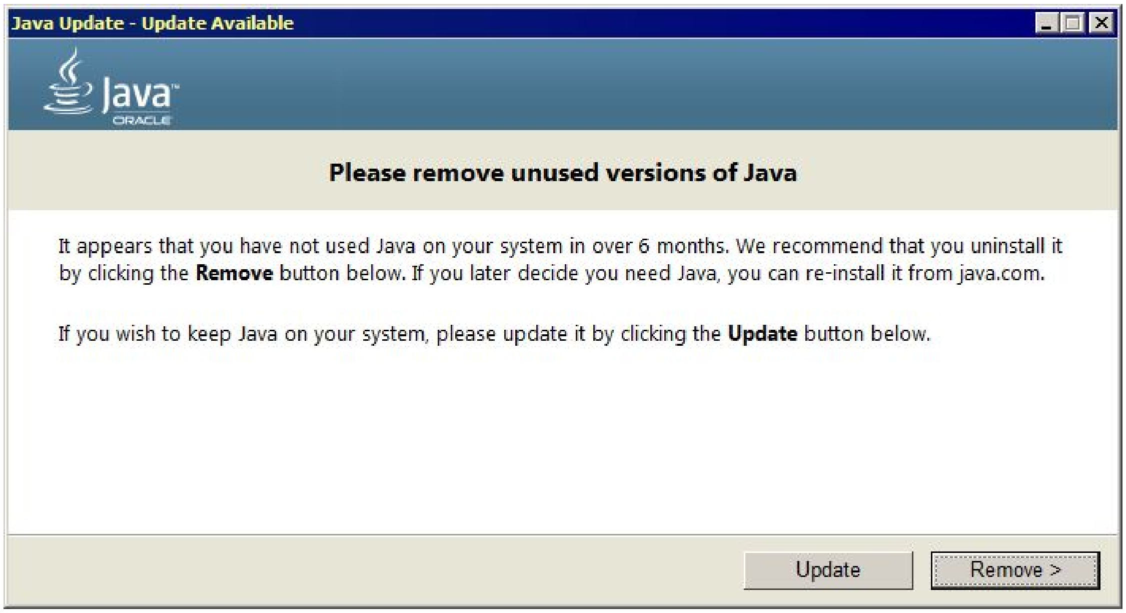
Security Now! #811 - 03-23-21

What the FLoC?

This week on Security Now!

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!

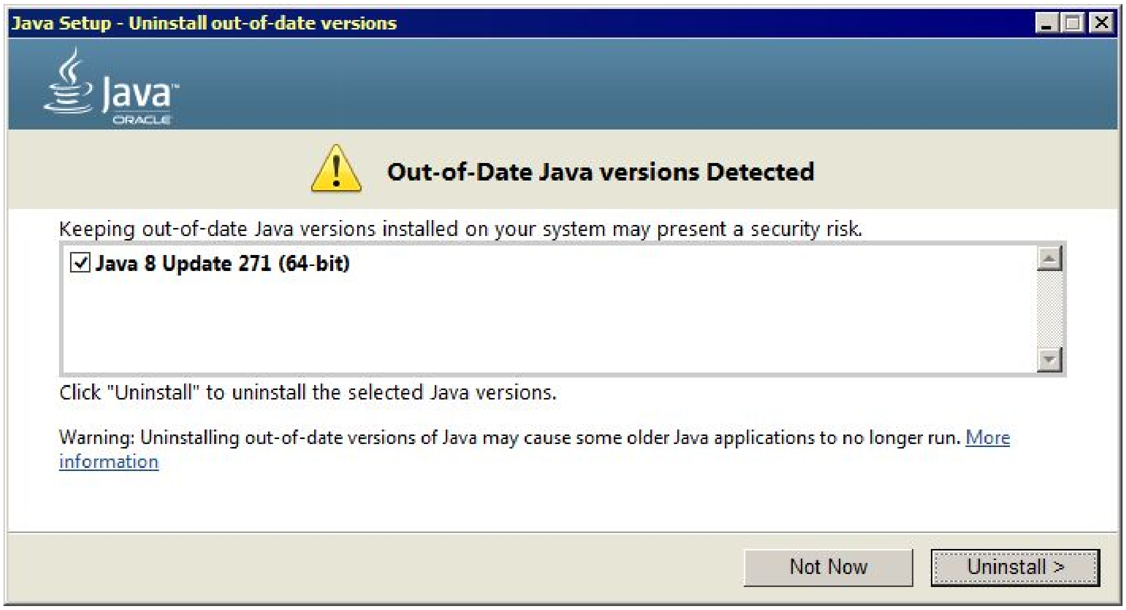


Java Update - Update Available

Please remove unused versions of Java

It appears that you have not used Java on your system in over 6 months. We recommend that you uninstall it by clicking the **Remove** button below. If you later decide you need Java, you can re-install it from java.com.

If you wish to keep Java on your system, please update it by clicking the **Update** button below.



Java Setup - Uninstall out-of-date versions

Out-of-Date Java versions Detected

Keeping out-of-date Java versions installed on your system may present a security risk.

☑ **Java 8 Update 271 (64-bit)**

Click "Uninstall" to uninstall the selected Java versions.

Warning: Uninstalling out-of-date versions of Java may cause some older Java applications to no longer run. More information

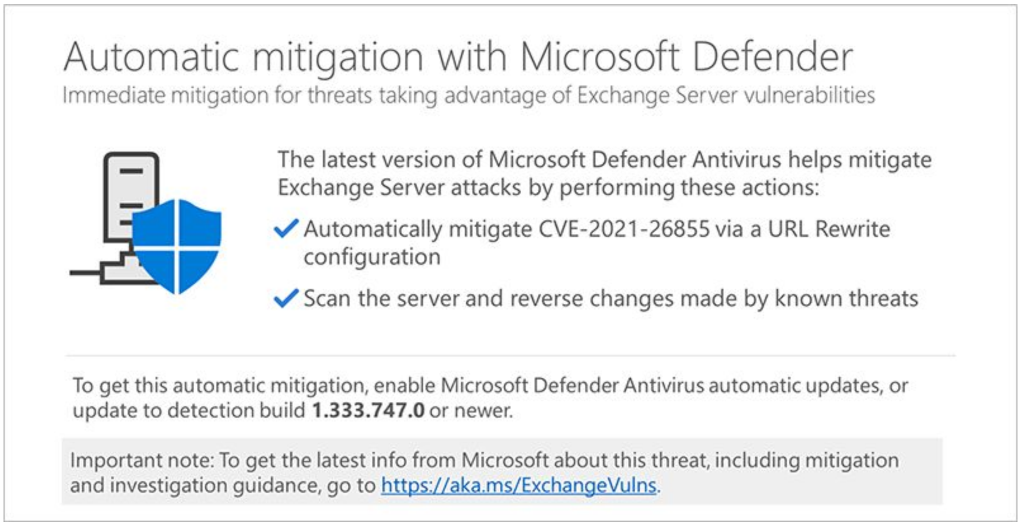
ProxyLogon Followup

https://www.microsoft.com/security/blog/2021/03/18/automatic-on-premises-exchange-server-mitigation-now-in-microsoft-defender-antivirus/

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, 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, 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 will automatically mitigate CVE-2021-26855 on any vulnerable Exchange Server on which it is deployed. Customers do not need to take action beyond ensuring they have installed the latest security intelligence update (build 1.333.747.0 or newer), if they do not already have automatic updates turned on.

I was initially puzzled about the use-case for 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 the posting makes their intent more clear:



Automatic mitigation with Microsoft Defender

Immediate mitigation for threats taking advantage of Exchange Server vulnerabilities

The latest version of Microsoft Defender Antivirus helps mitigate Exchange Server attacks by performing these actions:

✓ Automatically mitigate CVE-2021-26855 via a URL Rewrite configuration

✓ Scan the server and reverse changes made by known threats

To get this automatic mitigation, enable Microsoft Defender Antivirus automatic updates, or update to detection build **1.333.747.0** or newer.

Important note: To get the latest info from Microsoft about this threat, including mitigation and investigation guidance, go to https://aka.ms/ExchangeVulns.

So, it must be that the point here is (a) that it's one more thing they can do so that no one can say they didn't do something that they could have and (b) unlike the monthly patch updates which require an administrator's intervention and permission for a full system reboot, the use of Windows Defender will eventually be automatic. It's unclear to me how often Microsoft OSes check in for Defender updates by default. I couldn't find anything definitive about that online. But my own instance of Defender had run just 90 minutes before I checked. So it's presumably updating itself and scanning things often.

What this means, then, since Microsoft's blurb says “Scan the server and reverse changes made by known threats” is that eventually any Microsoft Exchange Server that hasn't had Windows Defender or its enterprise equivalent deliberately disabled — and I would assume that few would have — will eventually have its most critical ProxyLogon Remote Code Execution vulnerability neutered if not removed completely. And such a Server would also have any intrusion debris that Microsoft is aware of automatically removed. And that could also evolve over time.

I mentioned that the vulnerability would be neutered but not removed. Removal requires the replacement of DLLs that are resident in RAM and invoked by low-level services. They cannot be replaced without a full system reboot and reload. But Microsoft’s announcement mentioned the use of a URL Rewrite configuration. “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 than its normal scan and sequester. It’s tweaking the configuration of Exchange’s IIS web server to add a new URL Rewrite rule to prevent the still-exploitable underlying IIS server from being exploited. And it also goes beyond even that by seeking and reversing known malicious changes.

They’re essentially using Windows Defender, which checks for updates frequently, as a no-boot mitigation for Exchange Server systems that are critically and chronically un-administered. This is a good move on their part. It will allow Exchange Server to at least partially be brought back from oblivion without ANY administration. Of course, and unfortunately, this arrived on March 18th, 16 days after the emergency patches went public and a full week after mass scanning and compromise of all known Exchange Servers was well underway from as many as ten different very serious state sponsored threat actors.

I really REALLY 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 after being notified of a serious vulnerability. The INSTANT the knowledge exists in the world, the race is on.

Black Kingdom

The other interesting bit of ProxyLogon 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.

Over the weekend, our friend Marcus Hutchins of MalwareTechBlog tweeted that another threat actor was compromising Microsoft Exchange servers via the ProxyLogon vulnerabilities to deploy ransomware. Based on the logs his Honeypots were producing, Marcus said that the threat actor was employing the chained Exchange Server vulnerabilities to execute a PowerShell script that downloads the ransomware executable from 'yuuuuu44[.]com', then pushes it out to other computers on the network.

And this is confirmed by submissions to the ransomware identification site ID Ransomware with its first submissions seen on March 18th. Michael Gillespie, the creator of ID Ransomware whom we mentioned last week, told BleepingComputer that his system has seen over 30 unique submissions of the Black Kingdom ransomware campaign, with many submitted directly from mail servers. When encrypting devices, the ransomware encrypts files using random extensions and leaves a ransom note named “decrypt_file.TxT”. Hutchins states that he saw a different ransom note named ReadMe.txt that uses slightly different text.

But in any event, anything bad that can happen, has already happened or is going to happen to many tens of thousands of Exchange Servers.

Browser News

Firefox will be adopting a new privacy-enhancing Referrer Policy
We're currently on Firefox 86. 87 will bring a welcome change...

We've talked about the (misspelled) web browser referer header often. It's supposed to be "Referred" but the original specification and implementations used "Referer" in error... so that's what we have today.

The Referer field has long been controversial because it contains the URL of the page that "referred" its visitor to the resource being requested — another web page, a tracking beacon, or anything that the browser fetches from the referring page. Once upon a time this was 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.

Mozilla has announced that the next release of Firefox will introduce a more privacy-focused default Referrer Policy to protect Firefox users' privacy. The web browser will henceforth automatically trim user-sensitive information like path and query string information accessible from the Referrer URL.

Mozilla's spokespersons said: “Unfortunately, the HTTP Referrer 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.”

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 shows Referer headers containing, among many other tidbits, internal hostnames for government and enterprise entities that most likely should not be public. Yet this mechanism publishes them. And malicious actors might use such information to their advantage.

The first appearance of an explicit Referer Policy appeared in our web browsers between 2016 and 2018. Back then the web was still largely a hybrid of HTTP and HTTPS. So there was a concern that resources accessed over the less secure HTTP should have a restricted view of what was going on on any HTTPS page. So the policy known as ‘no-referrer-when-downgrade’ was enacted. Any query to HTTP would not have any referer header. So less secure resources would receive much less information about the page requesting their asset.

Today, Mozilla considers the ‘no-referrer-when-downgrade’ policy to be a relic of the past web because, as we know, today's web looks much different. We're finally on the 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 Firefox 87, the default Referrer Policy will be set to ‘strict-origin-when-cross-origin’ which will trim user sensitive information accessible in the Referer's URL. So, where previously Firefox Referer header might be: https://www.example.com/path?query

Starting with Firefox 87, when a query is being made to any other origin domain, the Referer header in the query will be set to: https://www.example.com/

Firefox will apply the new default Referrer Policy to all navigational requests, redirected requests, and a page's subresources — image, style, script, etc. — requests to provide a significantly more private browsing experience.

And the best news is that we Firefox users will not need to do anything.

Security News

This Week in RCE Disasters

The week before last, in the long shadow cast by the ProxyLogon vulnerabilities, Seattle-based F5 Networks disclosed patches for critical 9.8 scale vulnerabilities — five in all — in their BIG-IP and BIG-IQ devices.

On March 10th, F5 released an advisory stating that the REST interface of the iControl management interface is vulnerable to an authentication bypass and remote code execution. No detection rules or artifact information was initially provided by F5, and no public exploit was known at the time F5’s advisory was published. This potentially gave system administrators time to patch, and blue teams the space to research and implement detection capabilities. 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.”

112.97.56.78	(China)
13.70.46.69	(Hong Kong)
115.236.5.58	(China)

It may be necessary for the industry's security researchers to reconsider the timing of their release of proof-of-concept code and withhold their disclosures at least until non-script kiddies have themselves demonstrated that the vulnerabilities have been successfully reverse engineered. No ethical researcher wants to have their proof of concept code used, as is, in wide scale, devastating and damaging attacks.

MyBB

The free and open-source forum software "MyBB" was originally MyBulletinBoard, then it was shortened to MyBBoard, and now finally to MyBB. From this point it's going to be hard to make it much shorter. The BB is, of course, written in PHP with a MySQL database backend. It's not massively popular with around 2,100 potentially vulnerable domains having 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 (RCE) 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 were reported to the MyBB Team on February 22. And, as I said, on March 10th an update was released to close the holes.

According to the researchers, the first issue — a nested auto URL persistent XSS vulnerability — stems from how MyBB parses messages containing URLs during the rendering process, thus enabling any unprivileged forum user to embed stored XSS payloads into threads, posts, and even private messages.

MyBB's advisory stated: "The vulnerability can be exploited with minimal user interaction by saving a maliciously crafted MyCode message on the server (for example, as a post or Private Message) and pointing a victim to a page where the content is parsed."

The second vulnerability is an SQL injection in a forum's theme manager that could result in an authenticated RCE. A successful exploitation occurs when a forum administrator 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.

As a result, the researcher's write that "A sophisticated attacker could develop an exploit for the Stored XSS vulnerability and then send a private message to a targeted administrator of a MyBB board. As soon as the administrator 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."

https://blog.sonarsource.com/mybb-remote-code-execution-chain

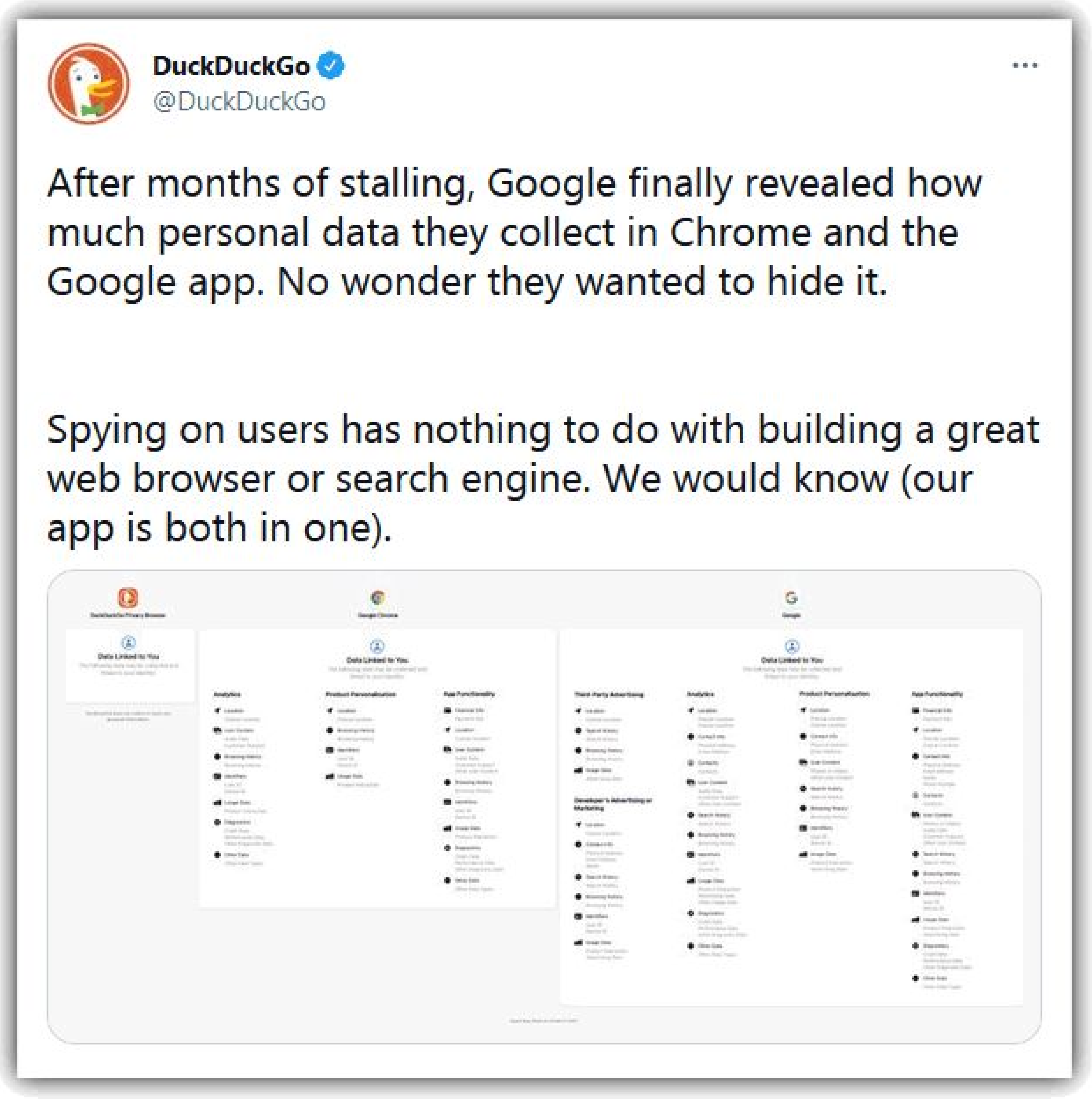
The researchers waited a respectful eight days after the patches were made available to publish their work, which included a complete soup-to-nuts description and discussion, with examples, of the exploit. So, while previously an attacker may have needed to be sophisticated, when armed with their complete 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? We can certainly hope so. But we know that won't have happened.

CAID is able

"CAID" is the China Anonymization ID which is an indirectly Apple-inspired workaround to Apple forthcoming plans to dramatically tighten-up the tracking allowed by App Store apps. I thought that this would be the perfect segue for this week's discussion of Google's FLoC initiative:

We'll be coming back around to this in a few minutes, but eight days ago, on March 15th, the privacy-focused DuckDuckGo search engine tweeted:



DuckDuckGo
@DuckDuckGo

After months of stalling, Google finally revealed how much personal data they collect in Chrome and the Google app. No wonder they wanted to hide it.

Spying on users has nothing to do with building a great web browser or search engine. We would know (our app is both in one).

I thought that this would be a useful preamble for our discussion of Google's planned Federated Learning of Cohorts — aka FLoC. And we should note that DuckDuckGo's comparison is unfair. **They** are not offering 15 gigabytes of free, fast and hyper-robust cloud storage. Nor do they provide the #1 — by far — most popular free eMail service in the world. Gmail has 1.5 billion (with a 'B') users, whereas Outlook is in the #2 spot with 400 million, and Yahoo! Mail somehow holds onto #3 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 catch all throwaway eMail, they are 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 — all 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 happened recently is that Google, after perhaps dragging its feet for three months following Apple's December 2020 announcement of its App Store privacy policy changes, has finally updated its Apple 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 with the forthcoming iOS v14.5, 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 — ID for Advertisers — which is part of a new framework Apple calls ATT — for App Tracking Transparency.

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 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 — fully **99% of users chose not** to give those apps that permission they were requesting.

In his speech delivered on January 28th during the Computers, Privacy and Data Protection (CPDP) conference, Tim Cook, who is, of course, 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."

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 that when users are informed and given a choice they decline to profiled and tracked. So it needs to be done, anyway.

And this is going to become a fraught issue. Just last Wednesday, France's competition regulator rejected calls from advertising companies and publishers to block ATT on antitrust grounds stating that Apple's ATT privacy initiative "does not appear to reflect an abuse of a dominant position on the part of Apple" though it did say that it would continue to investigate the changes to ensure that Apple does not apply less restrictive rules for its own apps.

And, where there's a will to track, and 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 (or CAID) that's aimed at bypassing Apple's new privacy rules — which, as we know, manages an app's access to Apple's officially sanctioned IDFA — the ID for Advertisers. The use of China's nascent CAID would enable companies to continue tracking users without having to rely on Apple's IDFA — and also without, I'm sure, asking for their user's permission.

The Chinese advertising technology firm, with the not-so-subtle name “TrackingIO”, said that "CAID has the characteristics of anonymity and decentralization, does not collect private data, only transmits the encrypted result, and the encrypted result is irreversible, which can effectively protect the privacy and data security of the end user; the decentralized design allows developers to be more flexible to meet business needs.” They added: “Because CAID does not depend on Apple IDFA and can generate device identification ID 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.”

So, although CAID is not yet formally implemented, it’s believed to be under testing by some of China's largest tech companies, including ByteDance and Tencent and several foreign advertising companies 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 reads: “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!

What the FLoC?

“Federated Learning of Cohorts”

On Wednesday, March 3rd, David Temkin, Google’s Director of Product Management, Ads Privacy and Trust, posted a statement about Google’s post-3rd-party cookie tracking plans titled: “Charting a course towards a more privacy-first web”

https://blog.google/products/ads-commerce/a-more-privacy-first-web/

Now, perhaps David is 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.” I would certainly agree that advertising has fueled a lot. And, in fact, advertising does fuel a lot. It fuels this podcast network. It’s arguably the reason that TWiT is still here and going strong after 15+ years. And we know that it certainly fuels Google. As I stated earlier, I’m the 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? Let’s ask David. He continues…

But 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 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.]

That's why last year Chrome 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, we continue to get 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 alternate identifiers to track individuals as they browse across the web, nor will we use them in our products.

[And notice that, coincidentally, we've recently been talking about 3rd-party cookie phase out. Those Firefox cookie same-site sequestration changes are all about phasing out the trackability of 3rd-party cookies. That handwriting really does seem to be on the wall.]

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.**

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 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 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.

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.

> [...] 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 to preserving a vibrant and open ecosystem where people can access a broad range of ad-supported content, with confidence that their privacy and choices are respected. We look forward to working with others in the industry on the path forward.

So, before we go any further, this raises an interesting philosophical question. How do we feel about non-tracking based aggregation about our interests? As individuals interacting with the Internet, do we demand full and absolute privacy — meaning that we are a completely opaque entity? Or is it alright for who we are to be known as an anonymous cloud of likes, desires and interests? As I thought about that, it seems to me that I have no problem with people 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 anything whatsoever about me... without my giving my explicit permission.

And given the title of the EFF's reaction to Google's FLoC, they apparently feel even more strongly. I should note that **the EFF does not like anything, ever.** The only thing I can recall them ever liking was Let's Encrypt. Ohhhh!, they LOVED themselves some Let's Encrypt! Everything else? **No!**

The EFF titled their reaction: "Google's FLoC Is a Terrible Idea." And they apparently wanted to be certain that no one came away from their posting feeling unsure of the details, so their posting is endless. So I'll share the way their posting begins, while interjecting liberally:

https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea

> The third-party cookie is dying, and Google is trying to create its replacement.
>
> 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 lynchpin 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. 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 (FLoC), which is perhaps the most ambitious—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.
>
> 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. Instead of re-inventing 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. This attitude unfortunately strongly biases their language since non-tracking is not new-tracking — it’s non-tracking.]

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 3rd-party cookie tracking. It was never meant to be... but we technologists allowed it to happen.]

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. No one needs to worry that their past browsing will be held against them—or leveraged to manipulate them—when they next open a tab.

[Wait a minute. “... users get to decide what information to share with each site they choose to interact with” ?? Huh? 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. Hmmm.]

In the other, 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. Their 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 must reject FLoC and other misguided attempts to reinvent behavioral targeting. We implore Google to abandon FLoC and redirect its effort towards building a truly user-friendly Web.

[Then, with that introduction, to offer a bit of background, which is interesting for reasons you’ll see in a second, they continue...]

In 2019, 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 primarily 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. Seriously. 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!! That’s why this abbreviation is so godawful! They had to reverse engineer something for FLoC to mean — it’s a FloC of birds! So we get the painfully horrible

> *“Federated Learning of Cohorts.” At least now we know where it had to have come from! Let’s hope it’s a working title. On the other hand, that’s what “MacIntosh was.”]*
>
> FLoC is designed to help advertisers perform behavioral targeting without third-party cookies. *[And I would interject, and also without tracking.]* 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 should belong to each cohort (though that’s not a guarantee).
>
> *[And the small size of that number both surprises and concerns me. I assumed that cohorts would be much larger groupings. But the motivation is clearly to keep them highly targeted. To so that you need small and specific groups.]*
>
> 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 the groups.
>
> *[“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.]*
>
> SimHash can be computed locally on each user’s machine, so there’s no need for a central server to collect behavioral data. However, a central administrator could have a role in enforcing privacy guarantees. In order to prevent any cohort from being too small (i.e. too identifying), Google proposes that a central actor could count the number of users assigned each cohort. If any are too small, they can be combined with other, similar cohorts until enough users are represented in each one.

When the EFF provides some useful and interesting detail:

> According to the proposal ***[which is public on GitHub, by the way. We’ll go there next]***, 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, or whether the ID will be shared in any other ways. 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 cohort identifiers, meaning that there were only 256 possible cohorts. In practice that number could be much higher; the documentation suggests a 16-bit cohort ID comprising 4 hexadecimal characters. The more cohorts there are, the more specific they will be; longer cohort IDs will mean that advertisers learn more about each user’s interests and have an easier time fingerprinting them.
>
> One thing that is specified is duration. FLoC cohorts will be re-calculated on a weekly basis, each time using data from the previous week’s browsing. This makes FLoC cohorts less useful as long-term identifiers, but it also makes them more potent measures of how users behave over time.

So far, despite the 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 any trade off, this really seems like a useful one.

The EFF does present an obvious negative:

> The first issue is fingerprinting. Browser fingerprinting 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 alone shouldn't distinguish you from a few thousand other people like you. However, that still gives fingerprinters a massive head start. 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).

And the EFF notes that:

> 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 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 should not create new fingerprinting risks until it's figured out how to deal with existing ones.

The EFF then highlights a new problem created by this this technology, which they call "cross-context exposure":

> 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.
>
> The project's Github page addresses this up front: "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 (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.
>
> The EFF noted: As described above, FLoC cohorts shouldn't 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.
>
> Two categories of information may be exposed in this way:
>
> 1. Specific information about browsing history. Trackers may be able to reverse-engineer the cohort-assignment algorithm to determine that any user who belongs to a specific cohort probably or definitely visited specific sites.

2. General information about demographics or interests. 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. Moreover, as your FLoC cohort will update over time, sites that can identify you in other ways will also be able to track how your browsing changes. Remember, 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 information about your health, but there's no reason it needs 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 treatment for depression.

FLoC erodes this separation of contexts, and instead presents the same behavioral summary to everyone you interact with.

I'd bet that everyone agrees with those points. I certainly do.

And then, tying back to the beginning, about the inherent problems associated with ANY types of targeted advertising, the EFF makes some additional disturbing observations:

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 other civil liberties.

The power to target is the power to discriminate. 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 it is to advertise shoes.

Over the years, the machinery of targeted advertising has frequently been used for exploitation, discrimination, and harm. The ability to target people based on ethnicity, religion, gender, age, or ability allows 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.

And I'm reminded that we've talked about this before. 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 demographics of any program’s audience, but they have no feedback beyond that. And that makes one wonder whether web and app advertisers shouldn’t be satisfied with choosing which websites to have hosting their ads... rather than having ads able to chase their targets across the web.