Poster: Reject All, Still Tracked? Cross-Layer Evidence from 10,000 CMP-Enabled Websites

Zakaria Mekelleche **Hassina Meziane**

LITIO Laboratory, University of Oran 1 Ahmed Ben Bella, Oran, Algeria

mekelleche.zakaria@edu.univ-oran1.dz meziane.hassina@univ-oran1.dz

Research Questions

Consent banners can expose a visible **Reject All** control, but this does not guarantee that the preference is technically propagated to cookies, redirects, scripts, tag managers, and server responses.

RQ1. What tracking-related behavior remains observable across the web stack under automated CMP rejection?

RQ2. Which sources provide direct stateful evidence versus weaker behavioral or stateless evidence?

RQ3. How can multi-layer tracking inference be made auditable while avoiding over-attribution from CDN use, cookie-name reuse, or benign browser-feature checks?

We interpret the crawl as **rejection-conditioned**, not as a strict after-click timing claim: OpenWPM does not expose a reliable event-level CMP-click boundary.

Measurement Pipeline

Tranco-ranked pool
EU-reachable candidates

↓

CMP validation
Clean baseline vs extension modes
selectors + frames + TCF + events

↓

OpenWPM crawl
Firefox/Xvfb + fresh state
rejection-conditioned visit

↓

Typed evidence events
HTTP, JavaScript, redirects

↓

Context enrichment
Category, party relation, lists, CDN

↓

Deterministic integration
D_j predicate + bounded score

Evidence Layers

Each observation is normalized as a typed event with site, host, party relation, operation, cookie/category context, resource type, list match, and CDN context.

- HTTP request cookies and `beacon/xmlhttprequest` context
- HTTP request and response cookie evidence
- JavaScript `document.cookie` reads and writes
- Fingerprinting-related JavaScript API access, storage operations, and transmission primitives
- Redirect-chain cookie evidence

Site-Level Outcomes

77.1% **Confirmed evidence** 7,706 websites

18.3% **Likely evidence** 1,826 websites

1.0% **Needs review** 104 websites

3.6% **No signal retained** 364 websites

Outcome	Websites	Share
Confirmed tracking evidence under rejection condition	7,706	77.1%
Likely tracking under rejection condition	1,826	18.3%
Needs manual review	104	1.0%
No tracking signal retained	364	3.6%

Signal Coverage Highlights

Signal source	Sites (%)	Events
JS fingerprinting signals	9,381 (93.8)	3,407,293
JS transmission primitives	8,816 (88.2)	417,360
JS storage writes	8,798 (88.0)	419,042
JS cookie evidence	7,631 (76.3)	348,961
HTTP `beacon/xhr` context	6,953 (69.5)	69,108
HTTP request cookies	6,766 (67.7)	746,675
HTTP response `Set-Cookie`	5,264 (52.6)	89,826
Tracker-list hits	3,946 (39.5)	318,124
CDN cookie context	3,030 (30.3)	23,666
Redirect cookie evidence	1,850 (18.5)	21,995

Rows are not mutually exclusive. Among sites with fingerprinting-related JavaScript signals, **8,697 (92.7%)** also invoked transmission primitives; the remaining 684 are treated as weaker evidence.

Technical Audit Units

Corpus unit: a retained site is EU-reachable, accessible, non-duplicative, and exposes a detectable CMP.

Validation unit: clean baseline and extension modes compare CMP state, TCF summaries, storage, events, and mutations.

Strong rejection evidence: observed reject/decline/refuse action or TCF data recording denial for most purposes.

Lower-confidence evidence: preference saves, storage changes, CMP hiding, or banner disappearance are not definitive rejection.

Direct evidence: non-essential cookie evidence in HTTP requests, responses, JavaScript access, or redirects activates D_j.

Guardrail: banner hiding, CDN context, repeated cookie names, or isolated API checks cannot decide the site label alone.

How We Avoid Over-Claiming

- **Cookie-name reuse is contextual:** identical names may carry different values and purposes.
- **CDN labels remain infrastructure context:** CDN involvement is not tracking evidence by itself.
- **Ownership and tracker-list roles are separated:** Tracker Radar supports same-company checks, while EasyList/EasyPrivacy provide tracker-list context.
- **Isolated API reads are weak:** fingerprinting APIs require storage, transmission, third-party, or list context.
- **Guardrail-only observations cannot label a site:** CDN/list/name/API context cannot produce confirmed or likely evidence alone.

Evidence Integration

Direct-evidence predicate. For site j, $D_j = 1$ when non-essential cookie evidence appears in request headers, response `Set-Cookie`, JavaScript cookie access, or redirects.

Label rule. If $D_j = 1$, label **confirmed**. Otherwise compute a bounded behavioral score and label **likely** only when $s_j > 5$:

$$s_j = \min(100, \sum_{k \in K} w_k n_{j,k} + B_j).$$

K	supporting signal families after direct evidence is absent.
$n_{j,k}$	normalized event count for signal k at site j.
w_k	ordinal strength: 1=weak context, 2=storage/transmission, 3=high-intensity or WebRTC fingerprinting-related signals.
B_j	bonus only for coupled behaviors: fingerprinting+transmission, storage+transmission, or third-party fingerprinting+storage.

Contributions and Project Status

1. Rejection-conditioned audit of 10,000 CMP-enabled websites.
2. Typed-event model across HTTP, JavaScript, redirects, and contextual enrichment.
3. Guardrail-driven interpretation of cookie names, tracker-list, and CDN context.
4. Deterministic integration separating direct evidence from behavioral evidence.
5. Empirical finding: 77.1% confirmed evidence and 95.3% confirmed-or-likely evidence under the rejection-conditioned crawl.

Project status. The crawl, evidence extraction, and site-level integration have been executed. The manuscript and reproducibility documentation are being finalized, including clearer descriptions of the scoring weights, guardrails, and corpus-construction scripts.

Takeaway: CMP effectiveness must be measured across the web stack, not only at the banner interface.