IPA

Design Goals and Tradeoffs

Proposed approach & design choices:

Use-case	Start with basic measurement
Event-level vs Aggregate	Aggregate
On-device attribution vs within MPC	Within MPC
How to protect the entire system against adversarial actors	● Add DP noise ● Manage a Privacy Budget ● Grain: ○ `user` **`x`** `site` **`x`** `time-window`

Which use-cases should we work on?

Start with basic measurement

Based on my observations, it seems to me that there exists:

- A consensus that basic measurement is a something we want to support
- A consensus that the best way to achieve this goal is to add new APIs to the web platform to support this use-case privately
- A consensus that technologies exist which make doing this in a private way possible

Move on to other things next

In my personal opinion:

- There does not appear to be consensus that we should add new APIs to the web platform specifically to support:
  - Remarketing
  - “Interest Based Advertising”, where “interests” are automatically inferred from passive web browsing behavior
  - Other more advanced use-cases
- I think that if we try to start with advanced use-cases, we will get stuck and not make much progress.

Build trust by actually shipping something together first

- I think we need to start by laying a foundation of trust.
- I think we need to demonstrate the ability to:
  - Work together
  - Reach consensus on something
  - Actually standardize and then ship an API
- This will help us build trust in one another, and prepare us for the more difficult discussions to come which will require a lot of openness

Basic measurement seems like the best place to start.

Event-level vs Aggregate?

Event vs Aggregate?

Let’s talk about Chrome’s proposed “Event level reports” in the “Attribution Reporting API”

- 64-bit identifier (“attributionsourceeventid”)
- Reports are associated with a specific ad-click
- Designed to **not enable record linkage**
- Differential privacy added (potentially) to add “plausible deniability”

Is “event level” within the “consensus window”?

- Based on the reactions from Safari and Mozilla, this appears to fall **outside of the consensus window**
  - Even if it cannot be used to establish record linkage…
  - Even if we add differential privacy to provide “plausible deniability”...
- Event level reports tell you (somewhat noisy) information about what a specific person did on a specific other website
  - e.g. “There is a 90% chance that Ben signed up for a subscription on news.example”
  - e.g. “There is a 75% chance that Ben bought something on shop.example”
  - e.g. “There is a 50% chance that Ben made a donation to political candidate X’s campaign”

There does not appear to be consensus that APIs which reveal this type of information about people’s behaviour should be on by default

How about Aggregate APIs?

- Aggregate measurement APIs, designed to reveal only total counts across larger groups of people **seem to fall within the consensus window**.
  - e.g. “This ad campaign led to 45 purchases”
- Safari has shipped both PCM and SKAdNetwork
- Chrome has proposed an “Aggregated Reporting API”
- Edge has expressed support for this “Aggregated Reporting API”
- Facebook and Mozilla have co-published the IPA proposal (also aggregated)

On-device attribution vs within MPC

Adversarial attacks against PCM

- Even with mitigations in place, it’s pretty easy to link an attributed attribution report with an individual:
  - “attributed_on_site”
  - 256 values for “source_id”
  - 16 values for “trigger_data”
  - Rough time window
  - Rough IP range
- "attributed_on_site" is the “registrable domain” where the attributed conversion occurred.
  - There are an unlimited number of “registrable domains”.
  - An attacker can register one unique domain for every 256 users
  - It’s absolutely possible to learn 100% accurate data about people (maybe a lot of people)

Chrome’s Aggregated Reporting API

- Chrome’s Aggregated Reporting API also proposes attributing trigger events to source events on device, then (after a delay) sending “Encrypted Aggregate Reports” directly to callers of the API.
- The combination of timestamp, IP-address, and any metadata included along with these “encrypted aggregate reports” might also risk them being fingerprinted to reveal which specific people had an attributed conversion.

IPA

- We proposed moving attribution into the MPC to avoid this problem.
  - Ad impressions (source events) return an encrypted report 100% of the time - which reveals no new information to the caller.
  - Ad conversions (trigger events) also return an encrypted report 100% of the time - which reveals no new information to the caller.
- As such, even if you look at the timestamp, IP address, or any metadata - these reports cannot reveal if there was a match (an attributed conversion)
- This leads to better privacy characteristics AND better utility
  - No need for **delays** (a big pain point today)
  - No need for IP-blindness (difficult)
  - No need to limit the entropy of source_ids / trigger_data (also a big pain point today)

Cross-device / cross-platform conversion attribution

- PCM stays in Apple land, and only works for same-device conversions
  - iOS App => Safari (same device)
  - Safari => Safari (same device)
- SKAdNetwork is the same
  - iOS App => iOS App (same device)
- Chrome’s aggregated reporting API
  - Android App => Chrome (same device?)
  - Chrome => Chrome (potentially different device IF both are signed-in to Chrome)

IPA

- The “I” in IPA stands for “interoperable”
- Source and trigger events have a standardized format, and can be generated by any browser / mobile operating system to support all the types of ad flows:
  - iOS App => Chrome (different device)
  - Android App => Safari (different device)
  - Smart TV (Android?) => Firefox
  - Facebook In-App-Webview => Chrome
  - Etc.
- This is a very important design goal. Omitting cross device conversions will severely undercount the total number of conversions an ad campaign drove.
  - It varies by the business buying ads, but I have seen examples of businesses for which Mobile App => Desktop browser comprises 40% of the total number of conversions

Differential Privacy + Privacy Budget

How much information is revealed about a person?

PCM:

An adversarial actor can collect an unbounded amount of perfect user-level information - limited only by the rate of clicks

- For each advertising business, just register a large number of domains, such that you only send 256 unique people to each domain.

How much information is revealed about a person?

Chrome's Aggregated Reporting API:

An adversarial actor can collect a large amount of high-confidence user-level information per time-window

- The privacy budget is at the grain of:
  - `browser x source-site x destination-site x time-window`
- There is no limit on the total number of "destination-sites", so an unlimited amount of information can be revealed
- If one entity controls N source-sites and can match user-identity across them (e.g. via login, or link-decoration based tracking), it can actually get N-times the amount of data about what a given person is doing on a given destination-site

How much information is revealed about a person?

IPA:

Aim: There is a fixed upper bound on the total amount of information an adversarial actor can collect about a specific person across ALL other websites within a given time-window

- The privacy budget is at the grain of:
  - `user` **`x`** `calling-site` **`x`** `time-window`
- Each website gets a per-user privacy budget. You cannot learn more information about someone by sending them to more unique domains.
- The information you learn is very noisy. The API is not designed to reveal user-level information.

Recap

Proposed approach & design choices:

Use-case	Start with basic measurement
Event-level vs Aggregate	Aggregate
On-device attribution vs within MPC	Within MPC
Differential Privacy + Privacy Budget	● Add DP noise ● Manage a Privacy Budget ● Grain: ○ `user` **`x`** `site` **`x`** `time-window`