Explaining the Chrome ‘explainers’ for advertising without third-party cookies

With the Chrome announcement of a two-year target timeline to kill the third-party cookie, I thought it would be useful to help make sense of Chrome’s proposals to address advertising use cases *without* third-party cookies.

The Context

First, at this time all four major browsers have publicly disclosed a privacy ideology. Chrome differs from the rest in that their position includes support for the economic benefits of digital advertising – in particular, an acknowledgment that audience-based advertising provides significant revenue lift to publishers.

The future of the web is private by default and that multiple advertising monetization methodologies will coexist and continue to compete. One such model seems likely to be “on-device,” in which data resides local to the client device and specific advertising use cases such as targeting, measurement and decisioning are implemented on-device through privacy-preserving means.

At this time the most comprehensive model for on-device is the “Privacy Sandbox” from Google Chrome. Currently, this exists as a set of “explainers” or draft technical specifications, most of which are still under discussion or otherwise likely to change. However, the underlying concepts and objectives certainly reveal the approximate destination.

A Potential Privacy Model for the Web provides a great starting point to understand the ideology and intention behind

the decisions made around the Privacy Sandbox. So, what is it?

Privacy sandbox

Chrome uses “privacy sandbox” as an umbrella term for the goals and technology proposals, which together endeavor to sustain the open web with the economic benefits of audience advertising, without the need for device identification, third-party cookies, or fingerprinting – and with much more decisioning and other execution handled by the browser.

It’s an ambitious engineering effort comprised of numerous proposals for distinct use cases, each of which we’ll go through below. But on the whole, the summary seems to be:

- Elimination of third-party cookies and all cross-site “tracking.”
- Dramatic reduction of the browser signals (that vary between users) so as to make “inferred IDs” (aka, device fingerprinting) unreliable and unusable.
- Client-side clustering of users into groups sufficiently large so as to offer scale while ascertaining anonymity of any one person/device within that group, but sufficiently small to enable relatively fine-grained reach.
- Utilizing these user groups instead of device-based or cookie-based identifiers as the basis for the core audience use cases (targeting, retargeting, measurement, optimization, etc.).

Privacy budget

First, the “privacy budget” is an effort to help eliminate fingerprinting. The proposal suggests that any data which is both (a) consistent across web domains and (b) potentially useful for fingerprinting (data which tends to be different across devices) will be removed entirely from the browser or

decrease in availability to address the goal of fingerprinting prevention. This seems likely to include:

- Detailed user agent strings including operating system and browser minor version;
- Screen resolution, installed system fonts, and similar data;
- Easily available client IP address information.

Privacy budget suggests that each session or page view will have a "budget" for such data, which means that sites might be able to access a few pieces of information about a device, but not all of it. For example, games that need to know the details of the display hardware may be able to do so, as long as they are willfully blind to the client IP address.

Q: The exact accounting of this budget remains unknown. For example, how much can a site "spend" on one page view, or one session? How much does it "cost" to read screen resolution, or to receive the IP address?

IP blindness

As a piece of the effort to reduce fingerprinting, the Willful IP Blindness document describes a mechanism by which domains can be voluntarily blind to the IP address of the client and can advertise this fact to the client, which itself may change behavior as a result. One imagines that IP blindness would add value to the privacy budget which might be spent elsewhere.

The proposal is for the network operator (a cloud operator like AWS or Azure), edge cache (Cloudflare, for example), or the client itself (a built-in VPN) to hide detailed IP address data from an implementing server; and for some form of audit to ensure compliance.

One imagines that cloud operators like AWS might offer IP blindness as an optional feature of ELB and similar network-

frontend tools.

Several possible solutions are offered for IP-geolocation:

geo-ip-granularity location could be made available through a client hint, which also allows the sites consuming the information to be tracked, measured and potentially denied. Alternatively, if policy allows it, the privatizer could provide geo-IP information to the hosted services.

Where Privacy Budget and Willful IP Blindness seek to take functionality away to prevent a specific type of activity (fingerprinting), the rest of the sandbox attempts to provide mechanisms to address advertising use cases without the use of third party cookies or cross-site identity in general.

Retargeting

The TURTLEDOVE ("Two Uncorrelated Requests, Then Locally-Executed Decision On Victory") proposal endeavors to allow consumer retargeting without any consumers actually identified or identifiable. The proposal defines a mechanism by which an advertiser could ask the browser to persist some piece of information about the user along with any "ad networks" who should have access to the data.

Let's look at the example:

```
var myGroup = {'owner' : 'www.wereallylikeshoes.com', 'name' : 'athletic-shoes', 'readers' : ['first-ad-network.com', 'second-ad-network.com']                                    };
window.navigator.joinPrivateInterestGroup(myGroup, 2592000);
```

What this means: for 2,592,000 seconds (30 days), the site www.wereallylikeshoes.com wishes to disclose to first-ad-network.com and second-ad-network.com that the client is in the "athletic-shoes" segment. The "athletic-shoes" segment name could be anything; the browser is (mostly) agnostic to its content.

Targeting time is a bit more complicated and consists of three steps:

1. A request for an *audience-targeted* ad from one of the ad networks listed above (“first-ad-network.com”), which is stripped of any contextual data such as the domain on which it will run and is returned to the client (as a web bundle) and saved in the browser for rendering at some later point in time.
2. A request for a *contextually-targeted* ad that does not receive audience data (like a “FLoC” defined below).
3. An in-browser auction to decide between the two ads with the auction logic defined by the network.

If this sounds complicated, it is. TURTLEDOVE looks like the most raw specification and the most likely to change in meaningful ways over the next two years.

Category targeting

“FLoC” suggests a path for the browser itself to use browsing activity to cluster devices into “cohorts” of devices with similar interests.

The browser uses machine learning algorithms to develop a flock based on the sites that an individual visits. The algorithms might be based on the URLs of the visited sites, on the content of those pages, or other factors. The central idea is that these input features to the algorithm, including the web history, are kept local on the browser and are not uploaded elsewhere — the browser only exposes the generated flock. The browser ensures that flocks are well distributed so that each flock represents thousands of people.

A “flock” identifier is short enough (“43A7”) such that it cannot — even with other data — be used to uniquely identify a particular device. It is not an index into a targeting taxonomy of any kind; the browser itself provides no

indication of meaning behind the cluster “43A7”.

The flock to which a device belongs (only one “flock ID” will be sent per request)) is sent from the device to servers using another header:

Sec-CH-Flock: 43A7

We can expect DMPs and others to provide a service matching flock identifiers to interest or content taxonomies. We might also expect machine learning models to optimize targeted advertising around these seemingly arbitrary flocks.

Q: How frequently will a user’s flock change?

Abuse prevention

Privacy sandbox suggests that Trust Tokens could be used for one domain (a trust provider) to issue a “trust token” to the device. This token is a magic tiny bit of storage intended to preserve a minimal piece of information (something like “is this user a subscriber?” or “is this device likely to be fraudulent”) the contents of which are entirely invisible to the user or device itself.

Trust tokens are “issued” from one site and may be “redeemed” on another. Redemption requires the consent of the issuer.

Measurement and reporting

The dual documents Aggregated Reporting API and Conversion Measurement API suggests a future for reach and conversion reporting in which aggregate data is available without event-level visibility. Unlike the targeting proposals that rely on user clustering algorithms and non-identification (inaccuracy) by design, these proposals focus on accuracy in aggregation.

For example, a client accepts an impression event:

```
var entryHandle = window.writeOnlyReport.get('campaign-123');
```

Along with a timestamp and any associated demographic or geographic data:

```
entryHandle.set('country', 'usa'); entryHandle.set('date', new Date().toDateString());
entryHandle.reportAfter(msecFromNowUntilMidnight());
```

And at some time in the future, the browser will aggregate all impression delivery events and submit them to a server-side aggregation service (presumably operated by Chrome), which itself submits reports to an advertiser service operated on a well-known URL.

Conclusion

Looking across the multiple proposals that make up Privacy Sandbox, and indeed represented within other browser privacy actions, clearly the browsers aim to remove third-party access to identifiers entirely (including the ability to even "guess"). Their objectives for doing so are to eliminate all web-scale, cross-site tracking of consumers, and achieve a global reset to "privacy by default" where trust is no longer a baseline assumption but is enforced by the browsers.

I appreciate these motivations behind Privacy Sandbox, and are participating in the process to shape its evolution. However, it would be helpful to see solutions that apply across the entire web platform and perhaps beyond.

To rebuild the advertising infrastructure supporting the open web, I hope we also return to trust and enable consumers, publishers, brands and tech platforms to engage in personalized content with greater trust and accountability throughout the system. Tech Lab recently announced Project Rearc with similar goals to increase trust and accountability within the digital advertising ecosystem.

Opinions expressed in this article are those of the guest author and not necessarily Search Engine Land. Staff authors are listed here.

About The Author

Sam joined IAB Tech Lab with the DigiTrust acquisition as a founding board member and volunteer CTO. Prior to his role as CTO of Tech Lab he was in engineering leadership roles for nine years at Rubicon Project, where he helped to grow the business from an ad network optimization startup to an independent global leader in Real Time Bidding. Sam has a BA in Communications from UC San Diego and an MBA from Cornell University.

http://feeds.searchengineland.com/~r/searchengineland/~3/EU8WJviTX3w/explaining-the-chrome-explainers-for-advertising-without-third-party-cookies-331647