Security Now! #815 - 04-20-21

Homogeneity Attacks

This week on Security Now!

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

Browser News

The Vivaldi Project's take on FLoC

Predictably, the Chromium-based privacy-oriented web browsers are all up in arms over Google's FLoC proposal. And the DuckDuckGo search site will be adding FLoC Blocker headers to prevent visits to their search engine from registering in Chrome's FloC aggregation.

https://vivaldi.com/blog/no-google-vivaldi-users-will-not-get-floced/

In the case of the Vivaldi browser project, their posting last Tuesday was titled: "No, Google! Vivaldi users will not get FLoC'ed." So we can guess how they're feeling about this. I want to share the intro of Vivaldi's posting now, and one of the points it makes later when we address this week's topic, the "Sensitivity of Cohorts."

Old habits die hard — Google's new data harvesting venture is nasty. Called FLoC (The Federated Learning of Cohorts), this new advertising technology intends to replace third-party cookies and related technologies like third-party localStorage. This clearly is a dangerous step that harms user privacy. Currently, it is being trialled in Google Chrome and is a part of the Chromium browser engine.

Now the real question: What is Vivaldi's position on this new technology by Google?
This is a valid question as we are based on Chromium. But the truth is that while we rely on the Chromium engine to render pages correctly, this is where Vivaldi's similarities with Chrome (and other Chromium-based browsers) end.

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

FLoC, a privacy-invasive tracking technology — Google will continue to build profiles, and track users, in the absence of third-party cookies and localStorage. It presents FLoC as part of a set of so-called "privacy" technologies, but let's remove the pretence here; FLoC is a privacy-invasive tracking technology.

Does FLoC work in Vivaldi? — The FLoC experiment does not work in Vivaldi. It relies on some hidden settings that are not enabled in Vivaldi. The FLoC component in Chrome needs to call Google's servers to check if it can function since Google is only enabling it in parts of the world that are not covered by Europe's GDPR. It seems there is still some discussion as to whether FLoC could even be legal under the GDPR regulations. We will continue to follow this closely.

Although Vivaldi uses the Chromium engine, we modify the engine in many ways to keep the good parts but to make it safe for users; we do not allow Vivaldi to make that sort of call to Google. We will not support the FLoC API and plan to disable it, no matter how it is implemented. It does not protect privacy and it certainly is not beneficial to users, to unwittingly give away their privacy for the financial gain of Google.

So... message received. The Vivaldi folks are not fans of FLoC.

Chrome continues to be THE high-value target

Last Tuesday, Google released Chrome v89.0.4389.128 to patch two newly discovered security vulnerabilities both of which it says exploits exist in the wild, thus allowing attackers to engage in active exploitation.

One of the two flaws leverages an insufficient validation of untrusted input in Chrome's V8 JavaScript rendering engine. This was the flaw demonstrated by researchers from Dataflow Security during the week before last's Pwn2Own 2021 hacking contest.

The other flaw resolved with this update is a use-after-free vulnerability in Chrome's Blink browser engine which was reported by an anonymous researcher on April 7. So in this case the Chromium team went from report to patch in under one week. Are you hearing this, Microsoft?

We're at Chrome v90

When I went to check up on the version of my instance of Chrome for the story above, I watched it update from that version — 89.0.4389.128 — which was already running, to v90 point something-or-other. Huh! A new major version of Chrome. I wonder what's new there?

HTTPS:// The long-awaited feature which appears for the first time in this release v90 of Chrome is that it will now, at long last, default to the HTTPS protocol scheme when none is specified by the user. So when just GRC.COM is entered into the URL bar, Chrome will first try to initiate a TLS connection to port 443 at the GRC.COM domain rather than first trying to initiate a plaintext connection to port 80. Note that connecting to port 80 will work at GRC as it will for many other websites. But GRC will reply with an 302 MOVED response providing the https:// protocol scheme in the redirection URL. Chrome's move to assume https:// first means that pages will come up more quickly because one entire and totally unnecessary redirection round trip will now be eliminated... for all websites where this had been happening.

Remote port 554 now also blocked to prevent NAT Slipstreaming

Google had previously been blocking port 554 but later removed the block after receiving complaints from enterprise users. However, after performing further analysis of this port, Google has determined that it is used for only approximately 0.00003% of all requests. Due to its low usage, coupled with its potential for abuse, Google is once again blocking it.

AV1: Chrome also gets the new AV1 audio/video codec for increased performance while using videoconferencing software with WebRTC. Google has indicated that the AV1 Encoder brings more of all the good things:

- Better compression efficiency than other types of video encoding, reducing bandwidth consumption and improve visual quality
- Enabling video for users on very low bandwidth networks (offering video at 30kbps and lower)
- Significant screen sharing efficiency improvements over VP9 and other codecs.

Link to Highlight: A feature that's supposedly rolling out in v90 — it's not working for me and

others yet — will be a new feature they call "Link to Highlight." Rather than just linking to an entire web page, when you right-click on a highlighted region, the context pop-up will have "copy link to highlight" which will place a URL with a pound (#) sign onto your clipboard. And supposedly, if you share this pound-sign-embellished URL with others, somehow their use of the link will jump them not only to the page but to the highlight in the page — with that region also highlighted. Except... that's not how pound signs have traditionally worked. Anyone who has coded HTML will know that it's possible to drop explicit anchors into HTML to which the text following the pound sign in a URL can refer, and all browsers will jump to the page and scroll to that previously-placed anchor. For example, GRC's Security Now! archive pages have always contained an ID tag with the episode number so that someone could jump directly to that episode.

That's been it since the dawn of the Internet. But that's changing. I did some digging and I discovered a very new — as in a W3C Working Draft dated last month — which proposes a dramatic extension to the syntax of what can follow a hash tag in a URL. It's supported in Chromium, so all of the Chromium browsers — meaning everything other than Firefox and Safari — also support it. And I imagine that Firefox and Safari may add their support since, unlike FLoC this one is not controversial. I have a link in the show notes for anyone who's curious about Google's proposed implementation: https://wicg.github.io/scroll-to-text-fragment/

And also in the show notes is an example of a link to Wikipedia which, in Chrome, takes its user to the page AND highlights the phrase that Google's search found and linked-to:

https://en.wikipedia.org/wiki/K-anonymity#:~:text=The%20concept%20of%20k%2Danonymity,subjects%20of%20the%20data%20cannot

Security News

Exchange Server Web Shells removed, with DOJ Permission...
Dated last Tuesday, April 13, 2021, FOR IMMEDIATE RELEASE, from the United States Attorney's Office for the Southern District of Texas. The release begins...

> Justice Department announces court-authorized effort to disrupt exploitation of Microsoft Exchange Server vulnerabilities
>
> Action copied and removed web shells that provided backdoor access to servers, but additional steps may be required to patch Exchange Server software and expel hackers from victim networks.
>
> HOUSTON – Authorities have executed a court-authorized operation to copy and remove malicious web shells from hundreds of vulnerable computers in the United States. They were running on-premises versions of Microsoft Exchange Server software used to provide enterprise-level email service.
>
> Through January and February 2021, certain hacking groups exploited zero-day vulnerabilities

in Microsoft Exchange Server software to access email accounts and place web shells for continued access. Web shells are pieces of code or scripts that enable remote administration. Other hacking groups followed suit starting in early March after the vulnerability and patch were publicized.

Many infected system owners successfully removed the web shells from thousands of computers. Others appeared unable to do so, and hundreds of such web shells persisted unmitigated. This operation removed one early hacking group's remaining web shells which could have been used to maintain and escalate persistent, unauthorized access to U.S. networks. The FBI conducted the removal by issuing a command through the web shell to the server, which was designed to cause the server to delete only the web shell (identified by its unique file path).

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

The Acting U.S. Attorney Jennifer B. Lowery of the Southern District of Texas was also quoted, saying: "Combating cyber threats requires partnerships with private sector and government colleagues. This court-authorized operation to copy and remove malicious web shells from hundreds of vulnerable computers shows our commitment to use any viable resource to fight cyber criminals. We will continue to do so in coordination with our partners and with the court to combat the threat until it is alleviated, and we can further protect our citizens from these malicious cyber breaches."

[I've skipped a bit of historical background that all of us have memorized. But then it continues to an interesting conclusion.]

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

The FBI is attempting to provide notice of the court-authorized operation to all owners or operators of the computers from which it removed the hacking group's web shells. For those victims with publicly available contact information, the FBI will send an e-mail message from an official FBI e-mail account (@FBI.gov) notifying the victim of the search. For those victims whose contact information is not publicly available, the FBI will send an e-mail message from the same FBI e-mail account to providers (such as a victim's ISP) who are believed to have that contact information and ask them to provide notice to the victim.

If you believe you have a compromised computer running Microsoft Exchange Server, please contact your local FBI Field Office for assistance. The FBI continues to conduct a thorough and methodical investigation into this cyber incident.

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

I've told the story on this podcast before about how I was involved in a multi-party conference call with officials from the DOJ, some of the politically-connected SANS Security Institute people and a number of other security researchers. The discussion regarded what to do about some of the Internet worm problems we were more often having back then. These were familiar names like "Code Red", "Nimda" and "MSBlast" which were, at the time, actively scouring the Internet seeking new targets while creating some major Internet traffic choke points that were resulting in spotty DoS for some sections of the Net.

So, we techies asked the government whether we could use the known vulnerabilities ourselves to remove the worms from known-infected systems. We believed that we could do it safely and in an entirely targeted fashion.

The response was an unequivocal "Don't even think about doing that!" Really. PERIOD!

In this case, the actions the FBI took were, of course, legal under US law where our courts have the power to selectively legalize activities—with clear boundaries and constraints—which would otherwise be illegal. Since courts have the power to authorize the disconnection of, and even the seizure of, an individual's or company's equipment, it should not be surprising that in this instance the courts authorized the FBI to perform a responsible surgical excision of the backdoors that the FBI's cyber team had identified.

WordPress joins the "FLoC No!" chorus

On Sunday morning, in a blog post titled "Proposal: Treat FLoC like a security concern" WordPress suggests four lines of code to block FLoC. After quoting some of the EFF's "Google's FLoC is a terrible idea" blog post, the WordPress post begins: "WordPress powers approximately 41% of the web – and this community can help combat racism, sexism, anti-LGBTQ+ discrimination and discrimination against those with mental illness with four lines of code." Wow. That would certainly be worth doing if it were in any way true.

```
function disable_floc($headers) {
   $headers['Permissions-Policy'] = 'interest-cohort=()';
   return $headers;
 }

add_filter('wp_headers', 'disable_floc');
```

The WordPress post proposes that this bit of code be added into the so-called WordPress core meaning that, by default, all subsequent updates to WordPress would include this header in its responses. We talked about this header last week. If Chrome sees that a server's response headers contain Permissions-Policy: interest-cohort=() then that site's visits will not be hashed into the FLoC SimHash.

I was trying to think of what the reactions to this put me in mind of, and I hit upon the Apple/Google contact tracing proposal. As we'll recall, none of the popular press, or even the tech press, took the time to understand how the system was designed. We did. And it was clear that it was designed very well with privacy protection as a central tenent. But the media just latched onto some of the scary words uttered by others, who also hadn't bothered to understand the system and ran with them. And this approach appears to be gaining traction.

I'll share something that just happened to me: Yesterday, I had a conversation with someone who had, so far, chosen not to get vaccinated against COVID-19 because he explained that the mRNA vaccines contained pig DNA and that he didn't want pig DNA mixed in with his human DNA. This person has been a friend for about forty years, so I didn't want to be rude, and I was caught a bit off guard. I was pretty sure that there was no pig — or other — DNA in those vaccines. So I explained the mechanism by which a deliberately engineered fragment of mRNA is injected. And how it briefly commandeers our cellular genetics to cause our own bodies to synthesize the characteristic COVID-19 spike protein, which our immune systems subsequently recognize as a foreign invader and consequently build antibody defenses for any future reappearance of the actual spike protein embellished COVID-19 virus. And I also explained that the injected mRNA fragments, which are not DNA, are rather quickly degraded and taken apart by the natural actions of the enzymes that operate our metabolism.

He didn't seem convinced until I explained where that weird rumor must have originated. Because one of the components of the vaccine is polyethylene glycol which, for convenience, is often abbreviated PEG. So, yeah, the vaccines do contain a small amount of PEG, but no PIG.

On this podcast, we're always going to look at the technology. And everyone knows that I have no problem if, once armed with the facts, people come to different positions. Leo, you and I have taken different positions from time to time. Who cares? For me it's the technology that's fascinating, and the future of Google's FLoC initiative will be interesting to observe. And opinions are, and should be, subject to change as facts emerge. Once I explained to Jeff that PEG was not PIG he said "Ohhhhhh!!"

So, at this moment, always subject to change if more is learned, it's clear that FLoC is different from tracking. And that in many ways it is VASTLY more privacy protecting than the cookies and the fingerprinting we have today. Assuming that we can truly kill all long-term tracking, it seems like an improvement. Today, using the existing true tracking technologies, not only EXACTLY who you are by web browser and EXACTLY everywhere you go, including how long you stay and what you do while you're there, is all being explicitly tracked and logged. By comparison, Google's proposal deliberately and significantly "fuzzes up" just some of that information by reducing that explicit identification and explicit website visiting — and activities, which disappear entirely — to a short hash token that indicates nothing EXACT about who you are, where you've been, how long you stayed and what you did while you were there.

But that said, it IS a profile tag. No argument there. And I understand that people don't want to be profiled. I don't want to be. But we keep being told that profiling is the price we pay for an otherwise free Internet. We're told that it supports the commercialization of the Internet and the content that we all take for granted. I've always been skeptical of that, but I have no way of gauging it. Perhaps it's just that those who already have enough, want still more. Soap operas had soap commercials because, back in the 50's, housewives formed their daytime viewing audience. Choosing the demographic of your viewers has always been enough. So perhaps this is just greed. I hope so, and I hope it ends. But it hasn't yet. I have no sense for how committed Google is to FLoC. Maybe it's DOA. Only time will tell. But it's an intriguing technology. One thing seems sure: Calling its tag a "FLoC ID" was a serious blunder. No one likes being "ID"d, especially when that's not what it is. A better name would have been a FLoC CIC — for Common Interests Cohort.

Miscellany

It's Humble Bundle Book Time

I should mention that many of these opportunities come along which I receive notifications of through various channels. But generally they don't quite clear the bar or they have only one day remaining before expiration. But in this case we have two weeks remaining on an O'Reilly Head First series of (mostly) Programming eBooks that looks worthwhile:

https://www.humblebundle.com/books/head-first-programming-oreilly-books

The total of $772 worth of O'Reilly eBooks are all DRM-free and multi-format. Just $1 unlocks:

Head First Ruby, Head First C, Head First PMP (project management), Head First SQL and Head First Statistics.

$10 or more additionally unlocks:

Head First JavaScript Programming, Head First Learn to Code, Head First HTML & CSS, Head First C# and Head First Agile.

I thought "what the heck is Agile? Is that a programming language I've never heard of?" So I read the description, and I still have no idea what it is: "In Head First Agile, you'll learn all about the ideas behind agile and the straightforward practices that drive it. You'll take deep dives into Scrum, XP, Lean, and Kanban, the most common real-world agile approaches today. You'll learn how to use agile to help your teams plan better, work better together, write better code, and improve as a team—because agile not only leads to great results, but agile teams say they also have a much better time at work. Head First Agile will help you get agile into your brain... and onto your team!"

So, like I said, I have no idea what that is, but at least you get JavaScript, Learn to Code, HTML/CSS and C#.

But wait, there's more! If you should choose to shoot the moon for $18, in addition to all of

those and apparently dramatically improving your agility, you'll also receive Head First Go, Head First Java, Head First Python, Head First Kotlin and Head First Android Development.

For coders who might be wanting to stretch out a bit, or for curious non-coders, this seems like a pretty neat deal. So, anytime in the next 19 days from today, April 20th, head over to Humble Bundle <dot> com and scroll a bit until you find the "Head First Programming by O'Reilly" item.

https://www.humblebundle.com/

Closing The Loop

dpmanthei / @dpmanthei
Could setting TTL to some low number help (but not solve) some security issues with web interfaces? Force admins to be within X hops to login to the web interface? Not a solution, but could reduce attack surface.

duckDecoy / @andrewCoyleNZ
@SGgrc I would be interested to hear your thoughts on how to make a "vaccination passport" that couldn't be faked. Any ideas?

StarKiss @StarKissedOne
@SGgrc I know Qnap deserved the beating you gave them last episode, but looking at system defaults, DLNA is turned off by default, so most systems won't be vulnerable. Same with the Plex bug a month ago, it's not even installed by default, let alone set for external access.

SpinRite

Yesterday morning I posted a Progress Report to GRC's spinrite development newsgroup. I'll share the first portion which will be of most interest to those here who are anxious to get their hands on the next SpinRite:

Work is proceeding quite nicely. I'm finally feeling as though I'm completely back in the groove with SpinRite's old code and its segmented, 16-bit coding environment. It's taken a while to make the switch, cognitively, since I code so much by habit and my habits were all wrong after coding, since 2004, exclusively for the 32-bit unsegmented flat model.

After a very good weekend, I have all of the drive discovery and enumeration, listing, selection, feature browsing and display working. I need to determine what I did to break the starting and ending percentage editing, since I've updated its display to show massive sector counts. But something I did back in the beginning broke its UI. This is not surprising, since I ripped out tons of code that was no longer relevant and I needed to make room (within the fixed 16-bit code segment) for all of the new code. The change-over to an entirely new drive database also impacted everything. So a lot of time has gone into finding and fixing everything that became broken.

Once I have the starting and ending percentage screen working, I plan to neuter item #3 from

the Main Menu -- the "Perform Drive Benchmarks" item. Then I'll release what I have for testing by everyone here.

Then, while that testing is underway, I'll work to bring the benchmarking back online. That's a perfect "read-only" solution for the next step, since it means that I need to have all of the various ways SpinRite can now access drives (six, at last count) working in order to perform that benchmarking. Then, we'll test that... which will be a significant milestone toward completion.

Homogeneity Attacks

Given the reactions we're seeing to Google's FLoC proposal, I wanted to introduce FLoC's deliberate awareness of its potential for divulging sensitive personal information. This, too, is something they've thought a lot about. Again, facts are our friends. So I want to start out by citing a part of the not-fully-grounded-in-facts rant from the Vivaldi project since it does reflect some widely voiced industry concerns.

Further down in the "FLoC off!" posting, they wrote:

> FLoC intends to do all of the profiling work within the browser. The browser sees everything you browse, so it gathers the data about your browsing habits and determines your preferences.
>
> This is not like a browser maintaining your browsing history for you. It is analysing your personal behaviour, for Google. It decides which aspects of your browsing behaviour are important, and if enough other people share that behaviour, it assigns you the same ID as all of them.
>
> Advertising companies no longer get to see a unique identifier so they cannot see exactly what you browsed — unless they also happen to be the same company that makes the browser you are using — so they cannot see you specifically. It does sound great.
>
> But they can see that every person who buys certain medical products seems to be in the group (FLoC) 1324, or 98744, or 19287. Now things start getting ugly.
>
> So if you have one of those FLoC IDs, they can display ads for that product — even if that particular medical condition is something you would rather keep to yourself.
>
> It's all anonymised. Sounds like it should be all right, but that is far from the truth.
>
> They can still work out that you have that certain medical issue. That you seem to be in a certain age group, or that you seem to have certain character traits because you share the same ID as other people who have those traits.
>
> Statistical analysis of those IDs is harder for small ad companies. They don't get quite so much data to work with. They don't see every website where that FLoC ID appears.
>
> The company that gets to know the most about that ID is the one that controls the largest amount of the advertising space — Google.

Okay.

Google's research paper is titled: "Measuring Sensitivity of Cohorts Generated by the FLoC API"

https://docs.google.com/a/google.com/viewer?a=v&pid=sites&srcid=Y2hyb21pdW0ub3JnfGRldnxneDo1Mzg4MjYzOWI2MzU2NDgw

The Abstract explains: "We present a discussion of the protections beyond k-anonymity that the Chrome implementation of the FLoC API will provide users. These protections mitigate the risk that a cohort number generated by this API in Chrome leaks sensitive information about the browsing behavior of a user."

Now, at first blush that seems to run counter to FLoC's entire goal. But it turns out that Google really wants this to work and that they have given this considerable consideration. I should mention something about this "k-anonymity." Wikipedia expresses it cleanly:

> k-anonymity is a property possessed by certain anonymized data. The concept of k-anonymity was first introduced in a paper published in 1998 as an attempt to solve the problem: "Given person-specific field-structured data, produce a release of the data with scientific guarantees that the individuals who are the subjects of the data cannot be re-identified while the data remain practically useful." A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from at least k − 1 individuals whose information also appear in the release. (Where 'k' is the individual count.)
>
> k-anonymity received widespread media coverage in 2018 when British computer scientist Junade Ali used the property alongside cryptographic hashing to create a communication protocol to anonymously verify if a password was leaked without disclosing the searched password. This protocol was implemented as a public API in Troy Hunt's Have I Been Pwned? service and is consumed by multiple services including password managers and browser extensions. This approach was later replicated by Google's Password Checkup feature.

So, k-anonymity is a form of statistically provable "fuzzing" of information to obscure explicit identification. Okay, so here's how Google places and frames this entire FLoC effort:

> Today, many publishers are able to leverage interest-based advertising as a source of funding. This revenue stream allows them to offer content free of charge to their users. Contrary to contextual ads, interest-based ads leverage information about a user's interests to decide what ads to show them. Interest-based advertising enables an overall better ad experience for users because the user is presented with more relevant ads than traditional run of network ads; for advertisers, who can better reach their target audience; and for publishers, who are allowed to earn more money, on average, per interest-based ad than a non-relevant ad. In fact, multiple studies from academia and industry have consistently demonstrated that personalized advertising can account for 50-65% of a publisher's revenue.

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

In order to accurately serve interest-based ads, ad tech companies use third-party cookies to generate user interest profiles. Thus, the planned deprecation of third-party cookies in Chrome puts interest-based ads, and the revenue publishers depend on, at risk. To ensure publishers continue to have options to fund their services, Chrome has proposed the FLoC API as a way to enable interest-based advertising in a private way. At a very high level, the FLoC API assigns users to a cohort in such a way that users in the same cohort have similar interests. An ad tech company can then use the API to advertise to an entire cohort.

It has been shown that the FLoC API allows ad tech companies to enable interest-based advertising **without** generating fine-grained browsing profiles of users. *[And remember that fine-grained browsing is what the tracking companies are collecting today. FLoC is only an improvement if we're also able to shutdown all other forms of tracking.]* The FLoC API achieves this by generating k-anonymous cohorts. That is, the API returns a cohort number shared by at least k users. This id can be used as an anonymous replacement of a third-party cookie, allowing ad tech companies to build cohort interest profiles without knowing the identity of a user.

While k-anonymity, especially for large values of k, protects users from reidentification, it is well known in the privacy community that this privacy notion can be vulnerable to so-called **homogeneity attacks**. In the context of the FLoC API, a homogeneity attack corresponds to a scenario where all users that share a cohort number also share a sensitive attribute. For instance, a cohort that consists only of users who visited a website about a rare medical condition. By revealing the cohort of a user, the FLoC API may inadvertently also reveal that a user has investigated that rare medical condition.

At a very high level, we want to make sure that no company, including Google, can correlate a particular cohort with any sensitive attribute.

*[Let me repeat that: At a very high level, we want to make sure that no company, including Google, can correlate a particular cohort with **any sensitive attribute.**]*

The purpose of this paper is to discuss the privacy protections that are needed in order to prevent this type of privacy leakage and what Chrome is doing to prevent homogeneity attacks from happening in the initial FLoC API origin trial. As the implementation of the FLoC API is the responsibility of each browser or software that supports the API, the description of the protections here describe only the implementation by Chrome and not necessarily characteristics that are intrinsic to the API itself.

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

Before we proceed I want to add one more variable; which is the short, one week lifetime of the browsing history that informs the FLoC ID. We've talked a lot in the past about the privacy implications of the potentially infinite age of personal information that's captured about us. One of the other egregious aspects of the current tracking and profiling paradigm is that we have no control over the length of time that our profiles endure. But FLoC sets this limit to a hard 7 days. When we're being tracked, where we go is potentially **never** forgotten. But with FLoC, all profile aggregation of any previous activity disappears after one week. Period.

Google then describes what they mean by "sensitive":

> Before describing the protections Chrome will put in place, we need to define what sensitive categories are. We will use the same sensitive interest categories defined by Google for its interest-based (personalized) advertising product: https://support.google.com/adspolicy/answer/143465
>
> This list of categories was chosen because Google already forbids showing ads related to them as well as targeting a user based on them. Examples of categories in this list are adult and medical websites as well as sites with political or religious content. We will use these categories to decide whether or not a web page is sensitive. While this list of categories certainly does not capture all the nuances of sensitive content (for instance, websites that are not sensitive but that a malicious actor might use, perhaps in combination with other data, as a proxy to infer sensitive attributes), we believe it provides us with a solid foundation that we can build upon. Moreover, the methodology presented here can be applied to any other ontology of sensitive categories as well.
>
> Now that we have established what content is sensitive, we define how we decide whether a cohort leaks sensitive information or not:
>
> ***At a very high level, we want to ensure that no cohort consists of users that have visited web pages related to a particular sensitive category at a much higher rate than the general population.***
>
> [In other words, there's nothing about any given cohort that makes them stand out.]
>
> More formally, we ensure that a cohort assignment satisfies the strong privacy notion of t-closeness.
>
> A cohort assignment is said to satisfy [the property of] t-closeness if it is k-anonymous, and for every sensitive category, the distribution of users who visited a web page related to that category has distance at most t from the general distribution. Intuitively, t-closeness ensures that an adversary that observes a cohort number cannot infer much more about the sensitive browsing behavior of a user than they could before knowing their cohort.

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

One criticism that immediately occurred to me would be that not everyone's "sensitivities" are the same. I might not care much about having my religious affiliation known, whereas I really

don’t want it known that I spend a lot of time cruising monster truck websites.

Ultimately, this does all come down to profiling — one way or another. If we’re to believe the academic sources Google cites, the ad personalization that’s available, thanks to the web’s current operation, at least doubles the revenue generated by ads. That’s a huge deal for all of the websites that have grown and hired staff and can afford bandwidth thanks to the revenue produced by the ads they also host.

So we have a choice: Be truly tracked and have our history amassed and stored without limit and entirely outside of our control — no matter WHERE we go on the Internet - to adult, medical sites or anywhere else; or support an entirely open-source, browser-side, time-limited and self-expiring transient profile tag that has been thoughtfully and carefully designed to explicitly preserve as much of its user’s privacy as possible, which deliberately attenuates strong signals representing sensitive websites, where the presentation of that tag might as much as double the revenue of the advertising websites we visit and wish to support.