybe another way you can further segment or add a little bit of signal to what may be a very noisy graph.

ASHKAN SOLTANI: And we heard mention of hashed emails and-- or emails and hashed emails. How do those play into how devices are linked?

JURGEN J. VAN STADEN: My understanding is that it's-- if a hashed email is shared, it's shared usually by a party that has the right to share it. So it's a first party comes in and shares that information. It doesn't have to be an email. It could be other data points. It could be a zip code or maybe a name or some other important customer identifier And is then hashed and encrypted and shared.

And then that's the point in which the different devices are linked together and said, OK well, I've got this identifier here with one device and another-- the same identifier with a different device. And then I'm able to link the two devices together.

ASHKAN SOLTANI: This is like if I'm logged into newspaper site, like The New York Times, for example-- as an example. They would-- on my web browser, they would send it to a third party. And then if I'm also logged in on my phone to The New York Times website, they would also send it to the same third party.

And that still falls under deterministic, right?

JURGEN J. VAN STADEN: Correct.

ASHKAN SOLTANI: And then the hashed version, is that still deterministic or is that probabilistic?

JURGEN J. VAN STADEN: No. The hashed version, according to Justin's and I think the way we understand it, is the deterministic version of it. It's something that you know. You're not guessing. You know that this is the same user.

JOSEPH LORENZO HALL: It's pretty fascinating. As nerds, we used to think that the coolest thing to have was a really short email address. So I used to know a guy at Hopkins, Steve Checkoway, who did a lot of the early car hacking work that was s@jhu.edu.

But it turns out, the shorter your email address, the easier it is to reverse, so to speak, use this kind of a hashing property. And just try basically brute forcing trying to find out-- and it's probablt-- you wouldn't call it exactly brute forcing, but it's a form of brute forcing, try and find that one that works.

And now, I guess, what we need are horrifically gory, long email addresses if you're worried about people reverse engineering these simple hashes. Usually, when we hash things and we want to make them hard to reverse, we use what are called password hashing functions, which do, say-- not MD5, but something like that, tens of millions of times, which means it would make someone trying to reverse it very, very-- it would be very, very hard to do that against a whole field of possible candidates.

Anyway, now I'm just rambling. Sorry.

ASHKAN SOLTANI: So actually that's a good point. So how much does hashing provide? Emails are caught at, what, 128 characters? Usually, you have the suffix, which is gmail.com or yahoo.com, so that brings it down to 120 characters. How much is hashed emails-- are hashed emails still PII? Are they still reversible or--

ANDREW SUDBURY: I guess that one of the risks isn't so much can you reverse it. So imagine a whole table of these hashed emails is dumped. So on one hand, you can apply a huge amount of computing power and try and reverse the hash. I presume the hashing method is well known because they want to share and so forth.

But the real issue is what is the threat. If I want to find out if Jonathan-- if that's his device graph, I can feed his email through that same hashing function and then come up with a hash very easily and quickly and cheaply and then look it up in the table and then find the device graph associated with him.

So it really depends on the threat in all of it. So going from the hash table to all the email addresses it represents is challenging. But going from an email address and finding the hash table is very easy.

JONATHAN MAYER: So the kind of-- when computer scientists think about the privacy properties of hash functions, there are generally two classes of attacks that they get concerned about. And I think we've heard both now. One is the person of interest attack, sometimes called an oracle attack, where you have someone you want to look up data about. You're already got their email address.

You've got this pile of data associated with hashed email addresses or any other identifier. And you want to figure out which data belongs to this individual. That's totally trivial. You just apply the hash function to the person's email address or other identifier and then look it up in the data. So that's the person of interest attack. Hashing buys you absolutely nothing against that attack.

Then, there is the brute force, or dictionary or whatever you want to call it, sort of attack where you don't know someone's email address in advance, but you want to reverse as many of those email addresses as you can. So you've got this pile of hashed data. Let's suppose there's a data breach someplace.

Bunch of hashed email addresses associated with web browsing histories get leaked, and someone wants to try to figure out whose email addresses those are. It turns out reversing these hashes is actually really pretty quick.

There are some technical measures you could take to make that a slower process. But it's an incremental benefit. It's very difficult to prevent these sorts of mass attacks. And so out of curiosity, this weekend, I actually ran some quick benchmarks to see how quickly could someone reverse a bunch of these hashes.

And I used the MD5 hash function, which is pretty popular. There are certainly others that folks could use.

ASHKAN SOLTANI: Actually, before you go on, what are the biggest hash algorithms used, MD5, SHA?

JONATHAN MAYER: Yeah. So the MD5 was very popular for a long time. It now is known to have certain cryptographic problems. For purposes of security, it might still be OK for-- the net benefit for privacy purposes maybe still OK. The family that folks have moved towards is a family of functions approved by NIST called the secure hashing algorithm family, SHA.

SHA 3 is the most recent. The entire SHA family is very popular too. So those are the specific algorithms you might hear about.

So OK, some benchmarks on MD5. So formulate in your head what you think a good number would be for how long it should take someone who gets a bunch of breached data to associate a lot of that data with individuals. How much time and effort should it require them?

And while you've got that number, let me give you the benchmarks. So if you wanted to try to match all of the email addresses and other information associated with a pile of data breaches, that kind of few hundred million emails names in some recent large data breaches, which is using that as a benchmark figure.

With the instance I set up using Amazon's Cloud-- very cheap. I think the entire thing costs me under $10-- you could reverse all of those hashes in about a fifth of a second.

So now, let's think about roughly top 1,000 first names, top 1,000 last names, all the combinations of them. And, maybe, let's add a bunch of numbers to the end of that and maybe the top 1,000 domains, just ballparking. That gets you to 15 minutes.

So there's really not a huge amount of privacy protection that comes out of this process. And so I think when and if you're evaluating a product or service that involves hashing to guarantee certain privacy properties, it's really important to ask follow-up questions about what exactly does that hashing buy.

And the answer is very likely none, nothing. But the answer is also very likely not too much.

ASHKAN SOLTANI: So we've heard a bunch of-- Are you going to comment? Yeah. Go ahead.

JURGEN J. VAN STADEN: Yeah. So the hashing algorithm, Jonathan-- just so I'm clear-- you used was an MD5 or SHA when you did this test?

JONATHAN MAYER: I did MD5.

JURGEN J. VAN STADEN: MD5. OK. You didn't-- did you account for the fact that companies might use salt.

JONATHAN MAYER: I didn't, although salt would not affect the computation meaningfully.

JURGEN J. VAN STADEN: Salt wouldn't. So if you didn't salt that computation-- if you did salt the computational, it would effect it?

ANDREW SUDBURY: Salted it per row like per email.

JONATHAN MAYER: Yeah. So I guess what Jurgen is getting at is the-- there was a historical model of attacking passwords offline where you build up a set of hashes that were pre-computed. And applying a salt, so a little bit of extra data to each ID, would provide some protection against that sort of pre-computation attack.

It doesn't provide protection against a live attack where you're calculating the path-- the various hashes on the fly because you can just calculate those with this extra bit of data included. So what Jurgen raises it is an important security protection, but in this particular case, it would actually not provide extra protection.

ASHKAN SOLTANI: So in all these methods, are there-- between hashed emails and regular just non-hashed emails, deterministic versus probabilistic on IPs, are there particular techniques that are more privacy preserving than others? So can we think about ways to do this in a way that, by virtue of the technology itself, provides privacy protections?

We'll talk about control mechanisms after, but just technical solutions to try and achieve cross device tracking. Are there certain ones that are better than others from a privacy perspective?

JOSEPH LORENZO HALL: So there are very exotic things from secure computation, things like private set intersection where-- and we have a blog. If you want to know about this, we have a blog post that explains it-- where two people want to know if they share the same data, but don't actually share that data. And there's a cryptographic method for doing that, and it's pretty intense.

It's only recently become scalable to many hundreds of millions of things. So there are things that are emerging. But I don't know if I consider that a good solution you can use right now.

ASHKAN SOLTANI: This is what-- are you talking about the thing that browsers use for blacklists?

JOSEPH LORENZO HALL: No. So that's a different thing that I don't think would work here. That's-- this is called private set intersection, PSI. It's just a--

ANDREW SUDBURY: The millionaires probably use also.

[INTERPOSING VOICES]

JOSEPH LORENZO HALL: Yes.

ANDREW SUDBURY: The only issue that that would help if companies wanted to share data amongst each other without exposing it. But it still wouldn't change the underlying collection.

JOSEPH LORENZO HALL: It would allow the matching to happen without sharing vast amounts of potentially risky data.

ASHKAN SOLTANI: That would solve the breach issue, no?

JOSEPH LORENZO HALL: I mean you'd still have a bunch of stuff on someone's-- you wouldn't have increasing pools of data because of the sharing so it might limit the size of one given breach to only the stuff that somebody had. But it wouldn't-- that's a perimeter security thing rather than an aggregation of additional information.

JONATHAN MAYER: So, in like a strange sense, things like audio beaconing, of the other sorts of deterministic modes of matching devices that-- or even some of the probabilistic modes-- really anything that doesn't involve sharing an email address, hashed email address, that sort of thing, does in a sense have a privacy advantage.

There isn't this extra bit of information that's associated that's really tightly linked with someone's identity. So I guess one technical approach here would be to avoid the sharing of those sorts of identifiers in the first place.

That, of course, doesn't address a broad range of privacy problems here. But, at least, that specific from a data breach, for instance, you could just solve by not sharing that information and using another approach.

ASHKAN SOLTANI: Like not sharing email, hashed email, IP, location, but sharing like random identifiers, randomly assigned identifiers or random audio beacon signatures, et cetera?

JONATHAN MAYER: Yes. And to be clear, in the IP address cross device tracking, the most common way that works, as I understand it, is one company collects the IP as the user uses multiple devices. It's not like they're swapping IPs on a very regular basis. So that has the advantage, in a sense, of information the company would obtain anyway.

ASHKAN SOLTANI: You guys are [INAUDIBLE]. Any thoughts for helping--

JURGEN J. VAN STADEN: I think when you-- these are great points. But I think when you extrapolate it out over internet scale and you look at the size of these companies, you look at some of the other privacy principles at work, I think it presents a little bit of a different picture. The data is shared-- I don't know if-- to begin with, I don't know of one ad network that has resulted in a big data breach where there's been a lot of emails that have been re-identified because of hashes.

I think that, for purposes of cross device linking, my understanding is that the data-- the hash is shared. The link is made, and then the hash is discarded. So you've got a short period of time, very, very short, in which that information is shared for the purpose of linking in order to create the graph and then that information is discarded.

So you've got a very short period amount of time of retaining the information, which further increases the-- decreases the likelihood of somebody breaching and getting access to information.

You've also got-- this whole scenario, I think, is based on the assumption that somebody is able to hack into the system and get that data or somehow able to see the channel of communication that sees that information, and that you, A, have the desire to go through that process to break that information or discover the emails because it provides some type of pecuniary gain or a value to you.

And I don't know that those assumptions are necessarily true in terms of you're looking at all the risk of vectors that are out there and attack vectors that are out there. I think there are others that are maybe more relevant to companies that if you've got to pick and choose where you're going to spend your security dollars, that's probably not the place where you're going to spend it.

So I think, from an ecosystem standpoint, if you have emails that are hashed PII, it is not perfect, as Jonathan has demonstrated. But it is way more than PII flying through the system. It's way better than having that as well as giving due credit to the fact that there are very strict administrative, contractual, and procedural controls within all of these companies and within the ecosystem as a whole.

The other thing that we've got, as NAI, for NAI members, the moment you take that information and you work to de-identify it-- I'm sorry. You re-identify it. If you were a company, that would require opt in consent. A, it would possibly be a violation of your contractual obligations.

But you also have to-- if it wasn't for some reason, you would still need opt in consent in order to use the PII. So I think on balance, if you look at the ecosystem as a whole, the privacy protections that it affords is, I think, appropriate for the use of it considering that we're not sharing my-- hopefully not-- my tax information isn't being shared. My Ashley Madison account-- I didn't go there, but I'm just saying-- that wasn't shared.

So it's-- what's shared is that I clicked on this shoe ad.

ASHKAN SOLTANI: Yeah. Now, that's a good point. So we had a question, which is that so you guys mentioned data breach. What's the harm scenario if someone gets-- someone malicious gets an email address? What are the incentive to reverse the hash? What's the sensitivity of the data? What could you really do in the event of a data breach with an email address?

ASHKAN SOLTANI: Certainly. So there's clearly pecuniary motives using increasing amounts of information to build profile on someone that you could use to submit fake tax returns. Like, you have to get a considerable amount of additional information than what you would typically get in an advertising environment.

The trick here is that this is now, and we can't assume that people have the goal of money or whatever behind it. Ashley Madison was not done, as far as I can tell, for a monetary reason.

And there's a bunch of others that we've seen that have been either-- a variety of other actors that don't care about money.

And so that depends on what they're motive is. And at least with Ashley Madison, for example, the stated case there was to embarrass or bring home a morality to those people. Anyway, so the point is we can't just focus on monetary incentives.

I think there's a lot of different reasons people might think to crack into your backend system and get access to that stuff.

JONATHAN MAYER: So I guess what the consequences might be-- so clearly the Ashley Madison breach had tremendous consequences for some individuals. It played out in the news. The Sony breach also not seemingly economically motivated, also tremendous consequences.

In one case, a private email exchanges were leaked. In another, the usage of this site that was very controversial. So just imagine there's a breach, and there's a list of folks' email addresses and the websites they went to and when. That would be pretty significant for some people.

Imagine you could look up a coworker's web browsing activity. There's nothing technically stopping someone from setting up a site in the event of a breach that allows you to look that up trivially. In the case of the Sony breach, someone actually did set up an easy search interface.

So I think the potential consequences here are really quite significant, and I guess I would like to know more about the security posture of companies in the space. It's no secret it's a rapidly evolving ecosystem, and that's great, a lot of young companies.

At the same time, it's also no secret that young companies are always the most sophisticated about security. And so are they getting it right? I have no basis of knowing. And at least back when I was working with the California Attorney General's Office, I don't think we got breach reports.

So data breach notification laws, many of you may be aware, are generally administered by the states. And I don't believe we got breach reports out of the ad sector. I think there was a coverage issue in their view. So I guess I don't know whether there was good security applied to this stuff.

ANDREW SUDBURY: I guess that the threats and the harms are-- this is a perennial question. It's hard to define because, as we've mentioned, it can vary so much. For a specific individual, it could be devastating whether it's what medical conditions they have, where they've been, their location. If someone has stalkers, they can be found. I mean it's lots of things. It's just hard to know.

Plus, you have to ask the question of exactly what are all the fields being collected, which I think is hard. I don't know that we know that. I don't know that. I can posit different things, but I'm not aware exactly what these companies have truly accumulated as well as the total-- the scale of just how far back all the data might go.

And then you have to ask yourself, what are all the possible ways this data could be used in the future, what decisions might be based on? And it's just hard to know. But, clearly, harm can be done.

JURGEN J. VAN STADEN: I think those are good points. And I think, generally, those are good concerns to have. And that's why I think the NAI has a lot of really good controls in that space and requirements for the companies that are responsible actors. We've a time-- a data retention requirement, so companies that do collect data have to provide consumers with very clear information about how long they keep that data.

There's also an access requirement. So any of you can go to any one of the NAI members and ask them exactly what data they've got about you. They also have additional requirements to disclose in any sensitive health categories that they may target on and require opt in consent for sensitive health categories and other sensitive categories so that consumers aren't harmed by some of the data sets that could be collected.

I think there are bad actors, and those companies should be dealt with. But I think you should also incentivize companies to do the right thing and make sure that those companies that do do the right thing are recognized for doing that.

ASHKAN SOLTANI: Is there a register of data retention practice when it comes to this data?

JURGEN J. VAN STADEN: We have not set an average one. We've set a definitive one. But I think you-- what I've noticed in the short time that I've been in the NAI is that the retention period seems to be getting shorter and shorter just because the data that is active-- that you can actively target on and help advertisers reach their consumers is getting shorter and shorter.

So I think from that standpoint, you'll probably start seeing companies sharing retention periods with less and less time. Sorry, time limits on that time.

ASHKAN SOLTANI: So you mentioned transparency. So I mean that's a good place to pivot. So, today, the DAA announced-- or released their guidance on-- guidance towards mobile device cross device tracking.

One of the-- the two points were that first parties should also provide notice on their website as well as third party technology providers. Are there-- beyond notice on the website, are there ways for consumers to, for example, know which devices are currently being included in a cross device graph?

So most of the notices are like we may include you in a cross device graph. But are there any transparency tools? Can I-- in the same way that I can know what advertiser set cookies on me or what third parties are present, are there any transparency tools available to consumers to, say, understand their footprint?

JURGEN J. VAN STADEN: Sure. As far as I know, as of today, certainly as an industry, we haven't yet developed that or released that. And I think that's something that we will definitely look at going forward.

But I think this is an area that's-- it's new. It's young. And I know several companies are innovating in the space. And I think you'll continue to see more innovation in the space going forward, so I guess my answer to that is watch the space.

And I think some companies will probably be releasing those kinds of tools sooner rather than later as well as making sure that they contribute as much as they can to helping consumers understand what those transparency tools are and what the disclosures are around cross device tracking.

ASHKAN SOLTANI: Any other add-ons, transparency tools that allow users to know what their associations are?

ANDREW SUDBURY: I mean, certainly, there's tools out there that let you see the different third parties that are involved in a-- when you browse the web. I mean we showed one-- one was shown during the presentation. We make one. There's others.

Generally speaking, those have been less effective on mobile devices, and they also don't tell you a lot about how the data is being used. Obviously, it's just this company was present on this website visit. And, until recently, they've been very challenging to use on mobile devices. You either need to use a non-default browser, for the most part, to even do this.

IOS 9 now makes that more possible on the iPhone. And I guess we'll get into controls later. But there's that.

[INTERPOSING VOICES]

ASHKAN SOLTANI: But these let you know that you made a connection to a company who may engage in cross device tracking, but not whether they are currently cross device tracking or to what probability they're looking.

ANDREW SUDBURY: And for the most part, consumers have no idea what these companies-- they are just these random names when they see them in-- and even when they're shown to them, they have no idea who they are, what they are, and what they do.

And I'm hard pressed to imagine that they actually go visit the sites and learn about all 29 different companies on The New York Times website or Washington Post.

JONATHAN MAYER: So we, in our lab at Stanford, tried to reverse engineer some of the protocols used for swapping information between your browser and some advertising companies. And we had a little bit of luck spotting similar identifiers getting sent across multiple devices where it seemed like cross device tracking had taken place.

That said, I mean we were by no means certain. We were piecing together what had gone on using what limited information we could get in our lab. So I think it's fair to say like when it's hard for researchers to figure out what's going on, it's going to be very hard for the general public.

And the state of research on what the general public knows is very strongly suggested that there is very, very little understanding of, for instance, what the little triangles in the corner of ads mean, what ad disclosures mean. A lot-- apparently there is this common misperception that if a website has a privacy policy, that means they're subject to certain very rigorous federal protections, which is generally the opposite of what the privacy policy says.

So, right, user understanding is very, very low. And then if by some small miracle they land on one of these control mechanisms, there's also a little cottage industry of academic research that suggests users have great difficulty actually exercising control using the mechanisms in a way that conforms the advertising environment to their preferences.

ANDREW SUDBURY: I don't know if we are going to jump into control mechanisms now, but one of the important points about them is that, for the most part, the control mechanisms are not available to the user by default. They literally have to add in software that's generally used for more information security purposes to do egress filtering on the browser, delete cookies, because that one's available to them.

But egress filtering on the browser and other features that are now being added, but are just-- they never-- it's not even-- they didn't even have the choice without adding special purpose software directly to their devices and get the choices.

JOSEPH LORENZO HALL: I'll just say really quick, I do think another part of transparency is notice. I would like to see much more notice about these things. I know there's quite a bit of good critique about the effectiveness of current itty bitty notices. But things like this website is using the following thing you don't understand, but that Joe might or Jonathan might or somebody might understand that may be able to explain it to you, browser fingerprinting or something that said in the audio stream, this app is using audio beaconing detection technology or something.

Those kinds of things are really helpful because it at least gives people the opportunity to ask, what the heck is that? Can I know more? Oh, well, that actually makes sure that I don't see the Progressive Insurance commercial eight times in a row during a hurricane. That's great. Versus other--

ASHKAN SOLTANI: And just briefly, how would that apply to like wearables and watches and--

ASHKAN SOLTANI: And this is really, really early.

ASHKAN SOLTANI: --toasters. I don't think Justin's toaster had a screen.

ASHKAN SOLTANI: You have to toast in a very specific pattern, so the first toast you got said--

JONATHAN MAYER: The first toast is a EULA.

JOSEPH LORENZO HALL: Is there four more loaves?

JONATHAN MAYER: By accepting this toast, you-- But, for instance, iOS and Android now have, I think, a very nice approach to handling this sort of thing. So playing it out in the audio beaconing context, you fire up the app and it says this app would like to use your microphone so that it can associate-- do you allow or disallow? That seems like a pretty good motive, notice and consent.

JOSEPH LORENZO HALL: Yeah. And there's actually-- if you open one of those apps and leave the app, there is a little thing in the top in big red that says this app is currently using your mic. And so it's even more persistent than-- Yeah.

JONATHAN MAYER: Yup.

JOSEPH LORENZO HALL: Awesome.

JURGEN J. VAN STADEN: So, yeah, just to come back to what was said here, control has-- I think there are control mechanisms. If a consumer opts out, that's a control mechanism. It might not give them-- you might need to download additional extensions and ad blockers, et cetera to be able to control other elements. But in terms of opting out of interest-based advertising, that control mechanism is available.

And I think for NAI, we've worked really hard to try and provide those levels of control and help the industry understand what consumers-- the level of control that consumers should have within the degree of possibility.

ASHKAN SOLTANI: So that's a good point. So, today, the DAA announced their control mechanism for cross-- specifically for cross device-- or their guidance for cross device.

And correct me if I'm wrong, my read was that they provide-- the instruct companies to provide an opt out for cross device on a per device basis so that the device or browser that the user exercise the control can opt out, for use and collection, with certain purpose limitations, which I'm not sure whether I-- ad delivery and reporting would fall under that or frequency capping, et cetera.

Maybe you could speak to that. And then they also highlight that there's no choice mechanisms with regards to creation of the graph. So while you can opt out of use and collection, the associations made, you cannot opt out of based on their guidance. So I don't know if you can speak to-- am I reading that right?

JURGEN J. VAN STADEN: Yeah. I think it's a great discussion for the-- great topic for the next panel. I haven't had a chance to read it. But the short paragraph that I did read was on the opt out. And I just read the first two paragraphs-- because it just came out this morning-- that said that the opt out should be for that device.

So my understanding is that that device will be opted out from getting interest-based advertising on that device, and that the interest-based advertising data collected from that device should, I think, not be used on other devices. But I could be getting that right. That's-- I'm getting a thumbs up from some people over there saying that I got it the right way.

So I'll let them speak to it a little bit more.

ASHKAN SOLTANI: OK. And has anyone else had a chance to read it?

JOSEPH LORENZO HALL: I did. There are so many internal references to the other thing that I didn't read, so--

JONATHAN MAYER: So I guess two quick thoughts on-- I also had a chance to read it. The first being, I guess, to emphasize the other way it could've gone. There are two directions to take for control here. One was on a per device basis. The other was on an all your devices basis, to the extent reasonably possible.

So you could have imagined a scenario where you opt out on your phone, now also opt your computer out, also opt your TV out and so on, exercising choice on one device percolates to all your other devices.

ASHKAN SOLTANI: And to be fair, we talked to a number of companies who sadly turned us down to participate. But they did indicate that at least some of the-- what it seemed like, most of the big deterministic companies and some of the largest probabilistic companies, at least one would opt out for the whole graph. If the user opts out of one device, they would opt you out of the entire graph, which is I think what you're talking about.

JONATHAN MAYER: Exactly. And, of course, nothing prevents a company from going beyond the guidance. And, of course, it's a best practice. Do the best you can. So that's the first observation I wanted to make. The second was, just to be very clear with the audience, about the scope of how low the self-regulatory principles are drawn. There's a world of data involving-- or data purposes involving the delivery of behavioral ads.

So you visited a bunch of websites, and inferences are made about what you're interested in. Then, ads are shown to you based on those interests. You like cars. We're going to show you car ads. That sort of thing. So there's that world.

And then there's the world of ad delivery and reporting and security and non-behavioral advertising uses of data. For instance, limiting the number of times you see a particular ad, trying to make sure you're not a fraudster.

And, generally, the scoping of the principles is restricting the behavioral targeting practice, but generally not restricting overmuch the non-behavioral targeting practice. And the reason I want to specifically emphasize that is if you're concern is a company holds some browsing history, some online activity about you-- and that might be breached. It might be accessed by some insider in a way that's unsavory. Again, spin out all of your various concerns short of it being used to show you an ad about that data.

The principles generally don't focus on those risks. And so while there's undoubtedly much good to be said about the principles, they are focused on a very specific set of advertising targeting problems and not necessarily some of the problems we've highlighted today.

ASHKAN SOLTANI: And so if my concern is the linking, I don't want a particular work device linked to my personal activity, are there controls? I guess this is a bit like, what controls exist? And what are some other-- how do the controls interplay with things like browser controls that consumers might use or add-ons or privacy blockers? What's the scope of control with regards to cross device tracking?

JURGEN J. VAN STADEN: Sure. Like I said before, on a mobile device, you've got platform level controls that are respected, at least, by NAI members and DAA members.

ASHKAN SOLTANI: And briefly, what percentage of the industry is NAI, DAA, and then outside of both? Can you speak to that? Like, so how much-- if I were to browse The Washington Post website or crawl the top 100 sites, what percentage of websites would be DAA members-- third parties would be DAA members, NAI members? And then what percentage would not in terms of ad tech?

Maybe Andy can speak to that too.

JURGEN J. VAN STADEN: Yeah, I don't know exactly what the percentages are. But I venture to say that we're the biggest-- or the NAI I membership entails and makes up the biggest swath of it. We've got, I think, 14 of the largest 20 ad networks.

And some of the really large players in the space and then a total of about 98, 97-98 companies that engage in interest-based advertising or similar related activities.

In terms of choice though in the control mechanism, for mobile-- on a mobile device in a cross device graph, if you've opted out of interest-based advertising on a mobile device, then you've opted out of cross device linking as well for mobile. And those controls are available.

And the same is true for regular browser activity. If you've opted out, then that's going to extend to the cross device graph as well.

ASHKAN SOLTANI: So is the linking necssarily-- if you've opted out of the device, the linking will also be--

JURGEN J. VAN STADEN: I think that that's where the companies will have, I think, a different interpretation of what they will do. Some companies, I think, will offer-- will kill the whole graph or ex out the whole graph. According to the DAA principles that we just talked about, it seems like it's just for that one device for interest-based advertising. So--

ASHKAN SOLTANI: And if I clear cookies or block cookies, how does that interplay with the opt out tools?

JURGEN J. VAN STADEN: Well, if the identifier is the cookie, then that device is gone, right? So you don't have that-- the opt out will be gone, too, if it's an opt out cookie. But the device identifier will be gone, so you-- in order to have-- to continue going forward, you would have to get new cookies in order for somebody to re-identify you and create a new graph from that cookie.

JONATHAN MAYER: So I think perhaps the direction Ashkan's suggesting is an important one, that by virtue of doing cross device tracking, one-off privacy controls on a device, unless they persist, you have some sort of blocking mechanism, which I imagine Andrew is going to be able to talk about, the device could just be re-associated with your personal device graph, and then all the data flows back.

So let's suppose you clear your cookies on your mobile phone browser. And then you go home tonight, and that browser gets re-associated with your personal device graph. Then, it's possible to re-associate all your mobile-- past mobile phone browsing activity and all the other cross device activity with this new mobile phone cookie.

So in a sense, the adoption of cross device tracking inherently poses a challenge to one-off technical protections the user might try to--

ASHKAN SOLTANI: Jurgen, you seem like you want to jump in there.

JURGEN J. VAN STADEN: Well, I'm just hearing what he's saying. And I think that's where you will-- why you'll see some companies-- or, I think, a lot of companies offering a full graph opt out. But I also think that if a company has identified you, and especially deterministically-- and that's maybe where some of the probabilistic part of it comes in too. I think companies are very careful to make sure that the representations are accurate.

And so if they're able to say to you, I've opted you out, I think they're going to try and make sure that that device remains opted out to the degree that they're able to. But that's, I think, a practice for them to address. And I think the marketplace will shake this out even more as we go forward.

ASHKAN SOLTANI: I'm going to let Andy jump in. But folks that have questions, do you guys want to just pass them to-- there will be some question takers. And we might have time for one or two more questions. Just pass them to the end of the row, and then these folks will grab them for you and bring bottom up.

ANDREW SUDBURY: So I guess-- I've forgotten what the initial question was. But I guess I'm going to say--

ASHKAN SOLTANI: The question was so what controls exits? How do they interplay with privacy blockers?

ANDREW SUDBURY: Sure. Sure. So as we mentioned, it's a few-- there's a few ways users can try and control just tracking generally, including cross device. One is trying to exercise these opt out choices, whether that's the advertising identifier in their mobile device or going to the NAA site and getting the opt out cookies and so forth.

The second one is to actually try and use more advanced features that they might have available on their browser, like clearing your cookies regularly, which, unfortunately, impacts negatively with some of the NAI cookies. And the third one is to actually try and employ specialized software, like the privacy blockers that we make, things you saw in the browser, that take more advanced steps.

And one of the reasons why I think that's not-- like, literally blocking their requests that might be made to any of these third parties. And, well, I guess I'm a little bit torn because I do agree with Jurgen that we should identify the ad companies that are trying to do the right thing and give them recognition in the marketplace, I think we've seen that what actually happens in practice in the advertising industry is an arms race between the ability to have the consumers to have these choices and then the development of new tracking method whether that flash cookies that we've seen LSOs, eTags, et cetera, that forces to really have the kind of choice that consumers might think they have or desire.

You really have to add lots of additional technical features into a user's device to even make that choice possible in terms of actually effectively blocking tracking. I would argue that something's, there really is no way to do user can have any control over it, specifically the Verizon-- I don't know if we're calling it the super cookie anymore, but whatever. The HTTP header that they were adding.

But I think that is actually problematic in that we really aren't giving users a choice to control this in any sort of reasonable manner.

JOSEPH LORENZO HALL: So I can talk about why I'm not calling the UIDH that Verizon uses as a super cookie, but ask me after this. I no longer call it a super cookie, but whatever, because changes they've made.

In terms of controls, I don't think anyone here wants to browse the web like I do, but you're welcome to come watch me, which is I block all JavaScript, and I block all third party loads into my browser.

And that's a state I don't want to see other people have to do. But it does block quite a bit of this stuff. It means you're not getting any content from someone you haven't specifically approved, and you're not getting any programmatic, dynamic code that might execute in your browser.

The tour browser, which is something we teach a lot of people, the journalists, citizens to use is very useful for this kind of stuff. It completely sets up a VPN you don't have to think about, through your local network, and locks down a lot of this stuff.

I would like to see more programmatic usage of the DNT flag, which is a platform level flag that says I would rather not be tracked. And we've seen-- healthcare.gov, recently, has a functionality where all of the leader and other kinds of code that they included to go get you elsewhere on the web if you flee and don't get health care at healhcare.gov, and actually remove all the script elements once they see that header DNT colon one fly through.

And that's the kind of thing that's really easy to do in most web front-ends, and I'd love to see more people do it. It may mean that you have to actually admit that DNT is useful for something, which may be a problem politically. But I think it's a really useful, technical hook you can hang your hat on and be confident that you're doing the right thing.

ANDREW SUDBURY: One point I forgot to make on the blocking request--

ASHKAN SOLTANI: [INAUDIBLE], real quick.

ANDREW SUDBURY: --and the way that, I guess, Joseph is describing his browsing is that we've found that it's actually really challenging to block content without breaking user experience of the web page. And once users encounter that, they're unlikely to continue to use proxy blockers. That's why we don't actually make an iOS 9 content blocker because we can't do it in a way that doesn't break the user experience.

JURGEN J. VAN STADEN: And just a point to the DNT point that you raised, Joe. You talked earlier about the design of the web and how it should be based on the end user. I think that's where cookies are really great. You have the end user. Whereas with DNT, I think there's a lot of technical steps in between and individuals who can potentially alter that.

And with a cookie ID, you don't have that. So respecting the endpoints of it, I think that's-- it goes to your earlier point, but like you said, it was a tough, tough political discussion.

ASHKAN SOLTANI: Great. Yeah. So I think that's all the time we have for now. We had one question, but it's more suited for the next panel, so I'll give it to-- it's on what the FTC should do. And so I'll leave that to the policy panel.

But we have a 10 minute break, and if you guys want to grab coffee, there's a little cafeteria down the way. We'll promptly be back here at 11:00. And I want to thank our panelists.