Privacy & Utility Goals

1. It must be difficult to re-identify significant numbers of users across sites using just the API.
2. The API should provide a subset of the capabilities of third-party cookies.
3. The topics revealed by the API should be less personally sensitive about a user than what could be derived using today’s tracking methods.

RESULTS

Noisy and Genuine Topics can be Identified

Binary Classifier: every topic that does not appear at least on 10 websites among the top 1M is considered noisy.

Repetitions leak genuine topics: Coupon Collector’s Problem.

One-shot: 25% of noisy topics removed.

Multi-shot (15-30 epochs): 49-94% of noisy topics removed.

Advertisers can Re-identify Users

The Topics API returns at least 1 true topic aligned with user interests in about 60% of cases.

Some Utility Retained, but the Topics API can be Manipulated

The Topics API returns at least 1 true topic aligned with user interests in about 60% of cases.

Topics (word): Comics (batman), Dance (dance), ...

Domain: example.com

Crafted Subdomains: batman.example.com, dance.example.com, ...

= 3.5M subdomains

CONCLUSION

The Topics API Discloses Users’ Online Behaviors

Natural properties about user interests can break Topic’s privacy claims of non-re-identification.

Specifically, users with stable interests are as uniquely cross-site trackable with Topics as with Third-Party Cookies.

Direct foreseen mitigations are only partial.

Other Avenues?

These modifications could impact billions of users for the better or the worst:

• Google with The Privacy Sandbox (Topics API, FLEDGE, …).
• Microsoft with TURTLEDOVE.
• Apple with Private Click Measurement.
• Brave with Brave Private Search Ads.
• Meta and Mozilla with Interoperable Private Attribution.
• …