SECURITY NOW!

Transcript of Episode #811

What the FLoC?

Description: This week we briefly, I promise, catch up with ProxyLogon news regarding Windows Defender and the Black Kingdom. We look at Firefox's next release which will be changing its Referer header policy for the better. We look at this week's most recent RCE disaster, a critical vulnerability in the open source MyBB forum software, and China's new CAID (China Anonymization ID). We then conclude by taking a good look at Google's plan to replace tracking with explicit recent browsing history profiling, which is probably the best way to understand FLoC (Federated Learning of Cohorts). And as a special bonus we almost certainly figure out why they named it something so awful.

High quality (64 kbps) mp3 audio file URL: http://media.GRC.com/sn/SN-811.mp3
Quarter size (16 kbps) mp3 audio file URL: http://media.GRC.com/sn/sn-811-lq.mp3

SHOW TEASE: It's time for Security Now!. Steve Gibson is here. We've got a new fix for the Microsoft Exchange Server flaw. This one's automatic, thanks to Microsoft. We'll also take a look at some nice new features in Firefox 87. You can get it right now. And then, what the FLoC? We'll take a look at Google's proposal for replacing third-party cookies. Is it better? It's all coming up next on Security Now!.

Leo Laporte: This is Security Now! with Steve Gibson, Episode 811, recorded Tuesday, March 23rd, 2021: What the FLoC?

It's time for Security Now!, the show where we cover your privacy, your security, your safety online with this guy right here, Steve Gibson from GRC.com. Hi, Steve.

Steve Gibson: Coming to you via Zoom for the first time ever.

Leo: Yeah.

Steve: I know that Alex has been doing this on MacBreak, and everyone's been saying, wow, look how good he looks. Now, okay, I don't look as good as Alex, but the picture is sharp.

Leo: Yeah. Skype has been going downhill, to be frank. And we tried some other alternatives, including an open source WebRTC solution called OBS.Ninja. And then Alex said, "Why are you trying everything else? I've been using Zoom all year," for all his stuff. You know, he does a daily eight-hour thing called Office Hours and stuff.

And he says it's the best. So I trust Alex. If anybody knows streaming, it's Alex. So we're giving it a shot.

Steve: Well, John sent me a link a couple hours ago.

Leo: Latency's low; right?

Steve: Yeah, yeah. And it just, you know, connected right up. And I had it - I think you and I, I guess did we use Zoom when we did that last TWiT special? We did the...

Leo: Oh, for the other, the panel?

Steve: Yeah, the panel. I think...

Leo: Did we? Maybe we did.

Steve: Because I had - I don't know why I would have Zoom on this little machine that I use only for our podcasts. It's like my TWiT box. So anyway...

Leo: Must have used it sometime, yeah.

Steve: We are at Security Now! Episode 811 for March 23rd. Oh, three days before I turn 66.

Leo: Happy birthday.

Steve: Thank you. And we're finally going to get to talk about the topic I've had on our radar for a couple weeks, except that ProxyLogon bumped it off for the last two weeks because we had to talk about that. And that's this so-called, god, the worst-named abbreviation ever. Actually, in discussing this we're going to figure out why it's called FLoC.

Leo: Birds of a feather FLoC together.

Steve: I realized why it happened. And it's like, oh, my god. Then they had to reverse engineer what this horrible abbreviation stands for. So we got Federated Learning of Cohorts, of all things.

Leo: Of course.

Steve: Anyway, we're going to briefly first - I promise briefly - catch up with ProxyLogon news regarding something Microsoft has done that's good with Windows Defender, and also something not so good involving the Black Kingdom. Then we look at Firefox's next release, which will be changing its Referer Header policy for the better. We look at this week's most recent RCE, you know, remote code execution disasters; a critical vulnerability in the open source MyBB Forum software; and China's new CAID, C-A-I-D, which is their - it stands for China Anonymization ID. Uh-huh. Good luck with that.

Then we're going to conclude by taking a long look and a deep dive into Google's plan to replace tracking with explicit recent browsing history profiling, which is probably the best way to understand FLoC, Federated Learning of Cohorts. Oh, and, yeah, as I said, we're going to figure out why they named it that. We've got a surprising Picture of the Week which I experienced, and so I took pictures of it. And I got like a surprise shout-out from an ex-longtime Microsoftie by the name of Dave Plummer, who is most notable for having written Task Manager.

Leo: Wow.

Steve: So some fun things to talk about on this podcast.

Leo: That'll be fun. Oh, all right. Picture of the Week time with Mr. G.

Steve: Java has been a mixed blessing. There was a time when the advice was you really didn't want your browser to be running Java apps behind your back, you know, Java being different and completely unrelated. Unfortunately there's a name collision between it and JavaScript. The two have nothing to do with each other from a technology standpoint.

Anyway, like for whatever reason, I saw the little orange coffee brewing icon on my tray a few days ago. And I thought, what? I've not used Java like forever. I don't even know why I have it on my machine. Actually, I think it was because of a utility that I used to use to extract from my TiVo when I was using TiVo, KTTMG or KMTTG or something like that. It was like TiVo to Go something. But it was a desktop app. I think that's what it was that used Java.

Anyway, so I click on the little orange icon, and up comes this dialog: "Please remove unused versions of Java." And it says: "It appears that you have not used Java on your system in over six months." That's right, I switched to Roku. No more TiVo. And it says: "We recommend that you uninstall it by clicking the remove button below."

Leo: Wow. That's unheard of.

Steve: I know. I thought, really? I'm very impressed. And they said: "If you later decide you need Java, you can reinstall it from Java.com." It says: "If you wish to keep Java on your system, please update it by clicking this update button." So the point is there was an update, but it looked and saw, well, okay, we've got something new, but this guy's not using it. So it'd be better if he just took it out.

So when I clicked on "Remove" because I wasn't using it, up comes the next one, with a yellow caution triangle: "Out-of-date Java versions detected." And then it said: "Keeping out-of-date Java versions installed on your system may present a security risk." And so then it listed the one I had, Java 8 Update 271. And it says: "Click Uninstall to uninstall the selected Java versions." And I just thought, wow, you know, got to give them some credit, where I've given them enough heat over the years. So when something like this happens where they're just being proactively security-conscious, I thought, okay, props. So very cool, Oracle. Good move.

Okay. Before I get into this, because I'm afraid I'm going to forget, Dave Plummer is, like, an original Microsoft, now retired, developer who was there for MS-DOS and Windows 98. He has launched a YouTube channel, Dave's Garage. And so like the first three episodes: "The Secret History of Windows Task Manager Origins," "The Secret History of Windows Task Manager Technology," "The Secret History of Task Manager Source." And actually the fourth one, I'm curious. I haven't watched it: "The Secret History of Microsoft Bob."

Leo: Wow. I knew he had secrets.

Steve: I don't know where that came from.

Leo: Wow.

Steve: Then "The Secret History of Windows Format FAT-32 Limits," which would kind of be interesting. But then we've got Hello, well, actually there was "Hello Windows! Retrocoding 'Hello World' for Windows with Dave," which apparently he did in C. And then I guess before that, or no, even more recently, I'll be interested again to see what he says: "Linux Versus Windows, Round 1, Open Source versus Proprietary, from a Retired Microsoft Dev." Again, interesting to know what he has to say.

Well, the one that happened yesterday was "Hello Assembly! Retrocoding the World's Smallest Windows App in x86 ASM." You know, assembler. And I learned about it because apparently a lot of our listeners have already figured out about Dave Plummer and are watching it. And they're going along, watching it, and if you were to jump forward, Leo, and turn audio on, to 28 minutes, go to 28 minutes, you'll catch the shout-out.

Leo: Right at the end. It's a 29-minute thing here.

Steve: Yup.

DAVE PLUMMER CLIP: 3,072 bytes. I'm pretty satisfied with 3K, but can anyone go smaller while still preserving all the functionality? There were a number of optimizations that I didn't take, such as tail call elimination, smaller strings, eliminating some air checks and so on. To me, anything under 4K smells like victory, but I'd be curious to see if anyone can go smaller than that 3,072.

Leo: I know someone who can.

DAVE PLUMMER CLIP: If we run the app, we find that it indeed works perfectly. It paints our greeting dead center in the main client area. It does it transparently over the gray background...

Leo: Wow, he uses the Windows UI.

DAVE PLUMMER CLIP: ...properly when we resize the window in either dimension. If we click on the Close widget or select Close from the system menu, the application shuts down, just as it was designed to do. And that is that, a complete working Windows application in 3K. Is it the world's smallest Windows app? I believe it is. And unless and until someone shows me a working demo that is less than 3,072 bytes, I stand by it. Notify Steve Gibson that there's a new king in town and bring me his crown and scepter. I hope you've enjoyed this episode of...

Leo: No, there's not. I thought of you. I was watching - so Harvard has an Introduction to Computer Science CS50, very well known. I'm using it to mentor a high school senior in programming.

Steve: Oh, cool.

Leo: Yeah, it's a really great class. It's actually the number one most attended class at Harvard, 800 students every year. Very famous. And at one point they talk about assembly language, which they say, "People used to write code in assembly language. Not anymore." And I thought, maybe not. Maybe there's at least two now I know, yeah. I'm going to watch that. It looks fun, yeah.

Steve: It really was. And there's an entire, you know, he shows you an entire Windows app in assembler. And so I thought, okay. So having thrown down the gauntlet, I went over, and I replied. I said: "Well, Dave did a terrific job with his smallest possible Windows app, except it isn't." I said: "It's small, yes. And he clearly made his point. As he was coding, and I was noting things that could be done better, meaning smaller, I was planning to just let them pass. Again, Dave did a great job. But then I got name-checked. Hold on a second."

Leo: Yeah, big mistake.

Steve: "Now, here's the problem. I know that Dave could make his smaller if he really needed to, if he thought about it. For example, I never compare a register to zero. That's wasteful. But Dave did it several times in his code. When checking a function return for zeroness, Dave compared EAX to zero. That generates the bytes 83F800. And it will set the zero flag if EAX equals zero. So it definitely works. But there's a smaller way to do the same thing. The more efficient way to check a register for zeroness is to OR that register with itself."

Leo: Right.

Steve: "That generates the bytes 0BC0. Yeah, it's only one byte smaller. But it's also 66.6% the size. And it's all about pride in one's code." And I finished: "As many of you know, I'm deep into updating SpinRite at the moment, still going strong on 16-bit DOS. So I'm going to let Dave's challenge stand and keep my focus. And again, we know that Dave could have also made it smaller if he really wanted to. All the best, everyone."

Leo: Is it OR or XOR?

Steve: XOR is the right way to zero a register.

Leo: Oh, okay, right.

Steve: That's better than moving zero into it. And really you could AND or you could OR the register with itself. Either has the same effect of just setting the status and not changing the register.

Leo: Yeah, that's awesome.

Steve: But anyway, I made it this week's shortcut. So for anyone who wants to jump to it, grc.sc/811, since this is Episode 811, grc.sc/811. And he's a neat guy, I mean, he's put in his time. He's worked for Microsoft forever, wrote Task Manager, and has lots of experience and opinions. So anyway, just a counter shout-out to Dave.

Okay. So the latest update on the ProxyLogon fiasco is from Microsoft last Thursday. They wrote: "As cybercriminals continue to exploit unpatched on-premises versions of Exchange Server 2013, 2016, and 2019" - and also apparently 2010, if anyone has something that's 11 years old. They said: "...we continue to actively work with customers and partners to help them secure their environments and respond to associated threats. To date, we have released a comprehensive Security Update, a one-click interim Exchange On-Premises Mitigation Tool for both current and out-of-support versions of on-premises Exchange Servers, and step-by-step guidance to help address these attacks.

"Today" - and this is the reason for their posting - "we have taken an additional step to further support our customers who are still vulnerable and have not yet implemented the complete security update. With the latest security intelligence update, Microsoft Defender Antivirus and System Center Endpoint Protection" - which is part of the enterprise version of that - "will automatically mitigate" - and they have a CVE number, the main entry point - "any vulnerable Exchange Server on which it is deployed."

Leo: Oh, that's a good idea. That's a very good idea, yeah.

Steve: Yes. "Customers do not need to take any action beyond ensuring they have installed the latest security intelligence update, if they do not already have automatic updates turned on." So again, this is a win. I was like, for a moment I was puzzled about this. If a system hasn't had its special early release emergency updates applied, nor the monthly March patches, then how does this help? But their little advertisement graphic in their posting makes their intent more clear.

They said: "Automatic mitigation with Microsoft Defender. Immediate mitigation for threats taking advantage of Exchange Server vulnerabilities." Then they said: "The latest version of Microsoft Defender Antivirus helps mitigate Exchange Server attacks by performing these two actions: Automatically mitigate that CVE via a URL Rewrite configuration" - which I'll explain in a second - and "Scan the server and reverse changes made by known threats."

Okay. So that's really good. Clearly, the point here is, A, that it's one more thing they can do, sort of by way of apology to the world, so that no one can say they didn't do something that they could have done; and, B, unlike the monthly patch updates, which require an administrator's intervention and permission for a full server reboot, the use of Windows Defender, as we know, will eventually be automatic. I mean, I'm sure it's on everywhere. It's unclear to me how often Microsoft OSes check for Defender updates by default. I spent some time trying to get something definitive, and I couldn't find anything. But when I looked on my own system's security widget last night, Defender had just run 90 minutes before I had checked.

Leo: Yeah. It's pretty regular, yeah.

Steve: So it's presumably updating itself all the time.

Leo: Yeah. It's pretty regular.

Steve: And so what this means is on systems that have, like, the administrators wandered off and like are on vacation or something...

Leo: Mm-hmm, exactly.

Steve: ...then this will come along. It will close the entry and scrub what it can.

Leo: Yeah, my last Defender scan was nine minutes ago. So, exactly.

Steve: Right, right.

Leo: They do it all the time.

Steve: Right. And so I mentioned that the vulnerability would be neutered, but not removed. As we know, removal requires the replacement of DLLs that are resident in RAM and have been invoked by low-level services. They cannot be replaced without a full system reboot and reload.

But Microsoft's announcement mentioning the use of a URL Rewrite configuration is interesting. URL Rewriting is an in-line, pre-server, pattern-matching filter that's able to transmute any matching URL into another. So they're really using Defender to do much more in this instance than its normal scan and sequester. It's tweaking the configuration of Exchange's IIS web server, and maybe it looks first to see if the patches have been installed and, like, leaves it - presumably, if there's no vulnerability, it doesn't. But I would bet that it's like, when it's vulnerable, it will add preemptively, proactively add a new URL Rewrite rule to IIS, their web server, to prevent the still-exploitable underlying IIS server from being exploited. And then it goes beyond even that by seeking and reversing known malicious changes that have been made.

So they're essentially using Windows Defender, which as we know checks for updates frequently, as a no-boot mitigation for Exchange Server systems that are critically and chronically unadministered. So this is clearly, as we've said, a good move on their part. It will allow Exchange Server to at least partially be brought back from oblivion without any administration. And unfortunately this arrived on the 18th of March, which was 16 days after the emergency patches went public, which triggered the explosion in scanning and attacks. And it's a full week after the mass scanning was seen that we talked about last week. So, but still, like, we know, what was it, 86,000 still unpatched servers last week, and an exponential drop-off in the rate of patching occurring.

So again, good move on their part. It's a little bit actually building new barn doors that are much stronger after the horses have escaped. But at least they're there now. And again, I really hope, truly hope that Microsoft is taking a serious inward look at what internal systems failed in order for this to have happened. As we've been saying, and as all recent experience shows, it's no longer sufficient to wait a few months, as it once was, two years ago, after being notified of a serious vulnerability. The instant the knowledge exists in the world, the race is on. And we're now seeing that more and more.

Also on the ProxyLogon front, we have Black Kingdom. The other interesting bit of news is that the original DearCry ransomware campaign, which was the first to impact vulnerable ProxyLogon servers, has now been joined by the so-called Black Kingdom ransomware. And we're not going to go into any profound detail about it. As we know, they're all pretty much the same. They're just scanned differently. They have different patterns. And so the various security firms who see them go, oh, look, we haven't seen this one before.

Over the weekend, just two days ago, our friend Marcus Hutchins of MalwareTech blog tweeted that another threat actor was now compromising Microsoft Exchange Servers via the ProxyLogon vulnerabilities. Notice that means that last weekend there were still plenty available to be compromised. He said, based on logs which his honeypots were producing, he said that the threat actor was employing the chained Exchange Server vulnerabilities to execute a PowerShell script that downloads this Black Kingdom ransomware from yuuuuu44[.]com - that's five U's - then pushes it out to other computers across the network.

And this was confirmed by submissions to the ransomware identification site ID Ransomware that we mentioned last week. That's Michael Gillespie's site. He told BleepingComputer that his system has seen over 30 unique submissions of this new Black Kingdom ransomware campaign, with many submitted directly from mail servers. So this is the tendency he's now seen for the last several weeks. When encrypting devices, the ransomware encrypts files using random extensions and leaves a ransom note named "decrypt_file.txt," although Hutchins states that he saw a different ransom note named ReadMe.txt which also had slightly different content. So it looks like they've been tweaking their ransomware a little bit as they've been going on.

In browser news, Firefox will be adopting a new privacy-enhancing Referrer Policy, which I'm glad to see. We're all currently on Firefox 86; 87 will bring this update. We've talked about the misspelled web browser "referer" - R-E-F-E-R-E-R - header often. It's of course supposed to be R-E-F-E-R-R-E-R, but it's not. It was misspelled in the original specification and implementations, so that's what we have today. It's long been controversial because it contains by design, you know, the original architects of the web thought, hey, when a query is sent for some other asset from a web page, it would be really handy for that asset being queried to know who's asking.

So the referer header in a web browser's query contains the URL of the page that referred its visitor to the resource being requested. That could be another web page, a tracking beacon, or anything that the browser fetches from the referring page. And, you know, once upon a time that would be a useful thing to know. But not surprisingly, the referer header has become a source of significant tracking information. Mozilla aims to trim its feathers back a bit.

And if you think about it, remember, like in the really bad old days, you could put a username and password in the URL. There was a syntax for that. And there was something I did for SQRL where one of the things that has always been adhered to is nothing after a pound sign had ever been in a referer. So you can put things that you absolutely don't want leaked behind a pound sign; but of course lots of other information like the query tail, all that stuff after the question mark.

So, for example, search engines often use the whole long search phrase following a question mark in the URL. Well, what that means is that everything that the browser on that page goes out to request - ads, beacons, everything - gets the referer header that until this change would contain that query tail, meaning they know what you searched for in order to get to the page which has then requested all this other stuff. So obviously this is a huge privacy mess.

Mozilla has announced that the next release of Firefox will introduce a more privacy-focused default Referrer Policy to protect Firefox users' privacy. And it's about time. The web browser will, from 87 on, automatically trim user-sensitive information like the path, which is great because it's no one's business, and the query string information, which is accessible otherwise and has been historically from the Referrer URL.

Mozilla's spokesperson said: "Unfortunately, the HTTP referer header often contains private user data. It can reveal which articles a user is reading on the referring website, or even include information on a user's website account." And of course it's actually somewhat surprising and disturbing to see just how much potentially useful to bad guys information is inadvertently leaked by browser referer headers. An examination of web server logs shows referer headers often containing, among other tidbits, internal host names of government and enterprise entities that most likely should never be public. Yet this mechanism publishes them, just sort of blindly, without thinking. And of course bad guys could easily use such information to their advantage.

The first appearance of an explicit Referrer Policy appeared in our web browsers, it varied by browser, but it was around 2016 to 2018, depending upon which browser implemented and which browsers followed. And back then, the web was still largely a hybrid of HTTP and HTTPS. So there was a concern that resources being accessed over the less secure, certainly it was unencrypted and unauthenticated HTTP, should have a restricted view of what was going on on any HTTPS page. So the Referrer Policy then was known as "no-referrer-when-downgrade." That was enacted. So any query to an HTTP page or asset would not have any referer header. It was just eliminated.

So the idea then was that less secure resources would receive much less information about the page requesting whatever their asset was, if that page was protected by HTTPS. So anyway, today Mozilla considers the no-referrer-when-downgrade policy just to be a relic of the past because, as we know, today's web looks much different. We're finally on a path to becoming HTTPS-only. And browsers are taking steps to curtail information leakage across websites.

So Mozilla has decided that it's time for Firefox's default Referrer Policy to be updated. Starting with 87, it will be using what's known as "strict-origin-when-cross-origin," which will trim sensitive information accessible in the referrer's URL. So, for example, where previously Firefox's referer header might be HTTPS, the referer header value would be https://www.example.com/, and then the full path, and the question mark, and then the query. Now, under Firefox 87, when the request is being made to any other domain, the referer header's value in that query from the browser will stop after example.com. That's it. Period. No path, no query. So significantly more privacy.

And really, when you think about it, it's a little jarring that this is only happening now. It's one of those things that, wow, you'd have thought this would have been done 10 years ago. But the problem is these sorts of things tend to break stuff. So pulling back to make privacy-centric changes is something that arguably needs to be done carefully so that things are not broken. So anyway, this new policy will affect everything - all navigational requests, redirection requests, all of a page's subresources, images, styles, scripts, everything, to provide a significantly more private browsing experience. And the best news is we don't have to do anything. This just gets changed in Firefox for us with the next release. So yay.

Leo: Which is out now.

Steve: Is 87?

Leo: Yeah.

Steve: It wasn't last night. Oh, cool. I did a check and an update, and I was at 86.

Leo: You made me check. And I said, oh.

Steve: New, cool. So this week in RCE disasters, remote code execution. The week before last, in the long shadow cast by the ProxyLogon vulnerabilities, the Seattle-based firm F5 Networks, quietly but necessarily, disclosed patches for critical 9.8-scale vulnerabilities, five in all, in their so-called Big-IP and Big-IQ devices. March 10th, F5 released an advisory stating that the REST interface of the iControl management interface is vulnerable to an authentication bypass, which is not something you ever want to hear in anything that is publicly exposed on the Internet, as these are, which includes remote code execution.

No detection rules or artifact information was initially provided by F5, and no public exploit was known at the time of F5's advisory publication. This potentially gave sysadmins time to patch, and blue teams the space to research and implement detection capabilities so that they could get ahead of the bad guys. But in the week that followed, several researchers posted proof-of-concept code after reverse engineering the Java software patch in BIG-IP. And that's all it took. The proof-of-concept code turned the exploitation of the vulnerability from something requiring some real skill into low-hanging fruit. And sure enough, last week the scans and exploitation began.

Last Friday, on the 19th, Bad Packets tweeted that "Opportunistic mass scanning activity detected from the following hosts checking for F5 iControl REST endpoints vulnerable to remote command execution." There was one IP at 112.97.56.78, located in China. There's one, 13.70.46.69 in Hong Kong. And the third, 115.236.5.58, also in China.

So it seems to me what we're seeing is that it may be necessary for the industry's security researchers to reconsider the timing of their release of proof-of-concept code, and to withhold their disclosures, not until the patch has been released, but at least until non-script kiddies have themselves demonstrated that the vulnerabilities have been successfully reverse engineered. That is, hold the proof of concepts until the cat's already out of the bag, until we see exploits in the wild so that it's no longer the case that, as a security researcher, you have actually enabled that to happen. I would argue that no ethical researcher wants to have their proof-of-concept code used as-is, in the wild, wide-scale, devastating and damaging attacks. Yet that's what we're now seeing.

So I think maybe it's going to be necessary, just as a consequence of the dynamics of today's world, to hold onto these things longer. I know that the security researchers are excited to share, like hey, we reverse engineered this from the compiled Java. We know what it was. But yeah, but script kiddies can't do that. So why help them? That just, you know, that doesn't make any sense. And the other thing we know is that just having a patch released is way different than having a patch applied. It's the application of the patch that then makes a proof-of-concept release okay. But release and patch, or release and application, are unfortunately widely spaced events.

Speaking of which, MyBB, the free and open source forum software MyBB, was originally MyBulletinBoard. Then it was shorted to MyBBoard, and now finally to MyBB.

Leo: Probably because no one knows what a bulletin board is anymore.

Steve: Exactly.

Leo: Kids, these used to be called "forums." Before there were forums, there were blackboards.

Steve: That's right. That's right. They're not going to be able to make it any shorter than MyBB.

Leo: MyBB, yeah.

Steve: Yeah. So of course it's written in PHP with a...

Leo: I used to use MyPhpBB. Is that the same?

Steve: Oh. No, I don't think it is.

Leo: Okay. There's another one.

Steve: There's that, too. That's another one.

Leo: Okay. All right.

Steve: This one is written in PHP with a MySQL database backend. The good news is it's not massively popular. It's got around 2,100 potentially vulnerable domains showing MyBB present. Until patches were released on March 19th, it had a pair of critical vulnerabilities that could be chained to achieve remote code execution without the need for prior access to a privileged account. The flaws were discovered by two independent security researchers, Simon Scannell and Carl Smith, and they were reported to the

MyBB team on February 22nd. And as I said, on March 10th, an update was released to close the holes. So that's a nice, you know, February 22nd, so what, 18 days? No, February only has 28 days. So like a little over two weeks; and, bang, we now have a hole closed.

So according to the researchers, the first issue, a nested auto URL persistent cross-site scripting vulnerability, stems from how MyBB parses messages containing URLs during the rendering process, thus enabling any unprivileged forum user to embed stored cross-site scripting payloads into threads, posts, and even private messages. That's not good. MyBB's advisory said: "The vulnerability can be exploited with minimal user interaction by saving a maliciously crafted MyCode message on the server," they said, "for example as a post or private message, and pointing a victim to a page where the content is parsed," which is trivial to do.

The second vulnerability is an SQL injection in a forum's theme manager that could result in an authenticated remote code execution. A successful exploitation occurs when a forum admin with the "Can manage themes" permission imports a maliciously crafted theme, or a user for whom the theme has been set visits a forum page. So by chaining these, it's pretty simple to do. As a result, the researchers' write-up, they said that: "A sophisticated attacker could develop an exploit for the stored cross-site scripting vulnerability and then send a private message to a targeted admin of a MyBB board. As soon as the admin opens the private message on his own trusted forum, the exploit triggers. An RCE vulnerability is automatically exploited in the background and leads to a full takeover of the targeted MyBB forum."

The researchers waited eight days after the patches were made available to publish their work, which included, unfortunately, a complete soup-to-nuts description and discussion, with examples, of the exploit. So while previously an attacker may have needed to be sophisticated, as they said in their write-up, when armed with their complete and detailed how-to, not so much. Was eight days long enough for them to wait? Did every instance of MyBB get patched and updated during the interim? Well, we can certainly hope so. But we pretty much know that that won't have happened. So maybe MyBB is not big enough a target to cause much pain.

On the other hand, if any high-value sites are running MyBB, I'll bet you that state actors have built themselves a database that cross-references all of the valuable targets with all of the publicly exposed technologies they have. For example, there's just no way that China and Russia, they seem to be where these attacks are coming from. I mean, and not just at the U.S., but globally. I'll bet you that their teams have a database such that when a problem is announced by F5, they type F5 into their database, and it tells them every valuable high-profile target using F5 hardware. And they immediately launch attacks based on the reverse engineering of that vulnerability. That's the world we're in today.

So if there were any high-profile users of MyBB, and this release came out, I'll bet you that it didn't take long for an attack to get launched. That's, again, isn't that what any serious attacker would do? They would build a reverse index of all the technologies used by all the targets they care about. And if we learned anything from the last few months of attacks, certainly from the SolarWinds, where we saw a seriously committed threat actor who we believe to have been state-sponsored, we saw them do absolutely everything right. Well, part of doing it right is building an index of who's using which technologies on the Internet. That has to exist.

So I couldn't resist, Leo, calling this one "CAID is able."

Leo: As in Cain and Abel, all right, I got it, I got it.

Steve: Of course. I know you would. CAID, C-A-I-D, is the China Anonymization ID, which is an indirectly Apple-inspired, well, because it's a workaround, to Apple's forthcoming plans to dramatically tighten up the tracking allowed by Apple Store apps. And this refers to what you were just talking about, about the recent change in Apple policy that forced Google to disclose what their stuff was doing. And oh, boy...

Leo: What a list.

Steve: ...is it a laundry list of, like, in fact, I was thinking of putting a picture in the show notes. But in order to show it all...

Leo: More than one page.

Steve: ...the print has to be so tiny that you can't read it. So I thought, uh, no. Okay. But I thought this would be the perfect segue for this week's discussion, which we will conclude with, of Google's FLoC initiative. Okay. So we'll be coming back around to that in a few minutes. But eight days ago, on the 15th, the privacy-focused DuckDuckGo search engine tweeted. Their tweet reads: "After months of stalling, Google finally revealed how much personal data they collect in Chrome and the Google app." And DuckDuckGo says, "No wonder they wanted to hide it."

So they said, and of course we know where DuckDuckGo is coming from; right? They actually showed this as a side-by-side, and I do have the image from their tweet. The very left-hand little thing shows DuckDuckGo.com, and there's like, nothing there. They're proudly collecting nothing. And then they have Google Chrome, and then next to it Google apps in general, with the laundry list of all the stuff, all the categories that they're collecting.

And on the 'Net, I heard this, or I saw this referred to as the "nutrition label," meaning it's sort of like the standardized list of ingredients which all entities are now required to put on a can of something, where all the ingredients they contain listed in the order of how much of that they have, from most to least. And in many cases, of course, how much percentages of this and so forth. Anyway, that's what Apple has created, you know, a standardized means of sharing with users what the apps in the App Store are collecting, what information, what privacy and tracking-related stuff they're collecting.

So anyway, I thought this would be a useful preamble for our discussion of Google's planned FLoC, their Federated Learning of Cohorts. And we should note that DuckDuckGo's comparison is unfair. They are not offering 15GB of free, fast, and hyper-robust cloud storage. Nor do they provide the number one by far most popular free email service in the world. Gmail has 1.5 billion with a "b" users; whereas Outlook is in the number two spot with 400 million, and Yahoo! Mail, of all things, somehow holds onto the number three position with 200 million.

And when I look at the amount of spam Gmail detects and eliminates for me hourly, even though it serves as my catchall throwaway email account, they're doing a phenomenal job for me. Not to mention that I still prefer Google's search, and that I get Google docs and spreadsheets and so much more for free. Or if not exactly technically free, at least without me needing to transfer any of my cash, which I'd much prefer to give to Starbucks. And this is obviously a bargain that I'm not alone in being quite happy with. I really don't give it a second thought, nor does most of the world.

But as we also know, there is also a creepy side to this. For many of us, just the idea that we're being tracked and profiled even if it's against our will and wishes, against all of our efforts to say no is enough to give us pause. So first of all, what is Google's big reveal that the DuckDuckGo people got themselves all lathered up over?

This transparency is all being driven by Apple. And what recently happened is that, after months, I guess DuckDuckGo was saying three months, following Apple's December 2020 announcement of its App Store policy changes, Google's finally updated its App Store apps to bring them into compliance. I'll fill in some background about this in a minute. But the most interesting data point to me was that the forthcoming iOS v14.5, with iOS 14.5, all apps will be required to explicitly request and receive their users' informed consent before they will be allowed to use the device's Apple-provided advertising identifier, known as the IDFA, the ID for Advertisers, which is part of a new framework which Apple calls ATT, which stands for App Tracking Transparency.

So here's the data point: An analysis by the mobile advertising firm AppsFlyer found that, once several third-party developers had integrated Apple's ATT system into their apps, thus making clear to those apps' users what was going on and requesting permission to share their anonymous identity with other Internet services in other words "tracking," while steering well clear of that term, they didn't call it that, but everyone knows fully 99% of users chose not to give those apps that permission which the apps were requesting.

In his speech, which was delivered on January 28th during the Computers, Privacy and Data Protection conference, Tim Cook, who as we know of course is Apple's CEO, said: "Technology does not need vast troves of personal data, stitched together across dozens of websites and apps, in order to succeed. Advertising existed and thrived for decades without it. If a business is built on misleading users, on data exploitation, on choices that are no choices at all, then it does not deserve our praise. It deserves reform."

Okay. So Apple, who sells hardware and privacy, is tightening the screws on those who adamantly insist that tracking and profiling are worth it; that it needs to be allowed to happen; that it's unfortunate, when users are informed and given a choice, they decline to be profiled and tracked, but that needs to be done anyway.

So clearly this is going to become a fraught issue. Just last Wednesday, France's competition regulator rejected calls from advertising companies and publishers who wanted them to block Apple's ATT on grounds of antitrust, stating that Apple's ATT privacy initiative "does not appear to reflect an abuse of a dominant position on the part of Apple," though the regulator did say that it would continue to investigate the changes to ensure that Apple does not apply less restrictive rules for its own apps. In other words, if Apple's going to do this, then it needs to play by its own rules. And of course that's, okay, Apple, because Apple is selling privacy as one of their products, or a clear feature of their products and platform, we don't expect that to be a problem for them.

So where there's a will to track, and there's no lack of enabling technology and innovation, there will always be a way. The Financial Times recently reported that the Chinese Advertising Association (CAA) has developed an identifier it calls the China Anonymization ID (CAID) that's aimed at bypassing Apple's new privacy rules, and that is to bypass the need for or use of the IDFA, this ID for Advertisers. The use of China's nascent CAID would enable companies to continue tracking users without having to rely on Apple.

Leo: So wait a minute. It's not a China Anonymizing ID. It's a China Deanonymizing ID.

Steve: This is China tracking. The China tracker.

Leo: That's hysterical.

Steve: I know. And of course it's not going to ask for users' permission.

Leo: No.

Steve: Because users say no.

Leo: Yeah.

Steve: If you ask people, do you want to be tracked, uh, gee, hmm, let me think. So get this, Leo. The Chinese advertising technology firm with the not-so-subtle name TrackingIO said that CAID, this C-A-I-D, "has the characteristics of anonymity and decentralization, does not collect private data, only transmits the encrypted result, and the encrypted result is irreversible" - maybe that means hashed. Anyway, "...which can effectively protect the privacy and data security of the end user." They said: "The decentralized design allows developers to be more flexible to meet business needs." Okay. Whatever that means. You know, it's Chinese translated into English. Actually it's a pretty good translation.

They added: "Because CAID does not depend on Apple IDFA and can generate device identification independently of IDFA, it can be used as an alternative to device identification in iOS 14 and form a supplementary solution when IDFA is not available." Right. So although CAID is not yet formally implemented, it's believed to be under testing by some of China's largest tech companies who think it's a pretty good idea. And that includes ByteDance and Tencent and several foreign advertising companies that have already applied on behalf of their Chinese divisions.

And following these reports that companies are readying workarounds in an effort to bypass Apple's forthcoming notification and consent requirements on tracking, Apple has sent cease-and-desist letters to two Chinese app developers known to be testing CAID. The email from Apple includes the language: "We found that your app collects user and device information to create a unique identifier for the user's device." And it went on to warn the developer to update the app to comply with App Store rules within 14 days or risk its removal from the App Store.

So does a solution exist, or can a solution be created, to provide advertisers with the information they crave about the apparent interests of web and app users under a model that learns of those interests without, in any way, tracking them? Google says yes. The EFF, they're not sure that even that is okay. So What the FLoC?

Leo: And now Steve will explain FLoC. I'm dying to hear it.

Steve: So on Wednesday at the beginning of the month, on the third, Dave Temkin, who's Google's Director of Product Management, Ads Privacy, and Trust, posted a statement about Google's post-third-party cookie tracking plans titled "Charting a course toward a more privacy-first web."

Now, perhaps David is a bit biased because his posting begins right off the bat with an assertion that I'm not certain holds up. He starts: "It's difficult to conceive of the Internet we know today - with information on every topic, in every language, at the fingertips of billions of people - without advertising as its economic foundation." And I would certainly agree that advertising has fueled a lot. And in fact advertising does fuel a lot. You know, Leo, it fuels this podcast network.

Leo: Yeah, yeah.

Steve: It's arguably the reason that TWiT is still here and going strong.

Leo: Exactly, yeah.

Steve: After 15-plus years. And we know that it certainly fuels Google. As I stated earlier, I'm a happy recipient of a ton of free Google stuff that I make great use of, apparently in return for allowing Google to track and profile me. But is there a better way to accomplish the same task?

Well, let's ask David. He continues. He says: "As our industry has strived to deliver relevant ads to consumers across the web, it has created a proliferation of individual user data across thousands of companies, typically gathered through third-party cookies. This has led to an erosion of trust. In fact, 72% of people feel that almost all of what they do online is being tracked by advertisers, technology firms, or other companies; and 81% say that the potential risks they face because of data collection outweigh the benefits, according to a study by Pew Research Center. If digital advertising doesn't evolve," he writes, "to address the growing concerns people have about their privacy and how their personal identity is being used, we risk the future of the free and open web."

And I'll just interject here that once again we see the all-too-human characteristic that it's often not until someone has a solution to a perceived problem that they are fully willing to acknowledge that the problem exists in the first place. So now that Google has FLoC, oh, look, third-party cookie tracking bad.

Okay. Anyway, he says: "That's why last year Google announced its intent to remove support for third-party cookies, and why we've been working with the broader industry on the Privacy Sandbox to build innovations that protect anonymity while still delivering results for advertisers and publishers. Even so," he says, "we continue to get questions" - questions and questions - "about whether Google will join others in the ad tech industry who plan to replace third-party cookies with alternative user-level identifiers. Today, we're making explicit that once third-party cookies are phased out, we will not build alternative identifiers to track individuals as they browse across the web, nor will we use them in our products." What? What will Google do?

Leo: I'm sure they'll find a way.

Steve: I think they're not going to lose track of us, yes. And notice that, coincidentally, we've recently been talking about third-party cookie phase-out. Those Firefox cookie same-site sequestration changes are all about phasing out the trackability of third-party cookies. So that handwriting really does seem to be on the wall.

So David continues: "We realize this means other providers may offer a level of user identity for ad tracking across the web that we will not, like PII (Personally Identifiable Information) graphs based on people's email addresses. We don't believe these solutions will meet rising consumer expectations for privacy, nor will they stand up to rapidly evolving regulatory restrictions, and therefore aren't a sustainable long-term investment. Instead" - and here it comes - "our web products will be powered by privacy-preserving APIs which prevent individual tracking while still delivering results for advertisers and publishers." How are they going to do that?

Well, he says: "People shouldn't have to accept being tracked across the web in order to get the benefits of relevant advertising. And advertisers don't need to track individual consumers across the web to get the performance benefits of digital advertising. Advances in aggregation, anonymization, on-device processing, and other privacy-preserving technologies offer a clear path" - well, it's clear now, apparently - "to replacing individual identifiers. In fact, our latest tests of FLoC show one way to effectively take third-party cookies out of the advertising equation and instead hide individuals within large crowds of people sharing common interests.

"Chrome intends to make FLoC-based cohorts available for public testing through origin trials with its next release this month" - and this was written this month - "and we expect to begin" - so they'd better get on it, we're at the 23rd here - "and we expect to begin testing FLoC-based cohorts with advertisers in Google Ads in the second calendar quarter this year. Chrome also will offer the first iteration of new user controls in April and will expand on these controls in future releases, as more proposals reach the origin trial stage, and they receive more feedback from end users and the industry."

He finishes: "This points to a future where there is no need to sacrifice relevant advertising and monetization in order to deliver a private and secure experience." And finally: "Keeping the Internet open and accessible for everyone requires all of us to do more to protect privacy; and that means an end to not only third-party cookies, but also any technology used for tracking individual people as they browse the web. We remain committed," they say, "to preserving a vibrant and open ecosystem where people can access a broad range of ad-supported content, with confidence that their privacy choices are respected. We look forward to working with others in the industry on the path forward."

Okay. So before we go any further, this raises an interesting philosophical question. How do we feel about non-tracking-based aggregation of our interests? As individuals interacting with the Internet, do we actually demand full and absolute privacy, meaning that we are a completely opaque entity to every site we visit? Or is it all right for who we are to be known as an anonymous cloud of likes, desires, and interests? And as I thought about that, it seems to me that I have no problem with people who I know and implicitly trust knowing a lot about who I am. But I feel much less sanguine about having totally unknown and unknowable strangers knowing much about me without my giving explicit permission.

Okay. So given the title of the EFF's reaction to Google's FLoC, they apparently feel even more strongly. And I should note that the EFF does not like anything ever. The only thing I can recall them ever liking was Let's Encrypt. Oh, they loved Let's Encrypt. Everything else, no. The EFF titled their reaction to Google's FLoC, they said: "Google's FLoC Is a Terrible Idea." And they apparently wanted to be certain that no one came away from their posting feeling in any way unsure of any of the details. So their posting is endless. When I went to it, the scroll thumb shrunk down to like a little itty-bitty square. So I'm going to share some of what they posted with liberal interjections.

They said: "The third-party cookie" - this is the EFF, you know, the Electronic Freedom Foundation. "The third-party cookie is dying, and Google is trying to create its replacement." Okay. Eh. But we understand where they're coming from. They said: "No one should mourn the death of the cookie as we know it. For more than two decades, the third-party cookie has been the linchpin in a shadowy, seedy, multi-billion-dollar, advertising surveillance industry on the web. Phasing out tracking cookies and other persistent third-party identifiers is long overdue," writes the EFF. "However, as the foundations shift beneath the advertising industry, its biggest players are determined to land on their feet.

"Google is leading the charge to replace third-party cookies with a new suite of technologies to target ads on the web. And some of its proposals show that it hasn't learned the right lessons from the ongoing backlash to the surveillance business model. This post will focus on one of those proposals, Federated Learning of Cohorts, which is perhaps the most ambitious," they said, "and potentially the most harmful. FLoC is meant to be a new way to make your browser do the profiling that third-party trackers used to do themselves, in this case boiling down your recent browsing activity into a behavioral label and then sharing it with websites and advertisers. The technology will avoid the privacy risks of third-party cookies, but it will create new ones in the process. It may also exacerbate many of the worst non-privacy problems with behavioral ads, including discrimination and predatory targeting." And we're going to talk about that. We'll get there.

They said: "Google's pitch to privacy advocates is that a world with FLoC and other elements of the Privacy Sandbox will be better than the world we have today, where data brokers and ad tech giants track and profile with impunity. But that framing is based on a false premise that we have to choose between 'old tracking' and 'new tracking.' It's not either/or," they allege. "Instead of reinventing the tracking wheel, we should imagine a better world without the myriad problems of targeted ads."

Ah. So there's a clear data point. The EFF takes the position that any and all targeting will inherently be fraught with targeting-related problems independent of tracking. So this attitude unfortunately strongly biases their language since non-tracking is not "new tracking." It really is non-tracking. Anyway, we'll understand this.

They said: "We stand at a fork in the road. Behind us is the era of the third-party cookie, perhaps the web's biggest mistake." And of course we all know my often lamented feelings about third-party cookie tracking. It was never meant to be. But as technologists we allowed it to happen. So maybe we are finally going to get rid of it.

They said: "Ahead of us are two possible futures. In one, users get to decide what information to share with each site they choose to interact with." They said: "No one needs to worry that their past browsing will be held against them, or leveraged to manipulate them, when they open a tab." Okay, now, wait a minute. Users get to decide what information to share with each site they choose to interact with? How's that going to work? Like it's been such a wonderful improvement to our lives that we now need to give every site we visit explicit permission about whether or not to use cookies. What a nightmare. But anyway.

The EFF continues: "In the other case, each user's behavior follows them from site to site as a label, inscrutable at a glance, but rich with meaning to those in the know." They said: "Their recent history" - meaning users' recent history - "distilled into a few bits, is 'democratized' and shared with dozens of nameless actors that take part in the service of each web page. Users begin every interaction with a confession: 'Here's what I've been up to this week. Please treat me accordingly.' Users and advocates," they say, "must reject FLoC and other misguided attempts to reinvent behavioral targeting. We implore Google to abandon FLoC and redirect its efforts towards building a truly user-friendly web."

Okay. So with that introduction, to offer a bit of background, which is interesting for reasons we'll see, they continue: "In 2019," they say, "Google presented the Privacy Sandbox, its vision for the future of privacy on the web. At the center of the project is a suite of cookieless protocols designed to satisfy the myriad use cases that third-party cookies currently provide to advertisers. Google took its proposals to the W3C, the standards-making body for the web, where they have primarily been discussed in the Web Advertising Business Group, a body made up mostly of ad tech vendors. In the intervening months, Google and other advertisers have proposed dozens of bird-themed technical standards: PIGIN, TURTLEDOVE, SPARROW, SWAN, SPURFOWL, PELICAN, PARROT. The list goes on."

Leo: Okay.

Steve: "Seriously. Seriously," they said. "Each of the 'bird' proposals is designed to perform one of the functions in the targeted advertising ecosystem that is currently performed by cookies." And then it hit me. Birds.

Leo: Birds.

Steve: That's why, Leo, and you already knew, that's why this abbreviation is so godawful. They had to...

Leo: It's a retronym.

Steve: Oh, they had to reverse engineer something for flock, which it's a flock of birds. So we get the painfully horrible Federated Learning of Cohorts. At least we now know where it came from. Let's hope it's a working title. On the other hand, that's what McIntosh was, and that wasn't so bad. But FLoC, ugh. Anyway, maybe we'll get used to it.

They said: "FLoC is designed to help advertisers perform behavioral targeting without third-party cookies." And I would strengthen also, without tracking. It actually is non-tracking. They said: "A browser with FLoC enabled would collect information about its user's browsing habits, then use that information to assign its user to a 'cohort' or group. Users with similar browsing habits, for some definition of 'similar,' would be grouped into the same cohort. Each user's browser will share a cohort ID, indicating which group they belong to, with websites and advertisers. According to the proposal, at least a few thousand users" - and actually it's thousands is what Google says because I read the spec - "should belong to each cohort," although they say, though, that's not a guarantee.

Okay. So first of all, the small size of that group that they're alleging surprises and concerns me. I assumed that cohorts should be much larger groupings. But the motivation is clearly to keep them both highly targeted, yet you don't want them to be too small because you still want anonymity. Anyway, they said: "If that sounds dense, think of it this way: Your FLoC ID will be like a succinct summary of your recent activity on the web. Google's proof of concept used the domains of the sites that each user visited as the basis for grouping people together. It then used an algorithm called SimHash to create groups."

And I'll interject. SimHash is short for Similarity Hash. It's an algorithm that Google has deep experience with, since it's used by the Google web spider to estimate the similarity of non-identical web pages, which it encounters as it's spidering the web.

They said: "SimHash can be computed locally on each user's machine, so there's no need for a central server to collect behavioral data," which I think is cool. "However," they said, "a central administrator could have a role in enforcing privacy guarantees. In order to prevent any cohort from being too small - in other words, too identifying - Google proposes that a central actor could count the number of users assigned each cohort. If any are too small, they would be combined with other, similar cohorts until enough users were represented in each one."

Then the EFF provides some useful and interesting detail. According to the proposal, which by the way is public on GitHub, most of the specifics are still up in the air. The draft specification states that a user's cohort ID will be available via JavaScript. But it's unclear whether there will be any restrictions on who can access it. And I would presume there will be none. Or whether the ID will be shared in any other ways, like in a header, for example. FLoC could perform clustering based on URLs or page content instead of domains. It could also use a federated learning-based system, as the name FLoC implies, to generate the groups instead of SimHash.

It's also unclear exactly how many possible cohorts there will be. Google's experiment used 8-bit - and I almost fell off my chair, 8-bit, because it's so small - cohort identifiers, meaning that there were only 256 possible cohorts. That would be wonderful, but that's never going to happen. In practice that number could be, they said, much larger. The documentation suggests a 16-bit cohort ID comprising four hexadecimal characters. Of course the more cohorts there are, the more specific they will be. Longer cohort IDs mean that advertisers learn more about each user's interests and have an easier time, the EFF says, fingerprinting them.

Leo: I disagree. That's completely illogical. But okay.

Steve: Right. One thing that is specified is duration. FLoC cohorts will be recalculated on a weekly basis, each time using data from the previous week's browsing. So that's another nice thing is that this creates a rolling identifier as the things you do differ, and your browser notices that. It updates weekly and moves you into a new cohort ID. They said: "This makes FLoC cohorts less useful as long-term identifiers." Right. "But it also makes them more potent measures of how users behave over time." Well, okay. On the other hand, who cares, maybe?

So anyway, so far, despite EFF's valiant efforts, I'm not convinced that this is a bad thing. It's bad, of course, if you're absolutely unwilling to be targeted in any way. But for anyone who's willing to make a tradeoff, this seems like the one to make.

Leo: I'd agree with you.

Steve: Yeah.

Leo: I'm not going to assume that any tracking is bad. I mean, if all it is is so that you have ads that are suiting your interests, and if you've been anonymized sufficiently. That's why a higher number would be...

Steve: A larger number.

Leo: Larger number of buckets would be bad.

Steve: If I were one in 64K of all users in the world, bring it on.

Leo: Who cares? And the thing is there's a disincentive for them to make it too granular because you don't want groups to be too small, either. I would submit that if there's an 8-bit, 16-bit, 32-bit, 24-bit, would really be a function of how big advertisers want those groups to be. Do they want 100 people in the cohort? A thousand? Ten thousand? And I think the demands of advertisers probably vary. But Google will optimize that for the right-size bucket. Not for the granularity, but for the size of the bucket, if you make it too granular. But it sounds like what they're going to do is like a red, green, blue thing. So they're going to have, you know, there'll be several axes. So on the income axis, on the age axis, maybe the interest axis.

Steve: Well, I don't think so.

Leo: No?

Steve: From what they're saying, it is based on interests reflected by the history of the domains you visit.

Leo: I see. So they're not going to collect demographic information at all.

Steve: Yeah, they're not going to collect demographics.

Leo: That's interesting, yeah.

Steve: Yeah.

Leo: Advertisers want demographics.

Steve: I know.

Leo: But more than that, they want - they certainly want interest. I mean, look. If you could tell an advertiser this guy's going to buy a car in the next three weeks, for a certain group of advertisers, that's all they care about. Maybe they care about your budget.

Steve: Well, but you might be able - I don't know, who knows if you're able to infer that from places you visit.

Leo: Other information, yeah.

Steve: So now they present some negatives which are interesting. They said: "The first issue is fingerprinting. Browser fingerprinting," they said - we all know this, but it's brief - "is the practice of gathering many discrete pieces of information from a user's browser to create a unique, stable identifier for that browser. EFF's Cover Your Tracks project demonstrates how the process works. In a nutshell, the more ways your browser looks or acts different from others', the easier it is to fingerprint.

"Google has promised that the vast majority of FLoC cohorts will comprise thousands of users each, so a cohort ID should not alone distinguish you from a few thousand other people. However, it still gives fingerprinters a massive head start."

Leo: Right.

Steve: "If a tracker starts with your FLoC cohort, it only has to distinguish your browser from a few thousand others, rather than a few hundred million others." And I would counter that by observing that it changes weekly. And who knows? I would imagine that the changes are asynchronous globally so that my cohort is going to suddenly be different by some measure at some point and not notify anybody when that happens.

Leo: Your real concern is deanonymization; right? Is that the real concern? I mean, that's what fingerprinting might do is identify you as Steve Gibson.

Steve: Kind of. So here's more. They note: "Fingerprinting is notoriously difficult to stop. Browsers like Safari and Tor have engaged in years-long wars of attrition against trackers, sacrificing large swaths of their own feature sets" - and we've documented that on this podcast - "in order to reduce fingerprinting attack surfaces. Fingerprinting mitigation generally involves trimming away or restricting unnecessary sources of entropy, which is what FLoC is. Google," they're saying, "should not create new fingerprinting risks until it's figured out how to deal with the existing ones."

And then they highlight a new problem created by this technology which they call "cross-context exposure." They said: "The second problem is less easily explained away. The technology will share new personal data with trackers who can already identify users. For FLoC to be useful to advertisers, a user's cohort will necessarily reveal information about their behavior." Right? That's what it is.

The project's GitHub page addresses this upfront. GitHub page says: "This API democratizes access to some information about an individual's general browsing history, and thus general interests, to any site that opts into it. Sites that know a person's PII" - their Personally Identifiable Information - "for example, when people sign in using their email address, could record and reveal their cohort. This means that information about an individual's interests may eventually become public." Which is interesting.

The EFF noted: "As described above, FLoC cohorts should not work as identifiers by themselves. However, any company able to identify a user in other ways, say by offering 'log in with Google' services to sites around the Internet, will be able to tie the information it learns from FLoC to the user's profile." And they said: "Two categories of information may be exposed this way. First, specific information about browsing history. Trackers may be able to reverse engineer the cohort-assigned algorithm to determine that any user who belongs to a specific cohort probably or definitely visited specific sites.

And second, general information about demographics or interests." Well, duh. That's what the cohort ID is; right?

But they said: "Observers may learn that, in general, members of a specific cohort are substantially likely to be a specific type of person. For example, a particular cohort may over-represent users who are young, female, and black; another cohort, middle-aged Republican voters; a third, LGBTQ youth. This means every site you visit will have a good idea about what kind of person you are on first contact, without having to do the work of tracking you across the web, or buying that service from an aggregator. Moreover, as your FLoC cohort will update over time, sites that can identify you in other ways" - like because you log into them explicitly - "will also be able to track how your browsing changes. Remember," they wrote, "a FLoC cohort is nothing more and nothing less than a summary of your recent browsing activity."

And here's their key point: "You should have a right to present different aspects of your identity in different contexts. If you visit a site for medical information, you might trust it with your information about your health, but there's no reason for it to know what your politics are. Likewise, if you visit a retail website, it shouldn't need to know whether you've recently read up on the treatment for depression. FLoC erodes this separation of contexts, and instead represents the same behavioral summary to everyone you interact with." Now, I would argue that they're reversing disadvantages of it, you know, picking it apart. But they make a valid point.

And then, tying back to the beginning about the inherent problems associated with any type of targeted advertising, they wrap up, or I'll wrap up what they're sharing by saying the EFF makes some additional disturbing observations. They said: "FLoC is designed to prevent a very specific threat, the kind of individualized profiling that is enabled by cross-context identifiers today. The goal of FLoC and other proposals is to avoid letting trackers access specific pieces of information that they can tie to specific people. As we've shown, FLoC may actually help trackers in many contexts. But even if Google is able to iterate on its design and prevent these risks, the harms of targeted advertising are not limited to violations of privacy. FLoC's core objective is at odds with civil liberties."

They say: "The power to target is the power to discriminate." And this sounds like something you'll want to talk with Jeff about tomorrow, Leo. They said: "By definition, targeted ads allow advertisers to reach some kinds of people while excluding others. A targeting system may be used to decide who gets to see job postings or loan offers just as easily as to advertise shoes. Over the years, the machinery of targeted advertising has frequently been used for exploitation, discrimination, and harm." And actually in their posting they have links for examples of all this.

They said: "The ability to target people based on ethnicity, religion, gender, age, or ability shows discriminatory ads for jobs, housing, and credit. Targeting based on credit history, or characteristics systematically associated with it, enables predatory ads for high-interest loans. Targeting based on demographics, location, and political affiliation helps purveyors of politically motivated disinformation and voter suppression. All kinds of behavioral targeting increase the risk of convincing scams."

So I'm reminded of this when we've talked about it before. We talked about how billboards along the highway don't know, at least yet, who's driving by. Nor do placards posted in store windows. We're all treated uniformly. Television advertisers have always been able to select the TV programs on which they will appear. The advertisers can presume the demographic of any program's audience, but they have no feedback beyond that. And it makes one wonder whether web and web advertisers wouldn't be satisfied with choosing which websites to have hosting their ads, based on the demographics of the visitors to those sites, rather than having ads able to chase their targets across the web.

Leo: Well, they're only willing to take it if that's all they can get.

Steve: Right, right.

Leo: You know, and it's been our experience, that's how podcast ads worked. But more and more, advertisers are demanding more information about listeners. And as they start to get it from platforms like Apple and Spotify, it makes it difficult for people who don't want to somehow track listeners. You just don't get ads.

Steve: Well, and you know, I had a creepy thought, Leo, when I was thinking about this. It just occurred to me, I wonder whether the inserts that are now sometimes being placed into podcasts, if everyone gets the same one.

Leo: Oh, no, no, no. Because they're by IP address. So it's geographic. It's not - I don't think they have any demographic information, but they have rough geography from the IP addresses.

Steve: Interesting.

Leo: But that's not at all unusual. An advertiser who sells cars in Northern California doesn't want to buy a podcast that's international. They want to buy Northern California listeners. And I don't think that's unreasonable, to be honest with you.

Steve: You know that I'm a big user of YouTube TV. And they sometimes have these weird blank times where they just talk, and it's like nothing is showing. And I said to Lorrie, I said, "I wonder if they know it's us, and the ad that they might have inserted there like, you know, is not meant for me."

Leo: No. They know it's web. So the television ad buyer buys MSNBC, says specifically "not for streaming." So when that ad comes up, and you're watching YouTube TV, you won't get it.

Steve: Okay.

Leo: And that's very common because they don't want to buy those numbers.

Steve: Why not? I mean, it's like TV watchers are - is there any other way to watch MSNBC other than streaming?

Leo: I can't speak for advertisers' logic. Sometimes it does seem illogical. But they're well within their rights to say that. And by the way, billboards are targeted demographically. For a long time there's a lot of evidence that menthol cigarettes billboards only showed up in black neighborhoods.

Steve: And so like in large syndicated radio shows - well, in fact I know that because I see ads for some car dealer around the corner.

Leo: Local.

Steve: Obviously they're not giving that to people in New York.

Leo: Right. So your cable company is selling those, Comcast and Cox and the others. When you're watching a channel on cable, there are local avails that the cable company can sell. And that's why you see local ads there. You know immediately when we switch to the local ads, the crappy ads.

Steve: And they're really cheesy, yeah. Oh, my goodness.

Leo: You know immediately.

Steve: Where did they get this guy?

Leo: This has gone on forever. I mean, the inserts in your newspaper, those vary depending on neighborhood. This has always happened. The problem is that as soon as digital technology came along, better and better ways of doing this appear.

Steve: A much sharper knife.

Leo: A sharper knife, thanks to Google and Facebook. The thing that puzzles me about FLoC, you know, Facebook knows a lot about you, more than anybody else. That's why their ads are so efficient. But they don't give that information to anybody. That's their secret sauce. If you're an advertiser, you can't go to Facebook and say, well, you know, tell me who's looking at this page. You buy a demographic. Facebook keeps that closely held. That information...

Steve: And have you been seeing their ads? They're, like, advertising advertising.

Leo: Yeah, well, we'll be doing that too, soon, because advertising has slowed down a little bit in COVID. So I'm surprised Google - I think I have to read this more carefully. I don't know if EFF has misunderstood FLoC. Why would Google give this information to a website? I don't think they would. Google would sell that website based on that information to an advertiser. But they don't want anyone to know what group you're in.

Steve: Ah. So maybe the encoding is proprietary.

Leo: Absolutely. Unless I'm misunderstanding it.

Steve: So you get the code from the person visiting, and then you've got to pay Google to find out what that means.

Leo: Not even that. Look, when you're buying a Google ad, you're buying it from Google. They know. So you tell Google, look, I want to advertise only on sites for males 25-54. And Google knows who that is. You're in that FLoC. So when they see you show up, not the website, Google knows. The website doesn't get that information. That's proprietary. Google spends a lot of money to get that information. They're not going to give it away.

Steve: Well, in that case, then, that's a big question because these guys are clearly saying, and the GitHub page makes it, I mean, they even show some sample JavaScript where any site could put some JavaScript in the browser and acquire the FLoC cohort ID.

Leo: That's interesting. I don't know why Google would allow that. Because that's like Facebook saying, yeah, I mean, that's what they're selling. That information is the gold. So maybe this is because it's an open standard?

Steve: Google is selling the appearance of the ads on the page; right?

Leo: Yeah, and they sell it based on the advertiser saying I want this particular cohort.

Steve: Well, you're right. So Google does its own tracking and builds its own profile.

Leo: Yes.

Steve: Which is how, through DoubleClick, which is how they perform their...

Leo: Precisely. That's how they do it now. Now, the third-party tracking cookies is interesting because when you have a "like" button from Facebook on every page, Facebook is gathering information about every page you go to. And in fact you're visiting Facebook all the time because that's, you know. So I understand the concern about that. Although, again, mostly this is used just to target advertising. And I agree what the EFF says about how that could be misused. But it's still only advertising.

Steve: How about any advertising targeting is misused, yeah.

Leo: Yeah. But, I mean...

Steve: That's the nature of it.

Leo: It's an ad. It's crap. It doesn't mean you have to pay attention to it. In fact, the larger story is people are just blocking ads in general. And that's, see, I think sometimes that gets conflated with tracking. They're two separate things. You and I block ads because they can carry malware, because they kill bandwidth, because they're annoying. There's too many of them. There's all sorts of good reasons to say I don't want to see ads.

Steve: And, oh, gee, they're not relevant to me. What do you know?

Leo: That's part of the problem; right? I don't want to see those ads. They're not for me.

Steve: It doesn't appear to actually work.

Leo: Yeah. Yeah. Well, that's true.

Steve: It's like all this machination, and it's like, wait, you know, I'm not buying baby diapers.

Leo: But put yourself in the position of the person buying an ad.

Steve: Or adult diapers yet.

Leo: Yes. I mean, we buy ads. We buy Google ads. TWiT does. If you put yourself in the position of an advertiser, they don't want an ad to show to anyone who's not interested in their product. They're trying to find people who are...

Steve: Because it's expensive for them.

Leo: It's a waste of money.

Steve: They have to pay for that, yes.

Leo: So efficient ads are targeted ads. That's what advertisers want, quite reasonably. I think end users, if they were properly targeted, would prefer ads for - you don't want to see an ad for diapers if you don't wear them. You don't want, I mean, that's almost insulting.

Steve: Actually, there are a lot of ads, come to think of it, these days that I'm seeing that I'm thinking, oh, no.

Leo: We're insulted by every ad we see. If you watch cable news, it's nothing but drugs you don't need.

Steve: Exactly.

Leo: Nonstop. But those advertisers...

Steve: Let's try some Vyvanse. Ask your doctor if that's right for you. And then...

Leo: Most advertisers would far prefer to advertise to people...

Steve: And then there's the stuff that falls off if you take it. It's like, oh.

Leo: I know. It's kind of the opposite of an ad, isn't it. It's like, I'm not taking that. I don't know what's wrong with that person. Why are they taking that? But honestly, the makers of Vyvanse would far prefer to sell it to people who are buyers of Vyvanse, or potential buyers. Not us. So I have very mixed feelings about all of this. It is, frankly, just speaking from my position as somebody who sells ads, that's all - advertisers push for that very hard. And it's really a hard thing to say. You know, you'll lose ads, we lose ads all the time.

Steve: We'll have to figure out who has access to the information.

Leo: That's interesting.

Steve: But it sure does sound like it's better than tracking.

Leo: Yeah. I think it's at least somewhat anonymous.

Steve: Yeah, and the idea that Chrome would just shut down third-party cookies. Of course then we have the whole CNAME problem again. It doesn't remove that, which is now we know 10% of the top 10,000.

Leo: Here's my complaint, really, is Google saying, okay, no third-party cookies. We'll do something only we can do, and good luck to the rest of you. To me, what really this is is saying, yeah, well, we can...

Steve: And so maybe that argues for them making it public, for them allowing the cohort to be public.

Leo: I really have to read more about this.

Steve: And that means you have to publish what the ID means. Otherwise nobody can make any sense of it.

Leo: Right. But why would they do that?

Steve: Well, and it also means that there will be sites that show you the meaning of your own browser's cohort.

Leo: Right, right.

Steve: I mean, you know...

Leo: Who are you? Here's what I know about you.

Steve: I won't have to do it. Troy Hunt will do it for us.

Leo: Yeah.

Steve: But, you know, you'll be able to go there and see, oh, look at the Venn diagram that I'm in.

Leo: You see, that's why I don't think this is going to work that way. I read this EFF article when it came out, and I was puzzled by some of it. So I don't know.

Steve: Well, I have a feeling this will not be the last "What the FLoC" we end up discussing.

Leo: Right.

Steve: And maybe we'll know more about...

Leo: What the FLoC.

Steve: What the FLoC.

Leo: Aptly named. Thanks, Steve Gibson. His home on the web is GRC.com. That's where of course you'll find SpinRite, the world's best hard drive maintenance and recovery utility. It's a good time to get SpinRite, if you don't already have it, because 6.1 is coming. And if you buy 6.0 now, you'll have a free upgrade, plus you can participate in the development of 6.1. And I think the best news of all that I've heard is that SpinRite is no longer just for spinning drives. It works very well in interesting ways on SSDs. So it's really of great use for anybody with storage. How about that? StorageRite. That's GRC.com.

While you're there, of course, you can get a copy of this show. Steve has a couple of unique versions of this show, a 16Kb version, little lower audio quality, but it's a much smaller file for the bandwidth-impaired. Well, you can tell how much lower. What is that, one-sixth the size. No, it's one-fifth the size, something like that. One-fourth the size, there we go. You can also get transcripts, which are probably a tenth the size because it's text, by Elaine Farris. She writes this all up. That's nice to have, if you read along while you listen or just read. I use it for searching because, if you search, you can find any part of the show, any show, all 811, all at GRC.com, along with 64Kb audio.

We have audio and video at our site, TWiT.tv/sn. You can also - that means you can download any episode from there. You can also get it on YouTube. There's a YouTube channel. All the episodes are up there, video. And of course subscribe in your favorite podcast application. That way you'll get it automatically, the minute it's available. We do the show on a Tuesday afternoon about 1:30 Pacific, 4:30 Eastern, 20:30 UTC. So if you want to watch us do it live, the unedited, unexpurgated version, you can get that at TWiT.tv/live, watching live, chat live at irc.twit.tv after the fact. Steve takes DMs at his Twitter site. You can slide into his DMs at @SGgrc, or leave feedback at GRC.com/feedback.

We have our own forums. Steve has his SpinRite forums. We have our TWiT community forums at www.twit.community. We also have a Mastodon instance. That's the federated Twitter-like Fediverse, and it's really a lot of fun in there. We just passed 1,000 users at TWiT.social. I'm @Leo at TWiT.social. People ask, how can we get Steve in here? I say, "Give me a break, it took me years to get him on Twitter." Slow down. Just wait a little bit. Maybe in a while. Steve has a lot of other things he's working on right now.

That's it for the show. Thank you for being here. We'll see you next week, Steve, on Security Now!.

Steve: Bye.