SECURITY NOW!

Transcript of Episode #815

Homogeneity Attacks

Description: This week we touch on the Vivaldi browser project's take on Google's FLoC. We look at Chrome's vulnerability-driven update to v89, and then its feature-embellished move to Chrome 90. We consider the surprising move by the FBI to remove web shells from U.S. Exchange Servers without their owners' knowledge or permission, and WordPress's consideration of FLoC Blocking. We also have an interesting-looking programmer's Humble Bundle, some interesting closing-the-loop feedback from our listeners, and a brief progress report on SpinRite. We finish by examining an important privacy guarantee provided by Google's FLoC implementation which prevents homogeneity attacks, where users presenting a common cohort ID also share a sensitive attribute.

High quality (64 kbps) mp3 audio file URL: http://media.GRC.com/sn/SN-815.mp3
Quarter size (16 kbps) mp3 audio file URL: http://media.GRC.com/sn/sn-815-lq.mp3

SHOW TEASE: It's time for Security Now!. Steve Gibson is here. Lots to talk about. Version 90 of Google's Chrome is out. He'll talk about some new features. He'll also talk about Google's Federated Learning of Cohorts, or FLoC. There's a heck of a drumbeat against FLoC, but is it all that bad? Steve examines the privacy implications from a very deep level. Always something of value from Steve Gibson on that. And then we'll talk a little bit about the Humble Bundle for Programmers. Steve's found some books he really likes. And he asks a question: What the heck is Scrum? It's all coming up next on Security Now!.

Leo Laporte: This is Security Now! with Steve Gibson, Episode 815, recorded Tuesday, April 20th, 2021: Homogeneity Attacks.

It's time for Security Now!, the show where we cover the security and privacy of you and your loved ones online with the man in charge of Security Now!, Steve "Live Long and Prosper" Gibson. Hello, Steve.

Steve Gibson: It was really interesting for me to see, you know, because that picture I shared around with Lorrie and me with our wedding rings, where we were both doing the Vulcan hand sign, I didn't know how obscure that was.

Leo: People got it.

Steve: Yeah, everybody got it.

Leo: The caption was "We plan to live long and prosper," so that might have helped.

Steve: Oh, yeah, that's a good point.

Leo: All right. What's on the agenda today, Steve?

Steve: We are Security Now! 815 for, what is this, the second to the last podcast of April. Leo, April is almost gone already.

Leo: I know. Amazing.

Steve: I don't know what's happening. This one, this was originally not titled "Homogeneity Attacks," until I got dug into the topic that I wanted to talk about, which is more an I think really interesting detail about the work Google has put into the privacy-preserving aspects of their FLoC proposal. And there's something known as a "homogeneity attack" which they are preventing, which is the true story behind a lot of the misinformation that's now on the 'Net, because of course this Google FLoC proposal is very controversial, and we'll be talking about some of that this week. Yeah, this week. Right now. In the next couple hours.

But we've got a lot to talk about. We're going to touch on the Vivaldi browser project's take, as I mentioned, on FLoC. We look at Chrome's vulnerability-driven update mid-version 89, and then its feature-embellished move to 90, that has some interesting new stuff, one that really puzzled me for a while. But then I drilled down, and I'll explain it to our listeners. We consider the surprising move by the FBI to removing web shells from U.S.-based Exchange servers without their owners' knowledge or permission, in what I think is pretty much a first for that kind of move.

We've also got WordPress's announced position on maybe adding FLoC-blocking to the WordPress core. We have an interesting look at a programmer's Humble Bundle that I think there's enough there to catch our listeners' attention. We've got some interesting closing-the-loop feedback from our listeners. I'll have a brief progress report on SpinRite to share. And then we're going to finish by examining this important privacy guarantee which is provided by Google's implementation of their FLoC, what they call their "FLoC API," which prevents homogeneity attacks, thus the title of the podcast, where users presenting a common cohort ID also share a sensitive attribute.

One of the critiques has been, and it's a misinformed critique, that sensitive information would be disclosed by the cohort that an individual is in. It turns out Google really wants this to work. And so it'll be clear, I think, by the end of the podcast, the efforts that they've gone to to preserve the privacy of the users, of anyone using FLoC who for now is going to be just Chrome. So I think a useful podcast. And again, I've got lots more to say, so we'll get to that in time.

Leo: Save it. And a Picture of the Week.

Steve: Yes. Okay, so this is a three-frame cartoon, and the first two frames show Mom and Dad clearly very upset at their kids, apparently, we're sort of led to believe. Dad is "We are so disappointed with you." And Mom, "How could you betray our trust like this?"

And then the second frame, Dad, "You're living under my roof. You should be following my rules." And Mom, "We expected better of you."

And then in the third frame, finally, "How dare you sell off maps of our home to the highest bidder?" And we see at their feet a robot vacuum cleaner that they've been talking to. And it's got thought bubbles. It's thinking, "Crybabies! They're the ones who signed off on my user agreement!"

Leo: Ah. Roomba. Roomba.

Steve: Yeah.

Leo: Got some 'splaining to do.

Steve: It's nice when we have long ago covered these topics on the podcast, and then they come out into the popular media. And it's like, oh, yeah, we talked about that.

Leo: Well, these guys, this is JoyofTech.com. This is Nitrozac and Snaggy. They're the ones who did the great moustache painting of you that's on the mugs.

Steve: Oh.

Leo: Yeah, yeah, yeah. They're friends of the network. Yeah, absolutely. They do a lot of our...

Steve: Yeah, that's a really snooty-looking picture. I've got to talk to them about that.

Leo: We can get a new one if you want. We need to gray the moustache a little anyway.

Steve: Kind of like looking down my nose, like okay.

Leo: I know. That's why I like it.

Steve: Okay. Speaking of looking down one's nose, we have the Vivaldi Project's take on the FLoC, uh, I guess maybe you could call it a controversy. It certainly has been picked up in the news a lot. And predictably. The Chromium-based privacy-oriented web browser forks are all up in arms over this proposal. And even the DuckDuckGo search site says they'll be adding FLoC blocker headers to prevent visits to their search engine from registering in Chrome's FLoC aggregation. So it's like, yeah, okay. But in the case of the Vivaldi browser project, their posting last Tuesday was titled "No, Google! Vivaldi users will not get FLoCed." And of course again another reason why this is just the worst acronym or abbreviation.

Leo: Oh, it just begs for that; doesn't it?

Steve: It does. It's just like, you know, don't FloC me and oh, my god.

Leo: Didn't they think about that?

Steve: FLoC no, or, oh.

Leo: Yeah, gosh.

Steve: Yeah. So we can all guess how Vivaldi would be feeling about this. And I want to share the intro of their posting now. But one of the points it makes later we need to address because it's - and this was my original title for the podcast. Google calls this "Sensitivity of Cohorts," which is like the technical term. But I thought homogeneity was more fun.

So here's Vivaldi. They said: "Old habits die hard. Google's new data harvesting feature is nasty. Called FLoC (The Federated Learning of Cohorts), this new advertising technology intends to replace third-party cookies and related technologies like third-party local storage. This clearly is a dangerous step that harms user privacy. Currently, it's being trialed in Google Chrome and is a part of the Chromium browser engine.

"Now the real question: What is Vivaldi's position on this new technology by Google?" And I think I must have skipped the title of theirs. Oh, yeah. I have it in a link above. Oh, yeah, right. "No, Google! Vivaldi users will not get FLoCed." So we can imagine what Vivaldi's position is.

They said: "This is a valid question as we are based on Chromium. But the truth is that while we rely on the Chromium engine to render pages correctly, this is where Vivaldi's similarities with Chrome, and other Chromium-based browsers, end." Although I did see a heading, a headline, and I did not get a chance to pursue it. And it just came out, I think, that apparently Microsoft has also said we may not be enabling FLoC, at least initially.

Leo: No, they're not going to put it in Edge. It's not in Opera, Vivaldi, or Edge. None of the Chromium users.

Steve: Offshoots, right. So they said: "FLoC off," of course. "Vivaldi does not support FLoC." They said: "At Vivaldi, we stand up for the privacy rights of our users. We do not approve tracking and profiling in any disguise. We certainly would not allow our products to build up local tracking profiles. To us, the word 'privacy' means actual privacy. We do not twist it into being the opposite. We do not even observe how you use our products. Our privacy policy is simple and clear: We do not want to track you."

FLoC - and they call it a "privacy-invasive tracking technology." They said: "Google will continue to build profiles and track users in the absence of third-party cookies and local storage. It presents FLoC as a part of a set of so-called 'privacy technologies,' but let's remove the pretense here. FLoC is a privacy-invasive tracking technology." So they said: "Does FLoC work in Vivaldi? The FLoC experiment does not work in Vivaldi. It relies on some hidden settings that are not enabled in Vivaldi. The FLoC component in Chrome needs to call Google's servers to check if it can function since Google is only enabling it in parts of the world that are not covered by Europe's GDPR. It seems there is still some discussion as to whether FLoC could even be legal under the GDPR regulations. We will continue to follow this closely."

And, okay, now that's really not fair. In the trial, Google does have a bunch of Google Chrome syncing stuff that is going on, which is part of the trial mechanism. But that's really separate from FLoC. So it's like, okay, fine. Anyway, they finish the part that I'm going to quote at the beginning, saying: "Although Vivaldi uses the Chromium engine, we modify the engine in many ways to keep the good parts, but make it safe for users. We do not allow Vivaldi to make that sort of call to Google. We will not support the FLoC API and plan to disable it, no matter how it is implemented. It does not protect privacy, and it certainly is not beneficial to users to unwittingly give away their privacy for the financial gain of Google."

Okay. So message received. The Vivaldi folks are not fans of FLoC. Just one more piece of the Internet checking in on how they feel about this. And I don't have a sense for how committed to this Google is. But as we'll see, because I'm going to quote them, a little bit of their background philosophy and sentiment, they really do get it that cookies are on the way out, and that traditional tracking is endangered. And we know that we have Google, they would argue, because of the revenue generated by advertising. And I actually do have some stats that they also quote about the amplification factor of revenue as a consequence of personalization, which is always - I've always been a little bit fuzzy about. But anyway, we'll get to that in a minute.

Talking about browser stuff, Chrome, of course, continues to be the high-value target on the 'Net. Last Tuesday, Google released - they initially released Chrome 89, which patched two newly discovered security vulnerabilities, both which it said exploits for existed in the wild, which allowed attackers to engage in active exploitation. They didn't quite call them zero-days. It wasn't clear whether they had seen them being leveraged, or whether they had seen that the exploits had been published. But still, in any event, nothing that we want.

So one of the two flaws leverages an insufficient validation of untrusted input in Chrome's V8 JavaScript rendering engine. This was the flaw that we actually talked about the week before, or last week's podcast. It was demonstrated by researchers from Dataflow Security during that Pwn2Own 2021 hacking contest. So that was immediately fixed. The other flaw resolved with this v89 blah blah blah dot 128. It was reported by an anonymous researcher on April 7th. So in this case the Chromium team went from report to patch in under one week.

And I can't pass up the opportunity to give Microsoft a little jab in the side and say, "Are you hearing this, Microsoft?" Less than a week from report to patch, and the whole Exchange Server mess would have never happened had they jumped on the early reports of it. But since then, and actually when I went to look at Chrome last night, talking about these vulnerabilities, it started to spin and update itself. And it moved me to 90. So, and it's 90 point whatever. So we have a major new version of Chrome with a bunch of interesting new stuff.

Finally, the long-awaited feature which has appeared for the first time in this v90 of Chrome, which everyone is now getting, at long last defaults to using the HTTPS protocol scheme, or implying it, when none is explicitly specified by the user. So, for example, when just GRC.com or TWiT.tv is entered into the URL bar, that is not specifying the whole https://, Chrome will now first try to initiate a TLS connection to port 443 at that domain, rather than first trying to initiate, as it always has until now, beyond reason, except just the only explanation would be inertia, until now it's gone to port 80 with a plaintext query. And at least in GRC's case, and in the case of many sites, anything coming into port 80 immediately gets turned around with a 302 Moved reply, telling it no, go to https://, and then the same place over on the TLS secure port for the site.

And so the good news is that Chrome's move finally with this v90 to assume HTTPS means that all sites which have been having to perform a similar redirect, bouncing their incoming port 80 queries over to 443, are now going to see a modest performance boost. The page won't have to go to the wrong place and then get rerouted to the right place before it shows. So, yeah, that's good for everybody.

Another change is that another port, I was going to say has been removed, but actually it's been re-removed for reason of NAT slipstreaming, and this is port 554. Google had previously been blocking port 554, but later removed the block after receiving complaints from some enterprise users that they were having a problem because they needed - something that they were doing needed Chrome to be able to initiate connections to port 554. However, after performing some analysis of the use of this port, they determined that it was used for only approximately, get this, 0.00003% of all requests in Chrome.

So they said, you know, given that it is subject to abuse through NAT slipstreaming, and they've got to prioritize security, that low level of incidence said to them there's got to be some other way to do whatever the enterprise users are doing over port 554. It is now blocked with v90, and hopefully the enterprise users will have figured out this was coming.

Also v90 brings us for the first time the newer AV1 AV codec, which increases performance for use in any videoconferencing over the WebRTC protocol. Google has indicated that AV1 basically brings more of all the good things that we want: better compression efficiency than other types of video encoding, reduced bandwidth consumption, and improved visual quality. They said enabling video for users on very low bandwidth networks, offering video as low as at 30 kilobits per second and lower. So more efficient codec means yes, you can deliver better quality at a lower bit rate. And also apparently significant screen sharing efficiency improvements compared to VP9, which had been the codec of choice until now, and other codecs.

Okay. And here's the one that was really puzzling to me. It's apparently rolling out in v90, but it's coming soon because it's not actually working yet for me and for others who were puzzled by this because they've announced it, they've talked about it, and actually it's existed over on the receiving side, but not on the sending side. I'll explain what I mean. This is a new feature which they call Link to a Highlight. Rather than just linking to an entire web page, when you right-click on a highlighted region of a page, the contextual pop-up from some coming version of 90 will have a new item in its context menu, copy the link to the highlight. And what that places on your clipboard is a URL with a pound sign.

And supposedly, if you share this pound sign-embellished URL with others - and this would of course be with Chrome-using others because as far as I know Chrome is the only browser to do this yet, or I should say Chromium, so all of the Chromium browsers probably will because there's no privacy downside. Their use of that link will jump them not only to the page, but to that highlight on the page, with it highlighted. And I actually have a sample of this in the show notes because later on, when I was putting the show notes together, I clicked on a link that came up in Chrome, in Google Search, to Wikipedia, that took me to exactly the phrase I'd been searching for on the Wikipedia page lit up in yellow. And maybe some Chrome users had been seeing that recently, thinking, oh, look at that.

Leo: I can't show it because I use Firefox.

Steve: I know you do, Leo.

Leo: But I'll take your word for it.

Steve: I understand. Okay. So of course this is not the way that pound signs have traditionally worked. Anyone who has coded much HTML will know that it's possible to drop explicit anchors into HTML to which the text following the pound sign in a URL can refer. And all browsers like from the dawn of time will jump to the page and then scroll to that previously placed anchor. For example, GRC's Security Now! pages have always contained an ID tag, which is the anchor, with the episode number in it, so that someone could jump directly to that episode description on the page. You know, it's just sort of proper HTML etiquette to offer that. And it's been there forever.

Well, that's now changing. Certainly that functionality will stay there. I did some digging, and I discovered a very new, as in a W3C working draft dated last month, which proposes a rather dramatic extension to the syntax of what can follow a hashtag in a URL. It's supported, as I said, in Chromium and all the Chromium browsers. And so that's everything except Firefox and Safari. But as I said, since there's no privacy downside - although this draft proposal is really complicated, so I guess they could just like take the source from the Chromium browser and move it over into their implementation because it maybe, you know, there's no reason it wouldn't become a standard. So for what it's worth, that will be coming in 90.

Chrome has for some time been able to scroll to and highlight a link which the Google search engine adds to the tags in order to support that. But users were not able to easily generate their own. Now, you'll be able to just, like if you want to send something to somebody who is a Chrome user, you can highlight a block on a page, right click, copy that special URL which adds that hashtag embellishment, and when they click on it, they'll be taken not only to that random place on the page independent of whatever ID tagging may be there, but also have that highlighted. So, you know, just kind of a cool feature as we move through the evolution of the Internet.

In the surprising news of the week - and this was happening as we were doing last week's podcast, the news was breaking, so we didn't cover it then. The official release, like the news release was dated last Tuesday, April 13th, titled like a press release, for immediate release, from the United States Attorney's Office for the Southern District of Texas. And the release begins, its title is: "Justice Department announces court-authorized effort to disrupt exploitation of Microsoft Exchange Server vulnerabilities."

And in their little brief summary they said: "Action" - that is, the action that was taken. They said: "Action copied and removed web shells that provided backdoor access to servers, but additional steps may be required to patch Exchange Server software and expel hackers from victim networks."

So there's some interesting stuff here that I want to share. They said: "Houston," as in like where this is being sent from. "Authorities have executed a court-authorized operation to copy and remove malicious web shells from hundreds of vulnerable computers in the United States. They were running on-premises versions of Microsoft Exchange Server software used to provide enterprise-level email service. Through January and February 2021, certain hacking groups exploited zero-day vulnerabilities in Microsoft Exchange Server software to access email accounts and place web shells for continued access." And I'll just, you know, to pick a nit, I'll just note that they were not zero-day vulnerabilities because Microsoft was told about them in December. So just said.

They said: "Web shells are pieces of code or scripts that enable remote administration. Other hacking groups followed suit starting in early March after the vulnerability and patch were publicized. Many infected system owners successfully removed the web shells from thousands of computers. Others appeared unable to do so, and hundreds of such web shells persisted unmitigated. This operation" - meaning the one that's being disclosed in this disclosure - "removed one early hacking group's remaining web shells which could have been used to maintain and escalate persistent, unauthorized access to U.S. networks. The FBI conducted the removal by issuing a command through the web shell to the server, which was designed to cause the server to delete only the web shell identified by its unique file path."

Okay. So then the Assistant Attorney General John C. Demers for the Justice Department's National Security Division is quoted in this, saying: "Today's court-authorized removal of the malicious web shells demonstrates the Department's commitment to disrupt hacking activity using all of our legal tools, not just prosecutions. Combined with the private sector's and other government agencies' efforts to date, including the release of detection tools and patches, we are together showing the strength that public-private partnerships bring to our country's cybersecurity. There's no doubt that more work remains to be done, but let there also be no doubt that the Department is committed to playing its integral and necessary role in such efforts."

And one last quote. T, the Acting U.S. Attorney Jennifer B. Lowery of the Southern District of Texas was also quoted, saying: "Combating cyber threats requires partnerships with private sector and government colleagues. This court-authorized operation to copy and remove malicious web shells from hundreds of vulnerable computers shows our commitment to use any viable resource to fight cyber criminals. We will continue to do so in coordination with our partners and with the court to combat the threat until it is alleviated, and we can further protect our citizens from these malicious cyber breaches." And so, okay, I'm skipping a bit of historical background that all of us have memorized by now. But then it finishes with an interesting conclusion.

It says: "This operation was successful in copying and removing those web shells. However, it did not patch any Microsoft Exchange Server zero-day vulnerabilities or search for or remove any additional malware or hacking tools that hacking groups may have placed on victim networks by exploiting the web shells. The Department strongly encourages network defenders to review Microsoft's remediation guidance and the March 10th Joint Advisory for further guidance on detection and patching."

And, finally: "The FBI is attempting to provide notice of the court-authorized operation to all owners or operators of the computers from which it removed the hacking group's web shells. For those victims with publicly available contact information, the FBI will send an email message from an official FBI email account (@fbi.gov) notifying the victim of the search," what they're calling a search.

"For those victims whose contact information is not publicly available, the FBI will send an email message from the same FBI email account to providers such as a victim's ISP, who are believed to have that contact information, and ask them to provide notice to the victim. If you believe you have a compromised computer running Microsoft Exchange Server, please contact your local FBI Field Office for assistance. The FBI continues to conduct a thorough and methodical investigation into this cyber incident."

So, okay, wow. This is the first such known effort to ever have been carried out under the auspices of the U.S. federal government and action on this scale of the FBI. And it's not entirely without some controversy since the federal government technically intruded, uninvited, into the Exchange servers owned by American citizens and altered them. I'm sure they were very careful to keep this on U.S. soil so that our FBI was not reaching out into the Exchange servers belonging to citizens of other countries.

But this is - it must be due to the case that was made to the court that authorized this about the ongoing threat leaving these backdoors in place represented. They referred to it as a "search"; right? So this, maybe they couched this as a search warrant, and the FBI entered the backdoor, found the web shell which they had shown the court was only there, not because the individual had deliberately put it there, but they were able to closely link it to the efforts of a hostile foreign power.

And then notice that they said they "copied and removed." So it must also have been that they said, if we remove this in error, we will have made a copy of it and notified the individuals owning that server that we did this and that we have a copy of it in case we removed it erroneously and so that they would be able to put it back, copy it back to their server. And then, okay, if you want this, it's yours to want it.

For example, you can imagine some security researchers. Who knows how much vetting the FBI did of the hundreds, they said, of instances of this. You can imagine that there may well have been some security researchers who were deliberately running these known compromised Exchange servers for the purpose of gathering data on the use of these backdoors. We don't know.

And in the case of good security researchers, they would not be associated with IPs of their security research firms; right? So the FBI might not have been able to figure out who they belonged to, just because they would want the bad guys to be able to figure out who they belonged to if they were running a honeypot operation. So you can imagine that this is dicey.

And I've told the story on this podcast before of how I was involved in a multiparty conference call, many moons ago, involving officials from the DoJ, some of the politically connected people from the SANS Security Institute, and a number of other security researchers. We'd all got together on a big conference to discuss what to do about some of these Internet worms that were really wreaking havoc back then - Code Red, Nimda, MSBlast. They were scouring the Internet seeking new targets. And at one point there were so many of them that their seeking other victim traffic was choking the Internet at some choke points that were creating spotty DoS for some sections of the 'Net. So, I mean, it was really a problem.

So I remember that we posed the question to the government whether we could use those same well-known by then vulnerabilities, they weren't secrets any longer, ourselves to go into these known-infected systems and remove the worms from them. We believed we could do it safely and in an entirely targeted version. And I well remember the unequivocal response from the DoJ was "Don't even think about doing that." Really. Period. Not even wink-wink. I mean, you will be breaking the law. And of course, yes, without a search warrant court order, that would have been the case. That would have been a cyber intrusion, even if it was people wearing hats that were scrubbed bright white. No, you can't do it.

But in this case the actions that the FBI took were of course legal under U.S. law where our courts have the power to selectively legalize activities within clear boundaries and constraints which would otherwise be illegal. Courts have the power to authorize disconnection, to authorize the FBI to go in and seize equipment that, again, if you convince a court that this is in the public interest to do so, courts can say, yes, you can go in and take all of their computer systems. We've talked about instances where that's happened in the past on this podcast. So I guess it's not surprising that in this instance the courts authorized the FBI to perform as responsible a surgical excision of those backdoors as was possible, and that happened. So, interesting, Leo. Wow. And, you know, probably overall a good thing.

So again, on this FLoC controversy, from a different angle, on Sunday morning, couple days ago, a blog post titled "Proposal: Treat FLoC like a security concern," and this was from WordPress. WordPress suggests four lines of code to block FLoC. They said, after quoting some of the EFF's "Google's FLoC is a terrible idea" blog post, their post begins: "WordPress powers approximately 41% of the web; and this community can help combat racism, sexism, anti-LGBTQ+ discrimination and discrimination against those with mental illness with four lines of code."

Okay, so certainly that would be worth doing if that were the case. The four lines of code relate to what we talked about last week. It's the PHP code which could be added to the WordPress core, which is what they're proposing, to cause the WordPress website to add that FLoC-blocking reply header or response header to everything going out from the WordPress site to the browser. And so this WordPress post proposes that this bit of code should be added to the so-called WordPress Core. So this 41% of the web would be saying we don't want the fact that we went to this WordPress site to be entered into the Chrome browser's aggregation of browser use history which is used to form this FLoC ID.

And I was trying to think, because, I mean, because I've looked at, you know, the 'Net is full of reactions to this. So if nothing else, I'm wanting to understand what the technology is, make sure our listeners understand what it is, that we're just informed. And then it's going to be interesting to see how this shakes out. I have no idea. You know, Chrome by far the majority browser on the 'Net. Probably not within the cohort of our listeners. I don't know. But the facts are what matters. I was trying to think of what the reactions that I've been seeing on the 'Net have put me in mind of. And the first thing that occurred to me was the Apple-Google contact tracing proposal, back when COVID was a new term for us.

And as we'll recall, none of the popular press or even the tech press took the time to understand how the system was designed. We did hear, and it was clear, if nothing else, that it was well-designed with privacy protection as one of its central tenets. But the media just latched onto some of the scary words uttered by others who hadn't bothered to understand the system. And yes, it was complicated. It was not simple to understand, but it was understandable. And as I've educated myself more about FLoC, I'm seeing that the same is true of it, and we'll get to that in a minute.

But I'll share something that just happened to me yesterday. I had a conversation with someone who had so far chosen not to get vaccinated against COVID-19 because he explained to me that the mRNA vaccines contained pig DNA, and that he didn't want pig DNA mixed in with his human DNA. Okay. This person's been a friend for about 40 years, so I didn't want to be rude, and I was caught a bit off guard. I was pretty sure that there was no pig DNA or any other DNA in those vaccines.

So I explained the mechanism by which a deliberately engineered fragment of mRNA is injected, and how it briefly commandeers our own cellular genetics to cause our bodies to synthesize the characteristic COVID-19 spike protein, which our immune systems then see and recognize as a foreign invader and consequently build antibody defenses for any future reappearances or appearances, actually, of the actual spike protein which embellishes the actual COVID-19 virus. And I also explained that the injected mRNA fragments, which are not DNA, are rather quickly degraded and taken apart. They're disassembled by the natural actions of the enzymes that operate our metabolism.

Well, he didn't seem convinced. Until I explained where that weird rumor must have originated. Because one of the components of the vaccine is polyethylene glycol, which for convenience is often abbreviated PEG. So, yeah, the vaccines do contain a small amount of PEG, but no PIG. And when I actually gave him that piece of information, he's like, oh, and he realized where this mistake had come from. He had some information now, and he seemed much more open to the idea.

So we're always going to look at the technology, which is what I want to do here again. At this moment, always subject to change if more is learned, it's clear that FLoC is different from tracking. The mechanism, for one thing, is all on the browser side, as opposed to being spread and distributed all across the Internet. And I would argue that in many ways it is vastly more privacy protecting than the cookies and the fingerprinting that we have today. Assuming that we can - and again, subject to any additional information that we receive, which is how we form our opinions. But assuming that we can truly kill all long-term, all other long-term tracking, to me it seems like an improvement.

Today, using the existing true tracking technologies, not only exactly who you are by web browser, that is, which web browser, and exactly everywhere you go, including how long you stay and what you do while you're there, is all being explicitly tracked and logged, you know, like not just where you went, but all the pages there that you visited. And the trackers are able to infer how long you remained by how long until another beacon from your browser pings somewhere else.

So by comparison, Google's proposal deliberately and significantly fuzzes up only a little bit of that information by reducing that explicit identification and explicit website visiting. And activities, like while you're there, disappear entirely. All of that is reduced to a short hash token that indicates nothing exact about who you are, where you've been, or how long you stayed, and what you were up to while you were there. I mean, it's a real improvement in privacy. But that said, it is a profile tag. No argument there.

And I understand that people don't want to be profiled. I don't want to be. But we keep being told that profiling is the price we pay for an otherwise free Internet. We're told that it supports the commercialization of the Internet and the content that we all take for granted. I've always been skeptical of that, but I have no way of gauging it. Perhaps it's just that it's something like those who already have enough still want more. We know that back in the '50s soap commercials were run during those daytime dramas because housewives in the '50s were watching those, and so products of interest to them were what was shown. Thus soap operas.

So it would be nice if this tracking, this profiling were to end. To me that's sort of a separate issue. It hasn't ended yet. And I have, as I said earlier, I've no sense for how committed Google is to this. Only time will tell. But for me, it's an intriguing technology. And mostly I just want us to understand what it is so, if nothing else, we have a true way of gauging it rather than saying, period, all profiling is bad, no thank you. If we are to believe what the research is said to show, adding profiling doubles the revenue that websites with ads receive. And so that's a big gain.

And for what it's worth, and we'll have a better sense for this as soon as you understand what I have learned about the efforts that Google has gone into protecting our privacy, which is how we're going to finish this podcast, I'd much rather have that, if all traditional tracking can really be killed, if we can kill cookie tracking and fingerprinting and local storage tracking. And the browsers know how we're being tracked. Certainly Google knows.

So anyway, I just - I'm looking at the reaction that is occurring and thinking, you know, folks, okay. We saw this back when we were talking about a very well-designed "who you had been in proximity to" system that Apple and Google together designed. I wish they hadn't called it a FLoC ID. I would argue no one likes being ID'd, especially when that's not what it is. A better name would have been a FLoC CIC, meaning Common Interests Cohort. But that's not what we got from Google. So anyway, we will wrap up by talking about one of the interesting things that they have done to further protect privacy. But I want to share a few other things first.

It is Humble Bundle book time. And I should mention that our listeners must be keeping track of the HumbleBundle.com site because I get references from them to this or that bundle from time to time. Most of them seem maybe of interest, but they don't grab me. We're talking about one this time because it did. This is O'Reilly's Head First series of predominantly programming eBooks that looks pretty much worthwhile. So if you were to buy all these, you would be laying out $772 worth of O'Reilly's books. They are all DRM-free, and they're available in multiple formats. So as we know, it's a tiered system; right? So just $1, pay $1 and you get Head First Ruby, Head First C, Head First PMP. I thought, what, is that a language I haven't heard of? No. That's project management. And also Head First SQL and Head First Statistics. So Ruby, C, and SQL.

And these Head First books are visual, heavy diagrammatic, sort of they're easy to wrap yourself around. So a dollar. If you go to $10, you get those and Head First JavaScript Programming, Head First Learn to Code, Head First HTML & CSS, Head First C#, and Head First Agile. And then again I thought, okay, Agile? Is that a programming language I've never heard of? No. I read the description, and I still have no idea what it is.

They said: "In Head First Agile, you'll learn all about the ideas behind Agile and the straightforward practices that drive it. You'll take deep dives into Scrum, XP, Lean, and Kanban, the most common real-world Agile approaches today." This makes me feel old, Leo. "You'll learn how to use Agile to help your teams plan better, work better, write better code, and improve as a team because Agile not only leads to great results, but Agile teams say they also have a much better time at work." Okay. That all sounds good. "Head First Agile will help you get Agile into your brain" - not into mine - "and onto your team." So anyway, like I said, I still have no idea what it is, but at least for $10 you get all the other ones and JavaScript, Learn to Code, HTML/CSS, and C#, which are all cool.

And wait. There's one more tier. Should you choose to shoot the moon for $18, in addition to all of those, and apparently dramatically improving your agility, whatever that is, you'll also receive Head First Go, of course Go is a cool language; Head First Java, the most popular language; Head First Python; Head First Kotlin; which actually is a language that's like a derivative of Java and actually runs on the Java VM; and also Head First Android Development. All for 18 bucks. So anyway, for coders who might want to stretch themselves out a bit, or for curious non-coders. I get a lot of people who say, like, hey, how do I start programming? I want to code. How do I start? $18, you know, you'll have all these eBooks to kick around and look at. And these are O'Reilly texts.

So anyway, if you just go to HumbleBundle (H-U-M-B-L-E-B-U-N-D-L-E) dotcom. And then you'll scroll down a bit, you'll find the Head First Programming by O'Reilly item. And that's your entry into all of this. So this one I really felt was worth sharing with our listeners.

Three interesting bits of closing-the-loop feedback. I love this one. @dpmanthei tweeted: "Could setting TTL" - that's the field in IP packets which, TTL, Time To Live, and we talked about this back in the How the Internet Works series ages ago. It's a packet which every router, I mean, every router, this is one thing, nobody skimped on this one. Every router decrements the value in that field. It's an 8-bit value. It decrements it by one as it accepts a packet and is getting ready to forward it. If in decrementing it to one, the value goes to zero, the router won't forward it. It just says no. And well-behaved routers will send back an ICMP packet saying "Expired," meaning that the time to live, this packet's life on the Internet expired right here at that router.

So that functionality has enabled some cool things; right? Like that's how trace route works. The way you can figure out the route your packets take on the Internet is by deliberately setting them to expire early and having the routers where they expire send back the equivalent of a ping, but it's a different type of ICMP packet, saying hey, this packet you sent me died here. And then you decrement the TTL one further so it expires on the previous router hop, and that one sends back an ICMP. And then you decrement it again and so forth, all the way back down to one, so you're able to map the route the packets take. So really cool from an engineering standpoint.

Anyway, he says: "Could setting TTL to some low number help, but not solve, some security issues with web interfaces?" He says: "Force admins to be within X hops to login to the web interface?" He says: "Not a solution, but could reduce attack surface." And Leo, this is something you and I talked about, again, so long ago. And, yes, it would be a great solution. I've long thought that, again, it's one of those don't let good be the enemy of perfect. It's like, yes, it's not a perfect solution; but, boy, would it be great. If you knew that you were not ever more than a few hops away, like if you were going to try to access your system through the same ISP as many people do, or that you were no more than four or five hops away, if you were to turn down the TTL on the packets that were leaving your system, your protected system, on a specific port, where you don't want to allow access from a greater distance, it's as good as blocking the port from all but a known IP.

In this case it's blocking the port, think of it as within a known radius, like a known Internet radius of that server. So that nobody further away is able to get to it. So it's, again, not a perfect solution. But I just thought, yeah, that's nice to see somebody else having the same sort of thought because it is another tool that could be deployed.

Someone tweeting, well, he has the name duckDecoy, but he's andrewCoyleNZ, so maybe New Zealand. He said: "I'd be interested to hear your thoughts on how to make a 'vaccination passport' that could not be faked. Any ideas?" Of course this has been in the news because of this idea that maybe at some point, I mean, there's lots of people who are up in arms at the idea that you would have to prove vaccination before doing something, traveling in an area where you would be exposing people to yourself, who knows what. But independent of that, again, don't care about the politics, let's talk about the technology.

Yes, thanks to crypto today and the concept of signatures, it would be absolutely possible to create, for example, a QR Code which contains the identity of an individual, their legal identity through some means, you know, name, birthday, whatever they want, whatever you want to use to anchor the person's real-world identity. And again, that could be whatever it should be. And that information is signed by an authority. And so in the same way that server certificates are signed by a Certificate Authority, and we trust the Certificate Authority, and we're able to verify their signature with their public key, it would be entirely possible to create a system of signed QR Codes or other form of signed digital information, like there are other types of barcodes which are not square, where the digital information contains a signature which cannot be faked. I mean, we know certificates can't be faked. There is no feasible means to fake a cryptographic signature.

So all of the technology exists. If we have vaccination passports, if that comes to pass, I hope they do it right. And it's hard to imagine that they wouldn't at this point because it's so easy to create a digitally verifiable spoof-proof passport or assertion of any kind. We are now able to assert things. That's what SQRL was; right? It was a signed assertion of a domain name that made that whole system work. So these problems are solved.

Oh, and lastly StarKiss tweeted @SGgrc, he says: "I know QNAP deserved the beating you gave them last episode; but looking at system defaults, DLNA is turned off by default, so most systems won't be vulnerable. Same with the Plex bug a month ago. It's not even installed by default, let alone set for external access." So I appreciated the feedback. I should have put that in errata. Not necessarily that it was wrong, but it's worth mentioning, as we know, the tyranny of the defaults.

And a quick progress report. Actually that's what I titled my posting yesterday, yesterday morning, to GRC's SpinRite development newsgroup. And so I'll just share the first portion of it, which will be of interest to our listeners. I wrote: "Work is proceeding quite nicely. I'm finally feeling as though I'm completely back in the groove with SpinRite's old code and its segmented, 16-bit coding environment." I said: "It's taken a while to make the switch cognitively since I code so much by habit, and my habits were all wrong after coding, since 2004, exclusively for the 32-bit unsegmented flat model."

I said: "After a very good weekend, I have all of the drive discovery and enumeration, listing, selection, feature browsing, and display working. I need to determine what I did to break the starting and ending percentage editing, since I've updated its display to show massive sector counts. But something I did back in the beginning broke its UI. This is not surprising since I ripped out tons of code that was no longer relevant, and I needed to make room within the 16-bit fixed size code segment for all the new code I was adding. The changeover to an entirely new drive database also impacted everything.

"So a lot of time has gone into finding and fixing everything that became broken. Once I have the starting and ending percentage screen working, I plan to neuter item #3 from the Main Menu, which is the 'Perform Drive Benchmarks' item. Then I'll release what I have for testing by everyone here," meaning in the GRC newsgroup. "And then, finally, while that testing is underway, I'll work to bring the benchmarking back online." I said: "That's a perfect read-only solution for the next step, since it means that I need to have all of the various ways SpinRite can now access drives - six now, six at last count - different ways for SpinRite to talk to drives working in order to perform that benchmarking. Then, we'll test that, which will be a significant milestone toward completion."

So anyway, I have to say I finally felt like I was back in the groove, really making very good progress. The coding is comfortable again. I've got this, you know, like I've unlearned the joy of having just a single 32-bit flat coding environment, and so I've got the 16-bit segment coding. That's now my default. I'll have to unlearn that again, thankfully, when I move SpinRite over to a 32-bit platform. But that will be joyful in and of itself.

Leo: Yeah, I bet.

Steve: Yeah.

Leo: Let's talk about homogeneity.

Steve: Okay. So again, given the reactions we're seeing to Google's FLoC proposal, I wanted to introduce, again, just like for just the facts, I wanted to introduce their deliberate awareness of its potential for divulging sensitive personal information. Which is one of the ways that it's being attacked. And this, too, is something that really got their attention. So I thought I'd start out by citing a part of, and this is, again, typical of what I'm seeing on the 'Net, part of the not-fully-grounded-in-facts rant from the Vivaldi Project, since it, as I said, it does reflect some widely voiced industry concerns.

So later in that FLoC Off! posting they said: "FLoC intends to do all of the profiling work within the browser. The browser sees everything you browse, so it gathers the data about your browsing habits and determines your preferences. This is not like a browser maintaining your browsing history for you. It is analyzing your personal behavior for

Google. It decides which aspects of your browsing behavior are important. And if enough other people share that behavior, it assigns you the same ID as all of them."

They said: "Advertising companies no longer get to see a unique identifier so they cannot see exactly what you browsed unless they also happen to be the same company that makes the browser you are using so they cannot see you specifically." And even Vivaldi said: "It does sound great." Then they said: "But they can see that every person who buys certain medical products seems to be in the group (FLoC) 1324, or 98744, or 19287." And they said: "Now things start getting ugly." And I'll just mention, I'll thank them for this example because they are completely wrong.

Then they said: "So if you have one of those FLoC IDs, they can display ads for that product, even if that particular medical condition is something you'd rather keep to yourself." They said: "It's all anonymized. Sounds like it should be all right, but that is far from the truth. They can still work out that you have that certain medical issue." Again, no they can't. But again, misinformation. They said: "That you seem to be in a certain age group." No. "That you seem to have certain character traits because you share the same ID as other people who have those traits." Okay, true.

"Statistical analysis of those IDs is harder for small ad companies. They don't get quite so much data to work with. They don't see every website where that FLoC ID appears." That's not quite the way it works, but okay. "The company that gets to know the most about that ID is the one that controls the largest amount of the advertising space: Google." Okay. Some good points, but mostly some anxiety. Right? And it's misplaced.

Their research paper, I've got a link in the show notes because, boy, the math and the charts and things, I'm not going to - I can't do that. I can't share that here. But I'm going to share the essence of what they've done, and the math is there to back it up. The research paper is titled "Measuring Sensitivity of Cohorts Generated by the FLoC API." And the abstract of the paper is two sentences: "We present a discussion of the protections beyond k-anonymity" - I'll explain that in a second - "that the Chrome implementation of the FLoC IC will provide users. These protections mitigate the risk that a cohort number generated by this API in Chrome leaks sensitive information about the browsing behavior of a user." In other words, they understood that anxiety long before it had been voiced, and they addressed it.

Okay. So at first blush that seems to run counter to FLoC's entire goal; right? I mean, profiling is the point. But it turns out that Google really wants this to work. They make this very clear, and that they've given this considerable consideration. Okay. So I should mention something about this "k-anonymity" term. It's an industry-wide, understood within some community term.

Here's what Wikipedia said: "K-anonymity is a property possessed by certain anonymized data." Oh, and by the way, "k" is the number of people in the group. So the concept of k-anonymity was first introduced in a paper published in 1998 as an attempt to solve the problem, and here's the problem: "Given person-specific, field-structured data, produce a release of that data with scientific guarantees that the individuals who are the subjects of the data cannot be re-identified while the data remain practically useful." So it's a formalized anonymity guarantee system.

They said: "A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appear in the release." Meaning cannot be distinguished from all of their others, where k is the individual count. And then they explain, and I thought this was interesting, and a number of terms our podcast listeners will ping on: "K-anonymity received widespread media coverage in 2018 when British computer scientist Junade Ali used the property alongside cryptographic hashing to create a communication protocol to anonymously verify if a password was leaked without disclosing the searched password. This protocol was implemented as a public API in Troy Hunt's Have I Been Pwned? service and is consumed by multiple services including password managers and browser extensions. This approach was later replicated by Google's Password Checkup feature."

In other words, it's the underlying technology that allows you to provide your password and know if it's represented in a group without it ever being disclosed. So cool stuff. Essentially it's a form of statistically provable fuzzing of information to obscure explicit identification.

Okay, so here's how Google places and frames this entire effort. They said: "Today, many publishers are able to leverage interest-based advertising as a source of funding. This revenue stream" - and remember this is coming from Google, and yes, they get all their revenues from advertising; right? "This revenue stream allows them to offer content free of charge to their users. Contrary to contextual ads, interest-based ads leverage information about a user's interests to decide what ads to show them. Interest-based advertising enables an overall better ad experience for users because the user is presented with more relevant ads than traditional run of network ads; for advertisers, who can better reach their target audience; and for publishers, who are allowed to earn more money, on average, per interest-based ad than a non-relevant ad. In fact, multiple studies from academia and industry have consistently demonstrated that personalized advertising can account for 50 to 65% of a publisher's revenue."

Okay. So Google is saying, take it for what it's worth, yes, we clearly realize that no one likes receiving personalized ads, since that means that the advertiser knows something about you in order to deliver something personalized. But independent studies continue to show that advertising that can arrange to be personalized by one means or another really is far more effective for advertisers, and that means generating at least or more than double the revenue for those sites hosting ads.

Okay. So in this paper, introducing, before we get into all the math that we're not going to get into, but to explain their intent here, they said: "In order to accurately serve interest-based ads, ad tech companies use third-party cookies to generate user interest profiles. Thus, the planned deprecation of third-party cookies in Chrome puts interest-based ads, and the revenue publishers depend upon, at risk. To ensure publishers continue to have options to fund their services, Chrome has proposed the FLoC API as a way to enable interest-based advertising in a private way. At a very high level, the FLoC API assigns users to a cohort in such a way that users in the same cohort have similar interests. An ad tech company can then use the API to advertise to an entire cohort."

They said: "It has been shown that the FLoC API allows ad tech companies to enable interest-based advertising without generating fine-grained browsing profiles of users." Okay. And just to stop for a second and remember that fine-grained browsing is what the tracking companies are collecting today. FLoC is only an improvement if we're also able to shut down all other forms of tracking, I mean, like truly kill it.

Okay. So they continue: "The FLoC API achieves this by generating k-anonymous cohorts. That is, the API returns a cohort number shared by at least k users. This ID can be used as an anonymous replacement of a third-party cookie, allowing ad tech companies to build cohort interest profiles without knowing the identity of a user."

They said: "While k-anonymity, especially for large values of k" - right, meaning a huge number of very different people sharing common interests, but who all have the same cohort ID - "especially for large values of k, protects users from re-identification, it is well known in the privacy community that this privacy notion can be vulnerable to so-called homogeneity attacks." Thus the title of today's podcast. "In the context of the FLoC API, a homogeneity attack corresponds to a scenario where all users that share a cohort number also share a sensitive attribute." And that being the key, "sensitive attribute." And of course in the example that Vivaldi gave, that being some medical condition.

Oh, and Google says: "For instance, a cohort that consists only of users who visited a website about a rare medical condition. By revealing the cohort of a user, the FLoC API may inadvertently also reveal that a user has investigated that rare medical condition. At a very high level," Google says, "we want to make sure that no company, including Google, can correlate a particular cohort with any sensitive attribute." Okay. Again, "At a very high level, we want to make sure that no company, including Google, can correlate a particular cohort with any sensitive attribute."

They said: "The purpose of this paper is to discuss the privacy protections that are needed in order to prevent this type of privacy leakage and what Chrome is doing to prevent homogeneity attacks from happening in the initial FLoC API origin trial. As the implementation of the FLoC API is the responsibility of each browser or software that supports the API, the description of the protections here describe only the implementation by Chrome and not necessarily characteristics that are intrinsic to the API itself." Meaning the API does what it does. It's an implementation issue. But if other browsers were interested in doing the same thing, Chromium is open source. So take it.

Anyway, they said: "The sensitive cohort detection described below considers the risk that certain cohorts might imply an elevated likelihood of sensitive browsing behavior. There is a separate threat, not considered in this analysis, of an attacker attempting to guess browsing history based on the details of how cohorts are created. That risk should be mitigated by other measures designed to ensure that the map from browsing history to cohorts is sufficiently lossy, even when conditioned on other information" - this is where the SimHash comes in; right? - "a site might have about one of its visitors. Such measures warrant further investigation, but are out of scope for this document." Meaning that's not what this effort is talking about.

And before I proceed I want to add one more variable, which, to remind people, which is the short, one-week lifetime of the browsing history that informs this FLoC ID in the first place. We've talked a lot in the past on the podcast about the privacy implications of the potentially infinite age of personal information that's being captured about us. Like even mass databases from a decade ago which are then breached on a website. Well, if the information had never been allowed to age, if it expired, then there would be a far reduced privacy compromise potential.

So of course one of the egregious aspects of the current tracking and profiling paradigm that we've all been under is that we have no control over the length of time that our profiles endure. But FLoC sets this limit to a hard seven days. When we're being tracked, actual tracking, where we go is potentially never forgotten. But with FLoC, all profile aggregation of any previous activity disappears after one week. Period. Which to me is a huge difference between the two. No comparison. And I don't know whether, for example, it's a rolling seven days where Google removes the oldest day and adds the newest day, and so your ID is actually changing every day, representing the previous week, or if it changes all at once for the previous week. I haven't got into it. But we do know nothing lasts more than a week.

Okay. So they then describe what they mean by "sensitive." And that's an important thing to understand because they are treating some sites different than others, specifically to protect people from these homogeneity attacks. They said: "Before describing the protections Chrome will put in place, we need to define what sensitive categories are. We will use the same sensitive interest categories defined by Google for its interest-based personalized advertising product."

Leo: So there's no cost to Google for this because they already block these kinds of ads.

Steve: Exactly.

Leo: So, nice.

Steve: Exactly. They said: "This list of categories was chosen because Google already forbids showing ads related to them, as well as targeting a user based on them. Examples of categories in this list are adult and medical websites, as well as sites with political or religious content. We will use these categories to decide whether or not a web page is sensitive."

Leo: These categories are determined by Google and no one else, although some of them are in response to regulatory and legal restrictions.

Steve: Right.

Leo: But there's no - it's not like I can say, oh, I don't want you to follow me based on my interest in assembly language. That's not on the list, yeah.

Steve: Correct. Correct. And what we do have from the proposal is the ability of websites not to have people's visits there be tracked. And so, for example, political, religious, adult, other websites as part of this do have the ability to say "Exclude any visitor tracking here." So that's, you know, very cool. And so sites could say, like somewhere in the fine print, profiling of you during your visit here does not participate. Which of course is not something that we have today. So it's coming from both sides.

So anyway, they said: "While this list of categories certainly does not capture all the nuances of sensitive content (for instance, websites that are not sensitive but a malicious actor might use, perhaps in combination with other data, as a proxy to infer sensitive attributes), we believe it provides us with a solid foundation that we can build upon." And again, important that they're thinking about this.

They said: "Moreover, the methodology presented here can be applied to any other ontology of sensitive categories, as well." They said: "Now that we have established what content is sensitive, we define how we decide whether a cohort leaks sensitive information or not." So they explain: "At a very high level, we want to ensure that no cohort" - meaning ID tag assignment - "consists of users that have visited web pages related to any particular sensitive category at a much higher rate than the general population." In other words, there's nothing about any given cohort that makes them stand out.

So they said: "More formally, we ensure" - and we have another term coming up - "we ensure that a cohort assignment satisfies the strong privacy notion of t-closeness." They said: "A cohort assignment is said to satisfy the property of t-closeness if it is k-anonymous; and, for every sensitive category, the distribution of users who visited a web page related to that category has distance at most t from the general distribution. Intuitively, t-closeness ensures that an adversary that observes a cohort number cannot infer much more about the sensitive browsing behavior of a user than they could before knowing their cohort."

So in other words, this is a formalized and statistically rigorous definition and enforcement of fuzziness with regard to those sites that are deemed to be sensitive. They said: "An adversary who observes a cohort number cannot infer much more about the sensitive browsing behavior of a user than they could before knowing their cohort."

So one criticism that immediately occurred to me would be that not everyone's sensitivities are the same. I might not care much about having my religious affiliation known, whereas I really don't want it known that I spend an inordinate amount of time cruising monster truck websites.

Leo: That's actually not in the categories. So if you start seeing ads for monster truck rallies, you'll know why.

Steve: That's right. I know that I've been - I am in a cohort where, you know, well, maybe I'm in good company.

Leo: Well, but that also points out that one of the problems with this is, if I know anything about you, I might know a lot about you. For instance, I'm suspecting, if you go to monster truck rallies, I know a few other things about even some of these categories they say we won't sell ads against.

Steve: Huh? No, I mean, Leo...

Leo: The real argument is that it's just more data about you that can be used in a fingerprinting attack. Because really that's what people are doing is fingerprinting you more than anything else.

Steve: Right.

Leo: It's good that it changes every week. So it's probably not that useful, yeah.

Steve: It's good it changes every week. And one wonders if - and Google is saying that to their satisfaction they've demonstrated this provides enough profiling to satisfy advertisers. So you have to imagine that, if Google really believed that, they really could thwart fingerprinting. That is, they're saying third-party cookies are going away. We talked about that CNAME horror which allows cookies to be subdomains with the websites in collusion as a way of avoiding the third-party cookie stovepiping problem, which is what even Firefox has begun to do.

So certainly, if it's additional profiling, then that's a lose-lose. If it's truly instead of, if they can robustly prevent cookie tracking, and where cookies again only become useful for state management with the first-party site you're visiting, which is how they were designed and originally intended, and if fingerprinting is fixed, if that problem really is fixed, then we've talked before, Leo, about like the ethics of blocking ads; right? Like if we were to block all ads, we really would be hurting the revenue of the sites we visit.

Well, imagine then if users were given the option of blocking profile-driven ad targeting where the profiling is no longer being tracked and having an endless history of everywhere you go being collected behind your back with no control, to the things you have done in the last week, grouping all of the people who did similar things into a group, and that being presented, and it doubling the revenue, therefore, of the sites that you visit. So ultimately it seems that we're moving haltingly toward people having more control, which is good. Do we want to cut the revenue of sites that depend on advertising in half?

Leo: Yeah, no, that's the problem, absolutely, yeah. And that's one of the reasons we have Club TWiT, because we can't compete at that level. I'm not going to track people to that degree. I just won't do it.

Steve: Right.

Leo: I don't care what advertisers want. And unfortunately that means a large number of our advertisers have gone elsewhere because they want that tracking information. It's sad. You know? It's worse for blogs. It's probably not as bad for podcasts, but it's worst for blogs.

Steve: It's going to be really interesting to see how this goes. And, you know, this thing may just never get off the ground because nobody ever takes the time to get underneath the scary stories about, oh, you're going to be tracked if you ever go somewhere and are identified as somebody with a certain medical condition, when in fact that absolutely has never been true about this.

Leo: Yeah, yeah. There's also, I mean, I think - I don't know. I don't want to say Google's disingenuous. I think they're probably sincere in what they're trying to do. But, I mean, okay, so I'm looking at the categories. And it's good. It's the obvious categories. We don't want to track you based on your financial worries or your psychological troubles or your sexual history, or there's no birth control, things like that. But at the same time, it wouldn't take much to figure out, if I go to a site for nose hair clippers, and then I go to a site for diapers, and then I go to a site for, oh, I don't know what, assisted living facilities, those are all legal. And I think you could make an assumption about people who are in that cohort, that we're probably all a little gray.

Steve: Okay, except there aren't that many bits available to represent cohorts.

Leo: Right, right, right. They're big, yeah.

Steve: They're really big. And so they really, I mean, probably the monster truck website example is kind of right inasmuch as I like big hats with big wide brims and...

Leo: Or to sell you belt buckles because we know you like them.

Steve: You know I've got these boots that come up nearly to my knees, and they've got, you know, I like spurs and, you know. Probably it's like a stereotype identifier more than anything else.

Leo: The problem Google has is they don't want to make it too effective because they want to still sell ads. So if they make it too effective, it's useless to advertisers, and might as well just not track at all. So they're trying to make something, let's not forget, that is useful to advertisers for tracking. For selling.

Steve: So imagine that you had, what, 65,000 categories.

Leo: Is it 16-bit?

Steve: Yes. I saw one ID, but it wasn't a real one. By the way, remember I asked last week, I asked our listeners to go check that amifloced.org that the EFF put up. Not a single - I didn't get a single reply from anyone saying, yeah. Because I was interested in how long the ID is. I saw one that was very scary. It was like 20 digits or something. It's like, woo, whoa, wait, wait, wait, wait, wait. Because, again, I feel very differently about this if there are few people in many cohorts. That's a whole different pie than if...

Leo: I don't think it's settled, to be honest. I think that this is an early testing. And I don't think it's settled how big the cohorts will be. They have not specified it; right?

Steve: It certainly does matter. But, okay. So the example I was going to draw was imagine that you were to design 65,000 characteristic sets. And, I mean, that's a lot; right? There's, like, age, economics, young yuppies buying things for the kids, I mean, if you had 65,000 categories, and each one of those categories could have any number of properties. So my point is you could describe a person really well as one of those 65,000, yet still have a huge number of people that fit that description better than any of the other 64,999 categories. And still, I mean, so yeah, you're learning something about that person. You definitely are; you know?

Leo: You have to, or it's useless.

Steve: Right.

Leo: I mean, that's just - we know it. Whatever Google ends up with, it's going to have to balance the interests of advertisers with our needs for privacy. And given that Google's business is selling ads, I can think who's going to win that one.

Steve: Well, and adoption, too; right? I mean...

Leo: Well, that's the big problem right now. It's already...

Steve: Even Edge is saying no.

Leo: It's already kind of a nonstarter out of the gate.

Steve: Yeah. Unfortunately, I'm reminded of DNT. I thought that was a great idea, too.

Leo: Yup, yup, yup. I think just the fundamental problem is advertisers want to know more. It's become pretty clear that users, not everybody, but a lot of users of the Internet don't want to be tracked. And those are fundamentally inconsistent.

Steve: Right.

Leo: And if your business is selling ads, you're going to have to figure out a way to fix that.

Steve: So advertisers want to know more. Because of lax technology, they have been allowed to know more.

Leo: It's way out of control now, yeah.

Steve: Yes, exactly. So it just kind of - it went wild, and nobody is saying no.

Leo: Right, yeah. And I think Google is, as I said, it's hard to tell, but I think they're sincere in trying to solve this. It's just it's a hard thing to solve, and it's also hard for me to forget that Google's entire revenue comes from selling ads.

Steve: Yes.

Leo: I mean, that's the fundamental bottom line.

Steve: Yeah. On the other hand, how many people use Gmail? How many people would, like, want to give up Gmail?

Leo: Right. Well, I did.

Steve: We are, from them, from Google, we're getting a bunch of amazing...

Leo: Oh, yeah. And we have to pay for that somehow.

Steve: Yeah.

Leo: And basically the way you pay is giving them information about you so they can advertise. Maybe advertisers at some point will just give up and say, look, we don't care, we're just - like in the old days. We're going to buy "I Love Lucy," and whoever watches that's going to see our ad, and we hope that some of them will buy our product.

Steve: And Leo, the advertisers who sponsor the podcasts on TWiT...

Leo: They know what they're getting.

Steve: They're looking at the demographic; right?

Leo: It's no accident. You'll hear our advertisers. They're aimed at tech enthusiasts, yeah. You don't hear a lot of jewelry and flower ads on here. Actually, we've had jewelry and flower advertisers, and they've done pretty well.

Steve: And some really good bedding, too.

Leo: Bedding? Oh, yes, bedding, yes. Well, everybody sleeps. That's an easy one; right? I'm so glad, you know, this is why we love and trust you, Steve, because you're willing to really look at it with an open mind, and talk about the technology, and explain it. And that's what we need. Because almost everything else is hot takes. I don't like Google. I want privacy. Or advertisers got to advertise, or sites got to make money. So it's nice to hear the actual nitty-gritty details. Thank you.

Steve: Well, exactly. And you can make up your mind after you know the facts.

Leo: Right. You need the facts first.

Steve: But you're really not making a decision if it's just, oh, no, that's just more tracking. It's like, well, okay, it's different. Let's look at it, and then at least we'll know.

Leo: It's a real attempt, and I think a legitimate attempt, based in statistics, to track in a less intrusive manner.

Steve: Yeah.

Leo: Steve Gibson. He's at GRC.com. That's his website, the Gibson Research Corporation. There you will find many things, almost all of them free, including the 16Kb versions of this show in audio, the handwritten human transcription so you can read along as you listen. He also has a 64Kb audio. There is one thing you pay for there, though, and that is his bread and butter, SpinRite, the world's best hard drive, any drive, what do we call it, storage maintenance, and recovery utility, I guess; right?

Steve: Yeah, mass storage, yeah.

Leo: Mass storage. Because it works just as well on SSDs. Well, it will. I mean, 6.0 works, but 6.1's really going to work, and he's in the process of developing that. If you buy now, you'll get 6.1 free, and you'll get to participate in the development. GRC.com. You should also check out all the other things he does, including ShieldsUP! It's just a wonderful little place to browse around. Plan to spend the afternoon: GRC.com. He's on Twitter, @SGgrc, if you want to slide into his DMs because they're open there. You can also message him at GRC.com/feedback. And as you can hear, Steve often includes that in the show later.

We have 64Kb audio versions, a little higher quality, and we also have video at our website, TWiT.tv, in this case TWiT.tv/sn. You can download any of the, what is it, 815 shows.

Steve: 814, yeah, 15.

Leo: They're all, well, 15 will be there any minute now. They're all there. You can also subscribe in your favorite podcast client, and that way you'll get it automatically. Steve, we'll see you, let's see, we do the show Tuesdays at 1:30 Pacific, 4:30 Eastern. We'll see you about 2:00 in the afternoon next week. Bye-bye.

Steve: Okay, buddy. Thanks. Bye.