

The Death of the Third-Party Cookie

May 2022

David Corish

Abstract

Many significant developments have followed in the wake of Google's proposed phase-out of third-party cookies. In this article, we examine the history of the cookie and explore the viability of Google's proposed replacement.

1 Introduction

Third-party cookies are one of the major tracking technologies which form the core of the web's revenue model. They've created a multi-billion-dollar industry by allowing advertisers to target individual users' browsing habits, revolutionising the world of marketing. However, many privacy concerns have been raised around the extensive amount of personal data collected in the process. Thus, the industry has sought to replace third-party cookies.

In January 2020, Google proposed a ban on third-party cookies by 2024, positing FLoC and later the Topics API as alternatives which allow ad targeting without compromising user privacy. This article will first examine the history of the internet cookie and its regulation, before assessing Google's proposed replacements.

2 Context

In 1994, programmer Lou Montulli, an employee of Netscape, set out to implement a virtual shopping cart. Early internet citizens were intensely privacy focused so privacy needed to be a central component of any internet protocols Netscape built. Keeping this in mind, Montulli rejected his initial concept of giving each user a unique ID that their browser would send to websites. He instead decided to use a small text file passed between a computer and a website to help the website remember information about its visitors. Thus, the internet cookie was born.



Figure 1. Website adherence to three core conditions of EU law

Figure 1. Website adherence to three core conditions of EU law [3]

We can broadly define cookies as belonging to one of two categories. First-party cookies are those generated and stored by the website you're visiting directly. These are used to provide useful functions such as remembering your login details. Third-party cookies are those created by domains outside the website you're visiting; generally placed upon a website through a script or tag (often without the user's knowledge).

Advertisers almost immediately began exploiting third-party cookies to track online users across different websites. Browsers accepted all cookies by default, leading to rampant tracking of user data. In recent years, Montulli has reflected on the early decisions made in the creation of the web. "[T]he web's reliance on advertising as a major revenue source has been very detrimental to society . . . I'm ... thinking the world might have been a better place if we had spent more time working on micropayments or subscription-based content" [1].

The privacy implications of cookies became public knowledge in 1996 and soon after, the Internet Engineering Task Force published a specification recommending that third-party cookies should be disallowed entirely. Internet Explorer and Netscape both ignored these recommendations. Cookies have since become a vital part of the web and its revenue model. The tracking of user data has led to wild profits for tech giants such as Google and Facebook, as well as mountains of information for advertisers to pore over, soon attracting the attention of regulators.

3 Regulation of Third-Party Cookies

3.1 European Union

The EU have introduced several regulations around cookies, mainly surrounding consent. In 2002, the ePrivacy Directive (also known as "the cookie law") specified that websites must allow users to opt-out of cookies if they so wish. This changed to an opt-in mechanism in 2009 [2]. GDPR later stopped websites from relying on implicit consent; where websites display a banner, but the user continues to browse without clicking accept.

Website compliance with European law has been incredibly weak. In a 2020 study of the top 10,000 UK websites, Nouwens et al. found "that dark patterns and implied consent are ubiquitous; only 11.8 percent meet our minimal requirements based on European law." [3]. Cookie walls and consent pop-ups have proven to be a great annoyance and most

users simply blindly accept all cookies. Thus, they've done little to curb the usage of third-party cookies in tracking user data.

3.2 Federal Trade Commission

While the US does not regulate cookies on a federal level, the FTC did propose the “do not track” (DNT) system in 2010. This took the form of a HTTP header field that allowed users to opt-out of tracking. The option was intended to be easy to find and use and be persistent across all websites [4]. All major browsers soon provided support for the header field. However, advertiser support was weak and the DNT system was soon abandoned due to a lack of any legal requirement for companies to respect its use.

3.3 Self-regulation

It is clear that government regulation has been ineffective, but as privacy becomes increasingly important to consumers, parts of the industry have taken steps against the tracking of user data (perhaps to avoid further regulation). In 2005, Safari became the first browser to block third-party cookies by default and in 2017, Apple announced Intelligent Tracking Prevention (ITP). ITP purges ad retargeting cookies after just 24 hours, vastly limiting the ability of advertisers to track users [5]. Firefox took similar steps in January 2019, turning on their Enhanced Tracking Protection by default, blocking all third-party cookies [6].

 Figure 2. Web browser market share worldwide, April 2022

Figure 2. Web browser market share worldwide, April 2022 [7]

All eyes turn to Google, the market leader. Presently, Google Chrome has 64.36% market share (70.5% if we include browsers which use Chromium's engine, such as Edge and Opera) [7]. Therefore, any moves made by Google will have major implications for the web. In August 2019, Google announced the Privacy Sandbox initiative. Here, they argue against the immediate blocking of third-party cookies, believing it would not only massively reduce the revenue of publishers, but also encourage the use of other techniques such as fingerprinting. Schuh points to data which demonstrates “when advertising is made less relevant by removing cookies, funding for publishers falls by 52% on average” [8]. Thus, Google set out to create a system which protects user privacy, whilst maintaining advertiser revenue. Google later announced their intention to phase out

third-party cookies once this new system has been successfully implemented (within two years) [9].

4 Replacing the Third-Party Cookie

4.1 Federated Learning of Cohorts

 Figure 3. An example of two different cohort assignments

Figure 3. An example of two different cohort assignments. Assignment A creates a cohort of abc.com visitors and a cohort of xyz.com visitors. Assignment B creates a cohort of car fans and a cohort of bike fans [12].

Josh Karlin uploaded an explainer of the Federated Learning of Cohorts API (FLoC) in August 2019. Karlin begins by explaining present methods of inferring people's interests, third-party cookies and device fingerprinting. Both of these involve sensitive user data being exposed to advertisers. The FLoC API allows for browsers to generate a key (a cohort id or "flock") which identifies the user as belonging to a distinct group. The browser uses a machine learning algorithm to develop this flock, based on the user's browsing habits. All inputs to the algorithm (such as web history etc.) are kept local on the device - the only piece of data exchanged is the cohort id [10].

The key component of FLoC is that each cohort is "k-anonymous", meaning that each cohort id is shared by at least k users. Google state "the higher the value of k, the more privacy protection we provide to a user" [11]. Under FLoC, you are no longer being tracked as an individual, but are instead an anonymous member of a larger audience. At first glance, FLoC appears to solve the longstanding privacy issues surrounding cookies; all whilst ensuring that advertisers' quarterly earnings do not fall.

4.2 Response to FLoC

FLoC has received widespread criticism from internet users and regulators for placing incredible power in the hands of the browser vendor. As has been established, Google's Chromium takes up a massive 70% of the browser market, giving Google an almost complete monopoly over the advertising industry. In this scenario, FLoC would massively hamper the ability of advertisers to collect data and force them to turn to Google. Google

themselves will continue to collect large amounts of information, meaning the question of privacy remains.

The Competitions and Marketing Authority (CMA) in the UK note a competing proposal from advertising company Criteo: SPARROW [12]. SPARROW builds on the FLoC API by introducing a third-party named the Gatekeeper. The Gatekeeper is an independent service which replaces the role of the browser in FLoC. The Gatekeeper must be independent from all advertisers and adtech companies. This upholds the steps forward in privacy, whilst ensuring that advertisers retain some level of power [13]. Google later responded with the DOVEKEY proposal, replacing the Gatekeeper design with a key-value server, saving on computation.

The FLoC API was dealt a major blow in June 2021 when Amazon opted to block FLoC from gathering data on its sites (only three months after Google began user testing). The same month, Mozilla released a scathing report analysing the privacy risks of Google's plans [14]. Google suspended development of the API in July 2021 and changed strategy, delaying their third-party cookie ban till 2023.

 Figure 4. A detailed overview of Criteo's proposed Gatekeeper design

Figure 4. A detailed overview of Criteo's proposed Gatekeeper design [14].

4.3 Topics API

Google are now replacing FLoC with the Topics API, incorporating feedback and ideas from across the industry. Topics works in a similar manner to FLoC - the browser determines topics that represent your top interests for that week (such as travel, movies etc.) Upon visiting a site, the browser picks three topics to share with advertisers. Josh Karlin argues "the Topics API significantly reduces the amount of cross-site identifiable information. The coarseness of the topics makes each topic a very weak signal" [15].

Topics are far less granular than FLoC cohorts, making it harder to identify individual users. However, privacy risks remain and industry response has been tepid. User testing is set to begin in the coming months. Could the Topics API revolutionise the advertising industry like the third-party cookie once did? Only time will tell.

 Figure 5. Example illustration of what users can see about third-party cookies (left) vs. Topics (right)

Figure 5. Example illustration of what users can see about third-party cookies (left) vs. Topics (right) [16].

References

- [1] Montulli, Lou. Interview. Conducted by Nicolás Rivero, 25 April 2021. “The inventor of the digital cookie has some regrets.” Quartz. <https://qz.com/2000350/the-inventor-of-the-digital-cookie-has-some-regrets/>
- [2] European Union. “Cookies, the GDPR, and the ePrivacy Directive.” <https://gdpr.eu/cookies/>
- [3] Nouwens, Midas et al. “Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence”. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.
- [4] Federal Trade Commission. “FTC Testifies on Do Not Track Legislation.” ftc.gov/news-events/news/press-releases/2010/12/ftc-testifies-do-not-track-legislation
- [5] Apple. “Safari Privacy” https://www.apple.com/safari/docs/Safari_White_Paper_Nov_2019.pdf
- [6] Mozilla. “Firefox Blocks Third-Party Tracking Cookies by Default” blog.mozilla.org/en/products/firefox/todays-firefox-blocks-third-party-tracking-cookies-and-cryptomining-by-default/
- [7] Statcounter. “Browser Market Share” gs.statcounter.com/browser-market-share
- [8] Schuh, Justin. “Chrome: Building a more private web” blog.google/products/chrome/building-a-more-private-web/
- [9] Schuh, Justin. “A path towards making third party cookies obsolete.” blog.chromium.org/2020/01/building-more-private-web-path-towards.html
- [10] Karlin, Josh. “Federated Learning of Cohorts (FLoC)” <https://github.com/WICG/floc>
- [11] Google Research & Ads. “Evaluation of Cohort Algorithms for the FLoC API” <https://github.com/google/ads-privacy/blob/master/proposals/FLoC/FLOC-Whitepaper-Google.pdf>

[12] Competitions and Markets Authority. “The role of tracking in digital advertising.” https://assets.publishing.service.gov.uk/media/5efb1d6ae90e075c53dfce67/Appendix_G__Tracking_and_PETS_v.16_non-confidential.pdf

[13] Criteo. “Certifying SPARROW: Trust but verify” https://github.com/WICG/sparrow/blob/main/certifying_sparrow.md

[14] Rescorla, Eric. “Privacy analysis of FLoC” <https://blog.mozilla.org/en/privacy-security/privacy-analysis-of-floc/>

[15] Karlin, Josh. “The Topics API” <https://github.com/patcg-individual-drafts/topics>

[16] Goel, Vinay. “Get to know the new Topics API” blog.google/products/chrome/get-know-new-topics-api-privacy-sandbox/

David Corish · david@corish.dev

The views expressed are my own and do not represent the views of any employer or organisation.

<https://corish.dev/writing/the-death-of-the-third-party-cookie>