Privacy-Preserving Attribution Proposed Roadmap

PATWG (one can dream), TPAC 2024

Congratulations on your new Working Group

Condolences about the delay.

Time to get to work

We can’t do this forever

	Data minimization (worst case failure)	Explainability to the user	Delays due to offline devices	Fraud vulnerabilities	Post-Hoc attribution logic	Replay protection cost	MPC cost
IPA	Reports tied to a persistent device-level match key	Not able to tell the user which conversions attributed to which impressions.	No delay to receive encrypted match key	Publisher queries are independent of one another.	Different attribution logics can be run adaptively on the same data	None	Aggregation dominates the cost
PAM	Histogram contributions per impression or conversion (“2-party cookie”)	Users can see which conversions attributed to which impressions.	Publisher reports scheduled, leading to delay from conversion to scheduled time, and potentially more delays due to offline device/app	Fraudulent and/or accidental clicks on other publisher sites can lead to publishers receiving fewer attributed conversions.	Attribution happens only once, but assuming PAM supports late-binding of histogram bins, it can enable adaptive reruns	Both impressions and conversions	Zero-knowledge proof and aggregation
IPA-PAM Hybrid	Reports tied to short-lived random identifiers (“2-party cookie”)	Users can see which conversions attributed to which impressions.	No delay for impression matchkeys. No delay for conversion reports	Publisher queries can be independent of one another(optionally). Cross-publisher attribution still vulnerable	Optionally in advertiser queries, selection of eligible ads can occur on device while attribution occurs in MPC	Conversions only	In honest majority MPC, similar MPC costs to IPA

Goals

Agree on a starting point

Meet basic needs and deliver something

Provide a platform on which to improve

Approach

Focus on core problems

Good, not perfect

Identify gaps and fix them when we can

Warning

Strawman incoming

Proposed Starting Point:
PAM
+Individual DP
-Hard Stuff
+Easy Stuff

PAM Recap

Impressions are saved by the browser

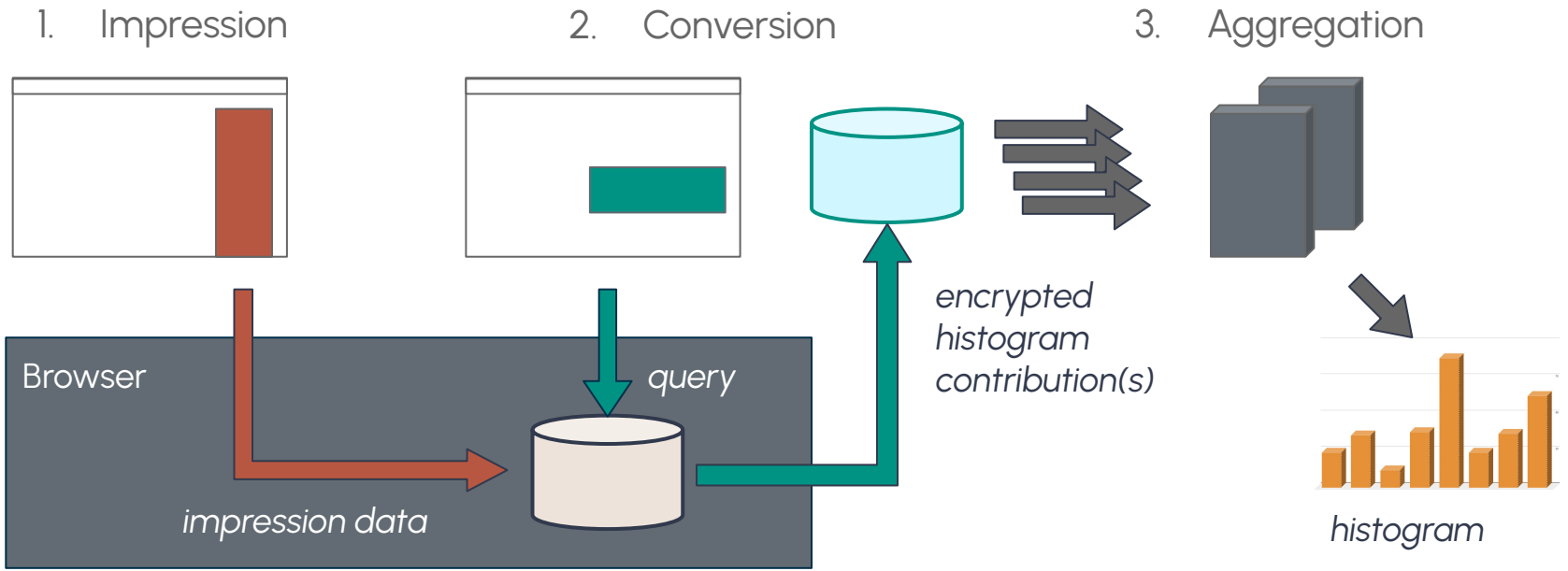
At conversion time, query impressions

Emit an aggregate histogram contribution

Aggregation services sums histograms

+ adds noise
+ prevents replay

PAM Overview



Save Impressions


```
const attribution =
navigator.privateAttribution;

attribution.saveImpression({
  bucket: 3,
  filterData: 7,
  conversionSite: “advertiser.example”,
});
```

Save Impressions

Histogram Bucket

(optional) filter data

Who Can Convert

```
const attribution =
navigator.privateAttribution;

attribution.saveImpression({
  bucket: 3,
  filterData: 7,
  conversionSite: “advertiser.example”,
});
```

Browser adds:

- top level site
- iframe site
- timestamp

Impression Store

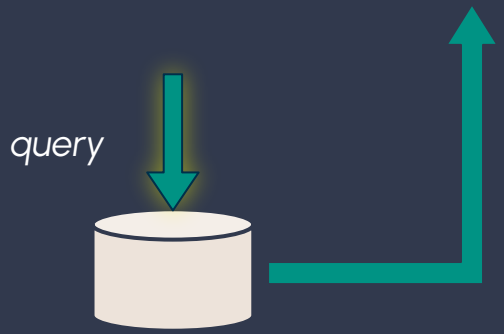
Browsers store impressions

- Storage associated with conversion site

Attributes include

- impression site (top-level)
- intermediary site (iframe)
- timestamp
- filter data

Measure Conversions




```
const report =
attribution.measureConversion({
  aggregator: “Honest Abe’s”,
  epsilon: 1,
  logic: “last-touch”,
  histogramSize: 20,
  value: 3,
  lookbackDays: 30,
  filterData: 7,
  impressionSites: [
    “example.com”,
    …otherSources
  ],
});
```

Measure Conversions

Privacy Budget Expenditure (How much and with which Helper Party Network)

```
const report =
attribution.measureConversion({
  aggregator: “Honest Abe’s”,
  epsilon: 1,
  logic: “last-touch”,
  histogramSize: 20,
  value: 3,
  lookbackDays: 30,
  filterData: 7,
  impressionSites: [
    “example.com”,
    …otherSources
  ],
  intermediarySites: [
    “adtech.example”,
  ],
});
```

Measure Conversions

Attribution Logic (The value to allocate and how)

```
const report =
attribution.measureConversion({
  aggregator: “Honest Abe’s”,
  epsilon: 1,
  logic: “last-touch”,
  histogramSize: 20,
  value: 3,
  lookbackDays: 30,
  filterData: 7,
  impressionSites: [
    “example.com”,
    …otherSources
  ],
  intermediarySites: [
    “adtech.example”,
  ],
});
```

Measure Conversions

Impression Choice (Which ads, how far back, site)

```
const report =
attribution.measureConversion({
  aggregator: “Honest Abe’s”,
  epsilon: 1,
  logic: “last-touch”,
  histogramSize: 20,
  value: 3,
  lookbackDays: 30,
  filterData: 7,
  impressionSites: [
    “example.com”,
    …otherSources
  ],
  intermediarySites: [
    “adtech.example”,
  ],
});
```

Delegation

Top-level site can use permission policy

- for both API functions, separately

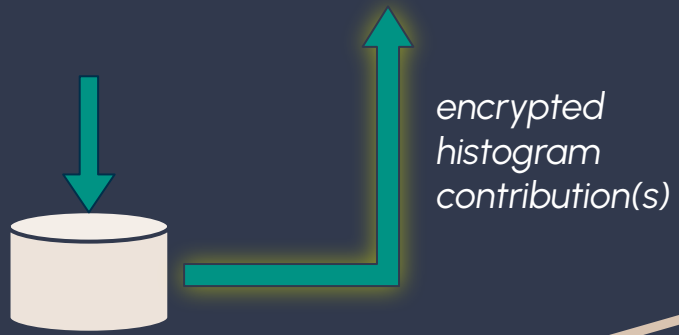
measureConversion uses advertiser's budget

- delegation maybe could include budget

Browsers decide maximums for

- How many delegations
- How much budget for each
- How much budget in total

Conversion Report



Contents will depend on aggregation service

Envelope will include metadata used

- Chosen aggregator
- (Potential) privacy budget spent
- Timestamp
- Conversion site
- Other aggregation protocol stuff

Encrypted histogram format will vary

Aggregation

Report collectors submit batches of conversion reports for aggregation

Implementation details flexible

- Option 1: DAP/MPC
- Option 2: TEE

Some work necessary for either

Aggregation: Common work

Aggregator oversight and governance

Define submission API (or APIs)

Common constraints on operation

Differential Privacy

Aggregation: MPC Work

Define how to use DAP (IETF Protocol)

Extensions to support privacy budgeting

- Queries that use less than full budget
- Tracking budget between queries

Improve scalability(maybe)

Batch submission (maybe)

Budget tracking extension

Aggregation: TEE Work

Implementation of code to run in TEE

Key release procedures

- Validate attestation from TEE
- N-of-M threshold keys?

Anti-replay design

Necessary Safeguards

Differential Privacy

A good model exists:
https://arxiv.org/abs/2405.16719

Noise added during aggregation

Application in TEE is OK

Application in MPC in progress

Anti-Replay

Conversion reports have limited uses

- Use once, or
- Use in parts without exceeding budget

Aggregation services need to ensure

- No duplicate reports in each query
- Budget is not exceeded

Transparency

Need to develop accountability plan

Provide what transparency we can

Propose:

- Aggregators publish all queries
- Include all non-**user**-private information

Possible data:

- Number of impressions and conversions
- Histogram Size
- Impression sites
- Privacy budget consumption

Harder Stuff:
Still Important

Propose to Explicitly Defer

Logistic Regression

Fancy, data-driven multi-touch attribution

Late binding

Massive histograms

Better fraud mitigations

Anything that could need more MPC

Deferral
Not Rejection

Let's start by shipping *_something_* quickly

Then let's get in the cadence of regular updates.

Let's not slow ourselves down by forcing all the features to be in the original version

We should keep working on these problems

Until they are mature

Add capabilities when ready

Easy Stuff

Some improvements could be easy

- e.g., ad tech/publisher/DSP reports

If it fits, add it

Constraints:

- Consider impact on shipping
- Needs implementation support

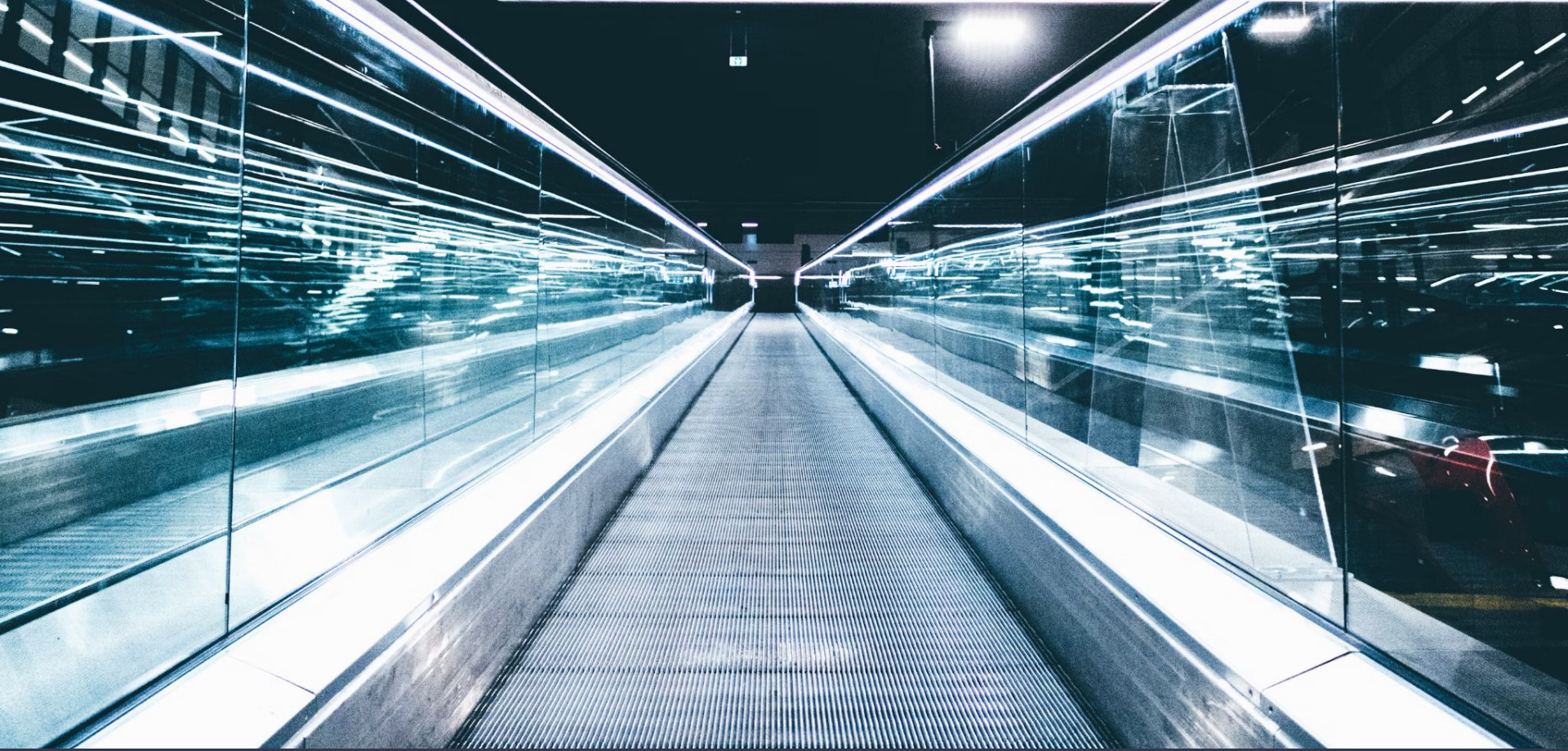
What else is easy?

Open question for discussion

Proposal:

- Build list of wants
- Prioritize on feasibility/importance
- Avoid anything that will delay shipping

What does the group want?



Discussion