Work-in-Progress: RegTrack: Uncovering Global Disparities in Third-party Advertising and Tracking

Tanya Prasad Rut Vora Soo Yee Lim Nguyen Phong Hoang Thomas Pasquier
The University of British Columbia, Vancouver, Canada

Abstract—**Third-party advertising and tracking (*A&T*) are pervasive across the web, yet user exposure varies significantly with browser choice, browsing location, and hosting jurisdiction. We systematically study how these three factors shape tracking by conducting synchronized crawls of 743 popular websites from 8 geographic vantage points using 4 browsers and 2 consent states. Our analysis reveals that browser choice, user location, and hosting jurisdiction each shape tracking exposure in distinct ways. Privacy-focused browsers block more third-party trackers, reducing observed *A&T* domains by up to 30% in permissive regulatory environments, but offer smaller relative gains in stricter regions. User location influences the tracking volume, the prevalence of consent banners, and the extent of cross-border tracking: GDPR-regulated locations exhibit about 80% fewer third-party *A&T* domains before consent and keep 89–91% of *A&T* requests within the EEA or adequacy countries. Hosting jurisdiction plays a smaller role; tracking exposure varies most strongly with inferred user location rather than where sites are hosted. These findings underscore both the power and limitations of user agency, informing the design of privacy tools, regulatory enforcement strategies, and future measurement methodologies.**

I. Introduction

Third-party advertising and tracking (*A&T*) underpin much of the web's business model, with these technologies present across the web ecosystem [1, 2, 3]. These mechanisms enable cross-site profiling, where advertisers and intermediaries infer browsing behavior, interests, and demographics from identifiers and interaction traces [2, 4]. Browsers exercise substantial control over user privacy by mediating page execution and network requests, exposing levers to shape storage and communication [4, 5]. This technical position allows browsers to block or rewrite requests, restrict cookies and other state, and deploy anti-fingerprinting and anti-tracking defenses [6, 7]. However, dominant browsers are developed by firms with significant advertising businesses, raising questions about incentive alignment between privacy protections and advertising addressability, motivating empirical evaluation of what users actually experience.

Different jurisdictions have introduced privacy regulations constraining tracking and data flows, with varying obligations and enforcement models [8, 9, 10]. The EU's GDPR emphasizes opt-in consent for tracking, whereas US laws like CCPA and CPRA center opt-out from "sale" or "sharing" [8, 10]. Empirical audits report substantial variation in compliance and dark patterns in consent interfaces, producing geographic disparities in effective privacy protection [9, 11, 12, 13].

These observations point to a methodological gap: prior work has not systematically compared the *relative* and *joint* effects of *browser choice*, *user location*, and *hosting jurisdiction* on tracking exposure using a single, controlled measurement design. Cookie banners and vendor responses may differ for EU vs. non-EU users [14, 15]; trackers may react to IP-level geo signals [12, 16]; and sites may deploy jurisdiction-wide policies based on where they host. We therefore ask three interconnected questions, ordered by decreasing user agency:

- **RQ1: Effect of browsers**—How do different browsers affect third-party *A&T* exposure?
- **RQ2: Effect of user location**—How does a user's geographic location (as inferred by sites) influence *A&T*?
- **RQ3: Effect of hosting location**—How does a website's hosting jurisdiction affect third-party *A&T* practices?

We begin with browser choice (RQ1), which users control directly; proceed to user location (RQ2), which users can sometimes influence (e.g., via VPNs); and conclude with hosting jurisdiction (RQ3), a structural factor beyond individual control. This ordering lets us assess how each lever contributes to observed tracking exposure under matched conditions. To support this, we design RegTrack, a consent-aware factorial measurement framework varying browser and browsing location over a shared list of popular sites, using synchronized crawls from 8 vantage points and attributing third-party requests to known *A&T* domains. Our main contributions are:

- We design RegTrack, a multifactor, consent-aware measurement framework systematically studying the effects of browser, user location, and website hosting jurisdiction on third-party *A&T*.
- We collect and analyze a dataset spanning 8 geographic vantage points, 4 browsers, and 743 popular websites, enabling within-factor contrasts and interaction analysis.
- We compare the magnitude of browser-, location-, and hosting-related differences in tracking exposure and identify where user-controllable choices provide meaningful leverage versus where structural forces dominate.

Our measurements yield three main findings. First, browser choice matters most in permissive environments: privacy-focused browsers (e.g., Brave) substantially reduce tracking in the US and opt-out contexts, while differences narrow in stricter regions like the EU. Second, user location has a large effect on baseline tracking and additional tracking unlocked after clicking "Accept," with EU vantages showing lower pre-consent exposure but large post-consent jumps. Third,

Workshop on Measurements, Attacks, and Defenses for the Web (MADWeb) 2026
27 February 2026, San Diego, CA, USA
ISBN 978-1-970672-06-0
https://dx.doi.org/10.14722/madweb.2026.23010
www.ndss-symposium.org

hosting jurisdiction plays a secondary role: most observed discrimination in banners and tracking behavior is driven by inferred user location rather than website hosting location.

II. Background and Motivation

The web economy relies on embedded 3rd-party services for advertising, analytics, and personalization, enabling cross-site identification and profiling at scale [2, 4, 17]. These services increasingly use request- and redirect-based techniques and 1st-party integrations rather than only classic 3rd-party cookies [1, 6]. As our goal is to *attribute* differences across browser choice, user location, and hosting jurisdiction, we focus on network-visible outcome measures comparable across factors.

We define *tracking exposure* as the set of 3rd-party apex domains contacted during page loads, labeled as *A&T* or "other" using curated public lists, along with the prevalence of consent banners and changes between pre- and post-consent conditions [1]. This definition favors coverage and comparability while acknowledging that some behaviors (payload content, fingerprinting) are not directly observable at network level. Since user studies show 72% choose "Accept all" [18], we treat *no-click* and *accept-all* as separate experimental states.

Browser choice is the lever users control most directly; mainstream browsers differ in default protections (3rd-party cookie restrictions, storage partitioning, anti-fingerprinting) that reshape request-level exposure [19]. *User location* conditions banner presentation and when 3rd-party requests are initiated, with field studies documenting location-dependent consent surfaces [8, 10]. *Hosting jurisdiction* shapes data recipients and applicable legal regimes; large-scale mapping shows cross-border transfers are common [1, 20]. RegTrack manipulates these three factors within a single framework, enabling contrasts respecting their user agency ordering.

Prior work typically vary one factor at a time or rely on setups hard to compare across papers [8, 20, 21]. Our 4x8 factorial design with two consent states reveals how *browser choice*, *user location*, and *hosting jurisdiction* each affect 3rd-party *A&T* exposure when other factors are held constant.

III. Measurement Methodology

To compare how browser, user location, and hosting jurisdiction shape third-party *A&T*, we built RegTrack, a consent-aware factorial framework concurrently crawling websites from eight vantage points using four browsers and two consent states (*no-click* and *accept-all*) (Fig. 1). Each (browser, location, site, consent) configuration runs in a fresh container, recording HTTP requests to HAR files and screenshots.

A. Measurement Variables

Browsers. We evaluate four mainstream desktop browsers with distinct privacy postures: Chrome, Edge, Firefox, and Brave, covering market-dominant and privacy-focused options [19, 22] (market shares: 65.54%, 13.89%, 6.36%, and 1%). All crawls run on Linux in default configuration without extensions to capture "out of the box" exposure [23].

Browsing locations. We deploy crawls from eight regions: California (USA), Ohio (USA), Quebec (Canada), Mumbai (India), Singapore, Frankfurt (Germany), Paris (France), and Dublin (Ireland), covering major privacy regimes including GDPR [24], CCPA [25], PDPA [26], PIPEDA [27], and DPDPA [28]. We treat these locations as representative bundles of legal obligations, enforcement practices, and consent norms, interpreting results at broad regulatory cluster levels (e.g., GDPR vs. opt-out frameworks).

Websites. To avoid regional bias, we combine globally and regionally popular sites: top-1K Tranco [29] augmented with top-100 per country (US, IN, SG, DE, FR, IE) from Cloudflare Radar [30], filtered to top-100K Tranco ranking, yielding 1,005 unique domains. Five additional domains (`ovh.net`, `hotstar.com`, `truecaller.com`, `swiggy.com`, `google.ie`) are from the country lists.

B. Crawling Architecture

Our crawler is built on Browsertime [31], running each visit in a fresh Docker container to ensure no client-side state persists. For each configuration, RegTrack loads the page, records network requests (HAR), and captures a screenshot. We focus analysis on apex domains (stable across visits) while using FQDNs for blocklist matching.[1] We visit each site 10 times per configuration; apex counts converge after roughly five visits, with additional visits providing robustness against transient failures.

C. Cookie Consent Handling

Cookie-consent banners can gate content access and substantially change which third-party requests are issued [8]. For each site and configuration, RegTrack performs two independent passes (repeated 10 times): a *no-click* pass without consent interface interaction, and an *accept-all* pass selecting the most affirmative option when a banner is present. We choose the most affirmative option as it provides an upper bound on tracking exposure and represents realistic user behavior (72% select "OK" or equivalent [18]).

Banner interaction. We automate "accept" clicks using CMP-specific selectors (Didomi, Quantcast, OneTrust, CookieBot) with text-based heuristic fallback for affirmative labels from a curated lexicon.[2] The detection logic is injected via Browsertime's JavaScript hooks and executed across all iframes. We manually audited a sample of pages to remove lexicon entries producing false positives. RegTrack successfully interacts with banners on 91–95% of banner-using sites; remaining cases are treated as no-click.

D. Data Cleaning

We exclude visits where the intended page did not load due to CAPTCHAs, block pages, or network errors. To detect

[1] E.g., we classify `ads.google.com` rather than `google.com`.

[2] We manually curated the lexicon by sampling sites with banners, collecting acceptance strings, and translating them into represented languages.

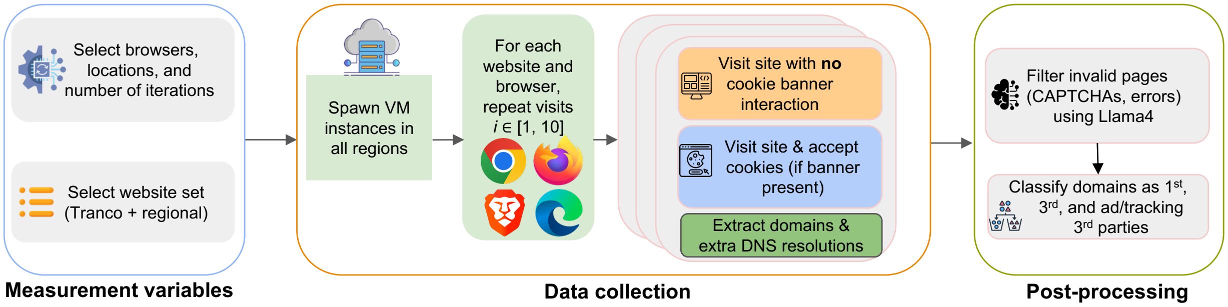

Select browsers, locations, and number of iterations
Select website set (Tranco + regional)
Measurement variables
Spawn VM instances in all regions
For each website and browser, repeat visits i ∈ [1, 10]
Visit site with no cookie banner interaction
Visit site & accept cookies (if banner present)
Extract domains & extra DNS resolutions
Data collection
Filter invalid pages (CAPTCHAs, errors) using Llama4
Classify domains as 1st, 3rd, and ad/tracking 3rd parties
Post-processing

Fig. 1: Overview of RegTrack. We fix browsers, locations, and consent states, then crawl websites from per-location VMs, spawning containers for each (browser, site, consent) configuration to collect HAR files and screenshots. Post-processing filters invalid pages with a multimodal LLM and labels domains as 1st-party, other 3rd-party, or *A&T* 3rd-party using public blocklists.

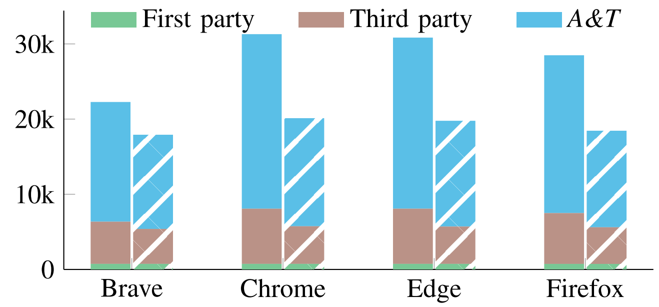


Fig. 2: Total number of *A&T*, by browser, after accepting all cookies in US-Ohio (plain) and France (striped).

such cases at scale, we classify screenshots using open-source vision-language models. After evaluating several models against 400 manually labeled screenshots (§C), we select Llama4, which achieves 99% accuracy (Cohen's $\kappa = 0.94$). A site is included if it loads successfully in $\geq 5/10$ visits for every (browser, location, consent) configuration. Of the initial 1,005 domains, 26% fail this criterion for at least one configuration (often due to CAPTCHAs or geo-blocking). Our final cleaned dataset contains 743 sites.

E. Domain Classification

From each HAR file, we extract all requested domains and classify them. A domain is *first-party* if its apex matches the visited site's pay-level domain; otherwise, it is *third-party*. Third-party domains are classified as *A&T* or "other" using a union of widely used public blocklists (EasyList/EasyPrivacy [32, 33], AdGuard [34], and others; full list in §D). If a domain appears on any list, we flag it as potentially *A&T*-related; remaining third-party domains are labelled "other." Since these blocklists are primarily designed for URL-level filtering, our domain-level matching may introduce measurement error; we discuss this limitation and our mitigation in §V-B.

IV. Data Analysis and Results

We now present our empirical results, organized around the three levers of user agency and the research questions in §I.

A. Browser Choice

Browser choice represents the factor over which users exercise the most direct control. Unlike user location or website hosting jurisdiction, users can freely select and switch between browsers with minimal technical barriers, legal restrictions, or geographic constraints. However, users typically commit to a single browser due to ecosystem lock-in and familiarity, making this choice particularly consequential for their long-term tracking exposure [35].

To understand how browser choice affects the number of third-party *A&T* domains, we compare two regulatory contexts: USA–Ohio, with relatively permissive privacy regulations, and France, governed by the GDPR. In USA–Ohio, Brave offers the lowest exposure, triggering 31% fewer *A&T* domains[3] than Chrome, the browser with the highest tracking levels, as illustrated in Fig. 2. On the other hand, the differences between browsers are less pronounced in France. While Brave still yields the lowest exposure, triggering 13% fewer *A&T* domains than Chrome, the gap between browsers narrows considerably in this GDPR-regulated region.

Browser choice significantly affects third-party *A&T* exposure, but other factors, specifically user location itself, appear to play an important role. We therefore turn to our second research question, examining how user location shapes tracking practices independent of browser selection.

B. Effect of User Location

We now turn to user location, a factor users can only partially influence (for example, via VPNs) but which determines the legal regime that applies to tracking and data protection. We concurrently visit the same 743 websites from eight vantage points spanning North America (Ohio, California, Quebec), Europe (France, Germany, Ireland), and Asia (Mumbai, Singapore). Unless otherwise noted, we focus on Chrome as a high-tracking baseline as shown in §IV-A, and we distinguish between *no-click* and *accept-all* consent states.

Cookie banner prevalence and tracking contribution. Fig. 3 reports, for each vantage point, the fraction of websites displaying a cookie banner (inner circle) and their contribution to total third-party *A&T* requests (outer circle). Banner prevalence varies substantially: EU vantages exhibit the highest rates (60–61%), consistent with GDPR's consent requirements, while permissive or opt-out locations (Ohio, Mumbai)

[3] Our *A&T* classification relies on domain-level blocklist matching, which may over- or under-count tracking in some cases; see §V-B for details.

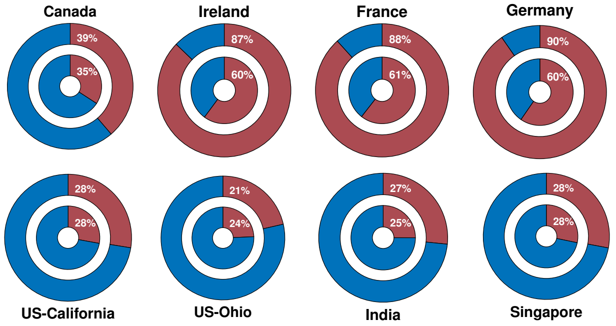


Fig. 3: Comparison between the prevalence of websites displaying a banner (inner circle) and their contribution to total *A&T* traffic (outer circle). Red 🟥 denotes sites with a banner, and blue 🟦 denotes sites without.

show significantly fewer banners, with Singapore and Canada (Quebec) falling in between. This pattern suggests websites selectively deploy consent interfaces based on visitor location. Critically, in EU, bannered sites contribute the majority of observed tracking, indicating these sites are responsible for most tracking activity. In non-EU regions, bannered sites contribute a smaller share and more tracking originates from sites without banners. This suggests consent mechanisms act as gatekeepers for high-intensity tracking deployments, particularly in jurisdictions where consent is legally required.

Effect of consent by region. We next compare exposure in the “no-click” and “accept-all” states for each location (Fig. 4). In GDPR-regulated vantages (FR, DE, IE), baseline tracking in the “no-click” state is low: mean *A&T* counts are significantly smaller than in OH or IN when visiting the same sites without consent. Once users accept all cookies, third-party *A&T* exposure in EU locations increases sharply: FR jumps from 9.3 to 33.9 domains (265% increase). In contrast, Ohio shows a modest increase from 48.5 to 56.2 domains (16% increase) because many trackers load even without explicit consent. This pattern extends to sites without detectable banners: in non-EU regions, these sites contribute substantially higher tracking (37–58 domains) than in EU regions (8–10 domains), further demonstrating how consent requirements shape tracking exposure across the web ecosystem. These results demonstrate that GDPR-style consent requirements provide a stronger privacy baseline until users click “Accept”, at which point much of that advantage erodes. Tracking exposure is also highly skewed: the top 50% of sites contribute 97% of all observed *A&T* domains across all locations (see §F for detailed distribution and category analysis).

C. Cross-border Data Flows

Privacy regulations such as GDPR govern not only *what* data may be collected, but also *where* that data may be sent. For example, GDPR restricts transfers of personal data outside the European Economic Area (EEA) unless the destination country provides an “adequate level of protection” or appropriate safeguards are in place [36]. Having examined how browser, user location, and hosting jurisdiction affect *A&T* volumes and cascades, we now ask a complementary question: where do third-party *A&T* requests go on the network, and how often does *A&T* traffic leave a user’s regulatory region?

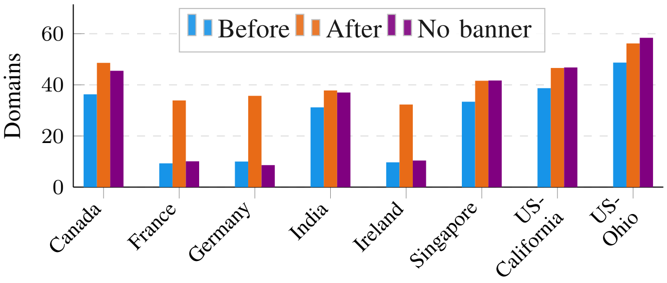
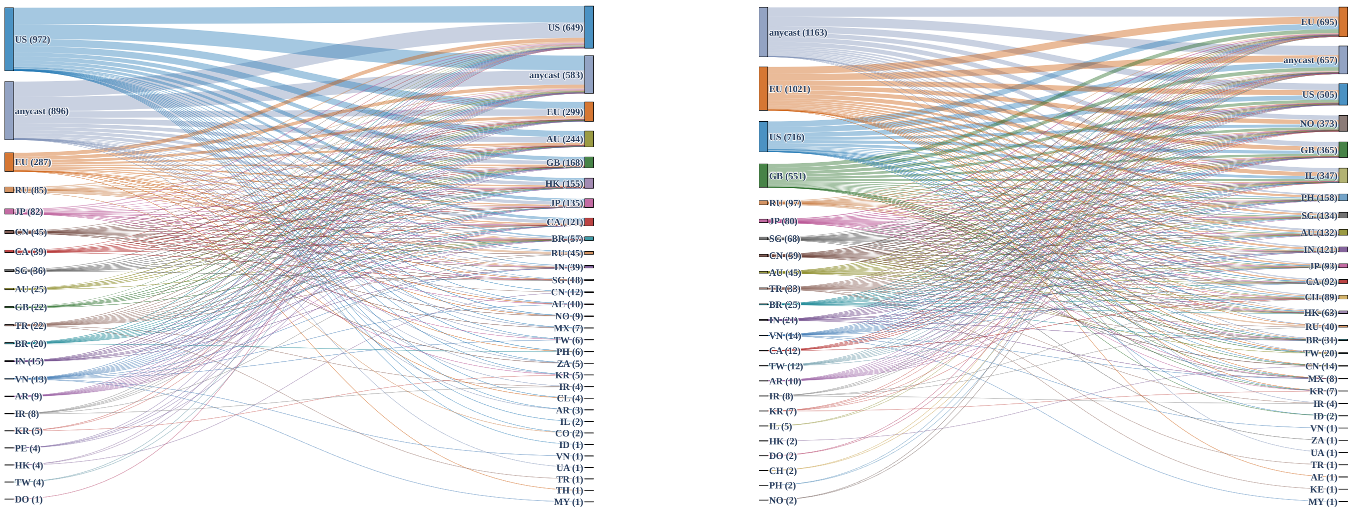


Fig. 4: Average third-party *A&T* domains before/after accepting cookie banners, and from sites with no banner, for Chrome.

IP geolocation. Determining server locations is challenging due to Content Delivery Networks (CDNs), anycast routing, and geographically distributed infrastructure. To obtain robust location estimates, we resolve each third-party apex domain using multiple popular recursive resolvers: Cloudflare (1.1.1.1), Google (8.8.8.8), Quad9 (9.9.9.9), and the default AWS resolver available within our measurement VMs so that we can harvest as many distinct IP addresses as possible for each domain. Using a diverse set of resolvers allows us to capture location-dependent DNS responses and improves coverage. Since IP-geolocation can be inaccurate for various reasons [37, 38, 39], we geolocate each resolved IP address using an ensemble of seven GeoIP databases and apply majority voting to select the most likely country for each IP. This ensemble approach avoids over-reliance on a single database and reduces the impact of individual misclassifications. Anycast prefixes [40] are filtered where possible to avoid ambiguous locations. For each (vantage point, domain) pair, we compare the geolocated server country to the user’s regulatory region (e.g., EEA versus non-EEA, or the user’s own country for non-EU vantages). If *any* resolver returns an IP outside the user’s region, we classify that request as leaving the region. For EU vantages, we treat the entire EEA as a single regulatory region.

Regional containment of tracking traffic. Table I shows that, from every vantage point, third-party *A&T* requests fan out to a large number of destination countries (45–48 unique countries). However, the fraction of requests that remain within the user’s regulatory region varies markedly across locations. EU vantages exhibit substantially higher regional containment: France and Germany keep 56% of requests within the EEA, while Ireland retains 41%. In contrast, non-EU vantages show more varied patterns: US vantages retain the highest domestic containment (69–72%), while India and Singapore show moderate containment (26–31%), and Canada shows the lowest (6.5%). In other words, US users see most tracking processed domestically, EU users see strong regional containment within the EEA, while users in other non-EU regions are considerably more likely to have their data routed to diverse foreign jurisdictions.

To relate this to GDPR’s cross-border transfer rules, we further examine whether EU-origin requests are sent to countries that the European Commission has designated as providing



(a) US-Ohio vantage point (more permissive regulation)

(b) France vantage point (GDPR regulation)

Fig. 5: Cross-border data flows showing first-party (left) and *A&T* (right) server locations based on IP geo-location.

TABLE I: Regional containment of third-party *A&T* requests by vantage point. For FR, DE, and IE, we treat the EEA as a single region, and for OH and CA, the US as a single region. The adequacy column is based on the EU Commission's Adequacy Decision [36].

Vantage point	Unique dest. countries	Requests in the same region (%)	Including Adequacy regions (%)
Canada	45	6.5%	–
France	48	55%	89%
Germany	47	56%	91%
India	47	26%	–
Ireland	48	41%	90%
Singapore	47	31%	–
USA–California	46	69%	–
USA–Ohio	47	72%	–

"adequate protection" under Article 45. When we count requests that stay within the EEA *or* go to adequacy countries (§E), the share of EU requests sent to legally "adequate" destinations increases substantially. This fraction rises to approximately 89% for FR and IE, and 91% for DE. While our GeoIP-based analysis cannot prove legal compliance, it is evident that *A&T* infrastructure for EU users is preferentially located in jurisdictions that GDPR recognizes as offering comparable protection. By contrast, traffic from other regions flows more freely to a broader set of destinations, including countries without comprehensive privacy regulations.

First-party vs third-party geographic distribution. To better understand these regional patterns, we examine how the geographic locations of first-party website servers differ from those of the third-party *A&T* requests they trigger. Fig. 5a and Fig. 5b visualize these cross-border data flows for Ohio and France (additional vantage points in §H). Each flow in the Sankey diagram represents the prevalence of a specific cross-border pattern: the flow thickness indicates how many websites in our dataset exhibit that particular combination of first-party server (primary server IP) location and *A&T* destination region, with each website contributing at most once per unique flow regardless of its tracking volume. When multiple DNS resolvers return different geographic locations for the same domain, we include all resolved locations, as we cannot definitively determine which location the domain actually serves from. This prevalence-based view reveals structural patterns in how websites route data across jurisdictions. Ohio and France exhibit comparable behavior: first-party servers are concentrated in the EU and US, and most *A&T* flows terminate in these regions. This dominance of the US alongside local infrastructure persists across all vantage points (see §H).

D. Hosting Jurisdiction

A site hosting location is beyond user control but may influence legal obligations and operational choices. We assign hosting jurisdiction using primary server IP geolocation, restricting analysis to sites with consistent IP-to-country mappings across vantages, after filtering CDNs and anycast [40]. Fig. 6 shows cookie banner prevalence vs. *A&T* contribution by hosting jurisdiction, aggregated by EU (blue) or non-EU (red) user location (detailed breakdowns in §I). We identify four patterns:

A/ Non-EU baseline: Non-EU users accessing sites hosted outside CN/EU see infrequent banners with low *A&T* contribution from bannered sites, reflecting permissive environments where consent is optional.

B/ EU adaptation: Regardless of hosting jurisdiction, sites present banners frequently to EU users (61–64% for US-hosted sites) with bannered sites contributing 81–83% of *A&T*, demonstrating location-based tailoring for GDPR compliance.

C/ EU-hosted baseline: EU-hosted sites display banners at moderate rates (42–44%) even to non-EU users, suggesting GDPR-compliant sites deploy consent mechanisms uniformly.

D/ CN hosts: CN-hosted sites show uniform 19% banner prevalence regardless of user location, consistent with China's PIPL [41] mandating GDPR-like consent applied globally.

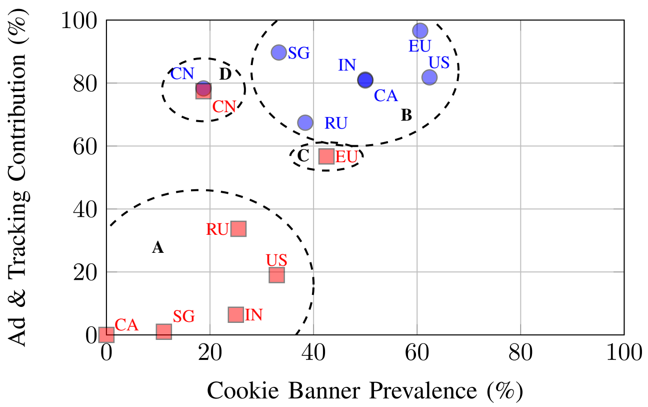


Fig. 6: Nodes represents the hosting locations (e.g., CA, CN, EU, etc.). Blue circles (●) are hosts accessed from EU whereas red squares (■) are accessed by non-EU vantages. The x-axis shows the prevalence of cookie banners, the y-axis shows the contribution of websites with cookie banners to the total *A&T* in their location. We identify 4 regions in the figure (A, B, C, and D). (More detailed data in appendix §I.)

V. Discussion

Our results show browser choice, browsing location and its regulatory framework, and hosting jurisdiction each affect different aspects of the *A&T* ecosystem: baseline volume, cookie-consent behavior, cascade structure, and cross-border data flows. We discuss practical implications and limitations.

A. Implications

Privacy-focused browsers reduce *A&T* exposure by 30% in permissive regions but offer smaller gains where GDPR already lowers baseline *A&T*. Accepting cookies erodes this advantage: in FR, *A&T* domains per site increase 3x after consent vs. smaller increases in OH. GDPR-regulated locations exhibit 80% fewer *A&T* domains pre-consent and keep 90% of requests within EEA/adequacy countries, suggesting opt-in consent and cross-border rules meaningfully reduce tracking and contain data flows. However, dark patterns steering users toward “Accept all” can undo much of this benefit. Location-based consent discrimination is widespread and easily measurable: websites present banners 2–3x more frequently to EU visitors, patterns amenable to regulatory scrutiny.

B. Limitations

As with any empirical study, our work involves trade-offs limiting generalizability. We focus on 1,005 popular websites and 8 vantages in North America, Europe, and Asia, potentially missing smaller or localized sites and some countries within each region. Our *A&T* blocklists (e.g., EasyList/EasyPrivacy [32, 33]) are designed for URL-level ad blocking; applying them at the domain level may over-count (flagging benign requests to listed domains) or under-count (missing path-specific first-party tracking rules). We mitigate this using Wally3K’s curated lists [42, 43], which filter entries unsuitable for domain-level matching, though some error may remain. This also explains why *A&T* domains appear in Brave crawls: Brave blocks specific URL patterns, while our classification flags any request to a listed domain. Crucially, comparative analysis across browsers and regions remains valid as the methodology is applied uniformly. Our lexicon-driven cookie-banner detection has high but imperfect coverage, and we model only two consent states; we measure tracking behavior rather than legal compliance. Our network-level observations may miss CNAME cloaking [44] and do not account for data volume or sensitivity; GeoIP-based cross-border analysis may contain errors from CDNs and anycast. Our factorial design observes associations rather than proving causal effects; site mix, business models, and regional ad markets may contribute. Finally, our setup cost $1K+ USD, generated 1.5 TB data, and required 2.5 weeks, limiting temporal repetition and additional dimensions (e.g., mobile browsers).

VI. Related Work

We focus on recent work that evaluates tracking under regulatory constraints, incorporates geographic scope, or foregrounds measurement methodology.

Studies operationalize GDPR/CCPA requirements through measurements of user-facing behavior: Sørensen et al. [45] quantify third-party presence before/after GDPR; Sanchez-Rola et al. [46] examine tracking persistence after opt-out; Liu et al. [47] audit consent choices across GDPR/CCPA contexts; Hausladen et al. [48] evaluate GPC signal compliance. Iordanou et al. [20] characterize cross-border tracking endpoints for EU users; Vallina et al. [49] study tracking across multiple vantages; Singh et al. [50] broaden coverage to 23 Global South countries. Urban et al. [1] measure third-party dynamics beyond landing pages; Stafeev et al. [51] systematize crawling design space; Hantke et al. [52] propose reproducible measurement tooling. RegTrack builds on these insights by jointly varying browser, user location, hosting jurisdiction, and consent state within a controlled factorial design. A chronological summary table of related work in comparison to RegTrack is in Appendix §K.

VII. Conclusion

We examined how third-party *A&T* exposure varies across browser, location, and hosting jurisdiction through synchronized, consent-aware measurements of 743 sites across 8 vantage points, 4 browsers, and 2 consent states. Browsing location is the strongest predictor, influencing pre-consent baselines, consent interface prevalence, and post-consent *A&T* levels. Browser choice provides context-dependent leverage, with larger gains in permissive settings. Hosting jurisdiction is weaker, suggesting sites adapt to inferred user location rather than hosting location. EU vantages show higher regional containment of *A&T* traffic, especially to EEA and adequacy destinations. Our results show user-controllable choices matter, but structural context, location-conditioned consent gating and region-specific infrastructure–often dominates, providing measurable compliance signals for regulators and emphasizing the need to treat browser, location, and consent state as first-class experimental variables.

Acknowledgment

We are grateful to the anonymous reviewers for their constructive feedback and guidance in shaping this paper. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC). Nous remercions le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) de son soutien. We acknowledge the support of the Canadian Foundation for Innovation (CFI). Nous remercions la Fondation canadienne pour l'innovation (FCI) de son soutien. Finally, we acknowledge the support of the British Columbia Knowledge Development Fund (BCKDF), UBC Advanced Research Computing (ARC), and Dell Computer. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not reflect those of the sponsors.

References

[1] T. Urban, M. Degeling, T. Holz, and N. Pohlmann, "Beyond the Front Page: Measuring Third Party Dynamics in the Field," in *The Web Conference (WWW'20)*. ACM, 2020.

[2] T. Bujlow, V. Carela-Español, J. Sole-Pareta, and P. Barlet-Ros, "A Survey on Web Tracking: Mechanisms, Implications, and Defenses," *Proceedings of the IEEE*, 2017.

[3] N. Demir, D. Theis, T. Urban, and N. Pohlmann, "Towards Understanding First-Party Cookie Tracking in the Field," *German Informatics Society – Sicherheit, Schutz und Zuverlässigkeit*, 2022.

[4] J. R. Mayer and J. C. Mitchell, "Third-Party Web Tracking: Policy and Technology," in *Symposium on Security and Privacy (S&P'12)*. IEEE, 2012.

[5] F. Roesner, C. Rovillos, T. Kohno, and D. Wetherall, "ShareMeNot: Balancing Privacy and Functionality of Third-Party Social Widgets," *USENIX :login;*, 2012.

[6] U. Iqbal, C. Wolfe, C. Nguyen, S. Englehardt, and Z. Shafiq, "Khaleesi: Breaker of Advertising and Tracking Request Chains," in *Security Symposium (USENIX Sec'22)*. USENIX, 2022.

[7] S. Siby, U. Iqbal, S. Englehardt, Z. Shafiq, and C. Troncoso, "WebGraph: Capturing Advertising and Tracking Information Flows for Robust Blocking," in *Security Symposium (USENIX Sec'22)*. USENIX, 2022.

[8] C. Utz, M. Degeling, S. Fahl, F. Schaub, and T. Holz, "(Un)informed Consent: Studying GDPR Consent Notices in the Field," in *Conference on Computer and Communications Security (CCS'19)*. ACM, 2019.

[9] C. Matte, C. T. Santos, and N. Bielova, "Purposes in IAB Europe's TCF: Which Legal Basis and How Are They Used by Advertisers?" in *Annual Privacy Forum*. Springer, 2020.

[10] H. Hosseini, C. Utz, M. Degeling, and T. Hupperich, "A Bilingual Longitudinal Analysis of Privacy Policies Measuring the Impacts of the GDPR and the CCPA/CPRA," *Privacy Enhancing Technologies Symposium (PETS'24)*, 2024.

[11] C. T. Santos, M. Nouwens, M. Tóth, N. Bielova, and V. Roca, "Consent Management Platforms under the GDPR: processors and/or controllers?" in *Annual Privacy Forum*. Springer, 2021.

[12] V. H. Tran, A. Mehrotra, M. Chetty, N. Feamster, J. Frankenreiter, and L. J. Strahilevitz, "Measuring Compliance with the California Consumer Privacy Act Over Space and Time," in *CHI Conference on Human Factors in Computing Systems (CHI'24)*. ACM, 2024.

[13] V. H. Tran, A. Mehrotra, R. Sharma, M. Chetty, N. Feamster, J. Frankenreiter, and L. J. Strahilevitz, "Dark Patterns in the Opt-Out Process and Compliance with the California Consumer Privacy Act (CCPA)," in *CHI Conference on Human Factors in Computing Systems (CHI'25)*. ACM, 2025.

[14] R. van Eijk, H. Asghari, P. Winter, and A. Narayanan, "The Impact of User Location on Cookie Notices (Inside and Outside of the European Union)," in *Security & Privacy Workshop on Technology and Consumer Protection (ConPro'19)*. IEEE, 2019.

[15] M. Degeling, C. Utz, C. Lentzsch, H. Hosseini, F. Schaub, and T. Holz, "We Value Your Privacy ... Now Take Some Cookies," *Springer Informatik Spektrum*, 2018.

[16] V. Mishra, P. Laperdrix, A. Vastel, W. Rudametkin, R. Rouvoy, and M. Lopatka, "Don't Count Me Out: On the Relevance of IP Address in the Tracking Ecosystem," in *The Web Conference (WWW'20)*. ACM, 2020.

[17] S. Englehardt and A. Narayanan, "Online Tracking: A 1-million-site Measurement and Analysis," in *Conference on Computer and Communications Security (CCS'16)*. ACM, 2016.

[18] H. Habib, M. Li, E. Young, and L. Cranor, ""Okay, whatever": An Evaluation of Cookie Consent Interfaces," in *CHI Conference on Human Factors in Computing Systems (CHI'22)*. ACM, 2022.

[19] R. Madhusudhan and S. V. Surashe, "Privacy and Security Comparison of Web Browsers: A Review," in *International Conference on Advanced Information Networking and Applications*. Springer, 2022.

[20] C. Iordanou, G. Smaragdakis, I. Poese, and N. Laoutaris, "Tracing Cross Border Web Tracking," in *Internet Measurement Conference (IMC'18)*. ACM, 2018.

[21] D. Cassel, S.-C. Lin, A. Buraggina, W. Wang, A. Zhang, L. Bauer, H.-C. Hsiao, L. Jia, and T. Libert, "OmniCrawl: Comprehensive Measurement of Web Tracking With Real Desktop and Mobile Browsers," in *Privacy Enhancing Technologies Symposium (PETS'22)*, 2022.

[22] "Global Desktop Browser Market Share," https://gs.statcounter.com/browser-market-share/desktop/worldwide, [Accessed February 19, 2026].

[23] H. Habib, "Evaluating the Usability of Privacy Choice Mechanisms," in *Symposium on Usable Privacy and Security (SOUPS'21)*. USENIX, 2021.

[24] European Parliament and Council of the European Union. (2016) General Data Protection Regulation (GDPR) –

Official Legal Text. Regulation (EU) 2016/679, OJ L 119, 4 May 2016.

[25] California State Legislature, "California Consumer Privacy Act (CCPA)," 2018, enacted 2018, amended by the California Privacy Rights Act (CPRA) in 2020; updated March 13, 2024.

[26] Parliament of Singapore, "Personal Data Protection Act 2012 (2021 Revised Edition)," 2012, no. 26 of 2012. Incorporates all amendments up to and including 1 December 2021.

[27] Parliament of Canada, "Personal Information Protection and Electronic Documents Act (PIPEDA)," 2000, s.C. 2000, c. 5. Assented to April 13, 2000.

[28] Government of India, "Digital Personal Data Protection Act, 2023," 2023, act No. 22 of 2023, Ministry of Law and Justice (Legislative Department), New Delhi.

[29] V. Le Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczynski, and W. Joosen, "TRANCO: A Research-Oriented Top Sites Ranking Hardened Against Manipulation," in *Network and Distributed System Security (NDSS'19)*. Internet Society, 2019.

[30] Cloudflare, "Cloudflare radar: Domain rankings worldwide," https://radar.cloudflare.com/domains, 2025, [Accessed February 19, 2026].

[31] sitespeed.io, "Introduction to Browsertime," https://www.sitespeed.io/documentation/browsertime/introduction/, Mar 2025, [Accessed February 19, 2026].

[32] Firebog, "Easylist-Default," https://v.firebog.net/hosts/Easylist.txt, [Accessed February 19, 2026].

[33] T. firebog, "Easylist-Trackers," https://v.firebog.net/hosts/Easyprivacy.txt, [Accessed February 19, 2026].

[34] AdguardTeam, "Adguard DNS Filter List," https://raw.githubusercontent.com/AdguardTeam/FiltersRegistry/master/filters/filter_15_DnsFilter/filter.txt, [Accessed February 19, 2026].

[35] "Beyond Choice Screens: Exploring browser choice design interventions – Mozilla Research — research.mozilla.org," https://research.mozilla.org/browser-competition/remedyconcepts/, [Accessed February 19, 2026].

[36] European Commission, "Adequacy Decisions," https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en, [Accessed February 19, 2026].

[37] P. Gill, Y. Ganjali, and B. Wong, "Dude, Where's That IP? Circumventing Measurement-based IP Geolocation," in *Security Symposium (USENIX Sec'10)*. USENIX, 2010.

[38] Z. Weinberg, S. Cho, N. Christin, V. Sekar, and P. Gill, "How to Catch when Proxies Lie: Verifying the Physical Locations of Network Proxies with Active Geolocation," in *Internet Measurement Conference (IMC'18)*. ACM, 2018.

[39] O. Darwich, H. Rimlinger, M. Dreyfus, M. Gouel, and K. Vermeulen, "Replication: Towards a Publicly Available Internet Scale IP Geolocation Dataset," in *Internet Measurement Conference (IMC'23)*. ACM, 2023.

[40] bgp.tools, "Anycatch v4 prefixes," https://github.com/bgptools/anycast-prefixes/blob/master/anycatch-v4-prefixes.txt, 2025, [Accessed February 19, 2026].

[41] "Personal Information Protection Law of the People's Republic of China - PIPL," https://personalinformationprotectionlaw.com/, [Accessed February 19, 2026].

[42] Firebog, "Wally3k github," https://github.com/WaLLy3K/wally3k.github.io, [Accessed February 19, 2026].

[43] ——, "Firebog website," https://firebog.net/, [Accessed February 19, 2026].

[44] D. L. Rebekah Houser, "CNAME Cloaking: Disguising Third Parties Through the DNS," https://unit42.paloaltonetworks.com/cname-cloaking/, 2022, [Accessed February 19, 2026].

[45] J. Sørensen and S. Kosta, "Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites," in *The Web Conference (WWW'19)*. ACM, 2019.

[46] I. Sanchez-Rola, M. Dell'Amico, P. Kotzias, D. Balzarotti, L. Bilge, P.-A. Vervier, and I. Santos, "Can I Opt Out Yet?: GDPR and the Global Illusion of Cookie Control," in *Asia Conference on Computer and Communications Security (AsiaCCS'19)*. ACM, 2019.

[47] Z. Liu, U. Iqbal, and N. Saxena, "Opted Out, Yet Tracked: Are Regulations Enough to Protect Your Privacy?" in *Privacy Enhancing Technologies Symposium (PETS'24)*, 2024.

[48] K. Hausladen, O. Wang, S. Eng, J. Wang, F. Wijaya, M. May, and S. Zimmeck, "Websites' Global Privacy Control Compliance at Scale and over Time," in *Security Symposium (USENIX Sec'25)*. USENIX, 2025.

[49] P. Vallina, Á. Feal, J. Gamba, N. Vallina-Rodriguez, and A. F. Anta, "Tales from the Porn: A Comprehensive Privacy Analysis of the Web Porn Ecosystem," in *Internet Measurement Conference (IMC'19)*. ACM, 2019.

[50] S. K. Singh, R. Ricci, and A. Gamero-Garrido, "Where in the World Are My Trackers? Mapping Web Tracking Flow Across Diverse Geographic Regions," in *Internet Measurement Conference (IMC'25)*. ACM, 2025.

[51] A. Stafeev and G. Pellegrino, "SoK: State of the Krawlers–Evaluating the Effectiveness of Crawling Algorithms for Web Security Measurements," in *Security Symposium (USENIX Sec'24)*. USENIX, 2024.

[52] F. Hantke, P. Snyder, H. Haddadi, and B. Stock, "Web Execution Bundles: Reproducible, Accurate, and Archivable Web Measurements," in *Security Symposium (USENIX Sec'25)*. USENIX, 2025.

[53] S. Black, "Hosts," https://raw.githubusercontent.com/StevenBlack/hosts/master/hosts, [Accessed February 19, 2026].

[54] T. Firebog, "Admiral Block List," https://v.firebog.net/hosts/Admiral.txt, 2025, [Accessed February 19, 2026].

[55] A. ND, “Adservers,” https://raw.githubusercontent.com/anudeepND/blacklist/master/adservers.txt, [Accessed February 19, 2026].

[56] T. Firebog, “Prigent-Ads,” https://v.firebog.net/hosts/Prigent-Ads.txt, [Accessed February 19, 2026].

[57] T. Frogeye, “First Party Tracker Hosts,” https://hostfiles.frogeye.fr/firstparty-trackers-hosts.txt, [Accessed February 19, 2026].

Appendix

A. Data and Code Availability

To support reproducibility and future research, we publicly release our dataset (raw HAR files, processed tracking-domain classifications, and aggregated statistics) and the RegTrack source code, analysis pipelines, and visualization scripts at https://github.com/ubc-spg/RegTrack.

B. Use of Generative AI

We used generative AI tools (ChatGPT, Claude) to assist with specific technical and editorial tasks during this research, and we document their use here for transparency. For writing, we relied on these tools to improve sentence clarity, correct grammatical errors, rephrase awkward constructions to enhance readability, and help maintain consistent terminology throughout the paper. For code, we used them to generate matplotlib plotting routines for data visualizations and to draft wrapper functions for data processing pipelines. All AI-generated code and text were reviewed, validated, and, where necessary, modified by the authors to ensure accuracy and appropriateness. Finally, we used AI tools to automatically classify invalid pages, with more details provided in the next subsection (§C).

C. Finding Invalid Pages using Llama4

We use vision-language models (VLMs) to automatically identify invalid pages in our crawl data. Invalid pages include CAPTCHAs, error pages, security warnings, and connection failures that would skew our tracking measurements. To select the best model, we evaluated five VLMs against 400 manually labeled screenshots (Table II). Llama4 achieves the highest accuracy (99%) and Cohen’s κ (0.94), indicating near-perfect agreement with human labels. Fig. 7 shows the prompt used to classify each captured page. We exclude websites whose landing page is classified as invalid (result = 1) in more than 50% of visits in any browser-location configuration, ensuring that our tracking measurements reflect actual website behavior rather than error states.

TABLE II: Performance of vision-language models for web-page screenshot classification (success vs. failure).

Model	Acc.	Prec.	Rec.	F1	Kappa
llama4	0.99	0.98	0.93	0.95	0.94
Qwen2.5 VL	0.98	0.95	0.93	0.94	0.93
llama3.2-vision	0.91	0.92	0.52	0.67	0.62
LLaVA 7B	0.91	0.76	0.63	0.69	0.63
Gemma3	0.32	0.19	0.97	0.32	0.06

```
Does this webpage show ANY of these invalid
    page indicators:
- CAPTCHA verification or "I’m not a robot"
    checkbox
- "Please verify you are human" messages
- Security warnings or "Potential Security
    Risk"
- Connection errors like "This site can’t be
    reached" or "can’t reach this page"
- DNS errors or technical error codes
- 404/403/500 error messages or "Not Found"
- Generic error messages like "Something went
    wrong" or "We’re having trouble"
- "Unable to connect" or connection timeout
    messages
- Blank or mostly empty pages with minimal
    content
- Browser error pages or access restrictions
- "Access Denied" or permission error messages
- Security warnings like "This site has been
    reported as unsafe"
- Technical service pages showing raw data or
    configuration
- Completely blank white/empty pages with no
    content
- Microsoft Defender or browser security
    warnings

Answer in JSON format:
{
  "result": 1 or 0 (1 for YES, 0 for NO),
  "reason": "brief explanation in 30 words or
      less"
}
```

Fig. 7: LLM prompt for invalid page detection.

D. Blocklists Used for A&T Classification

Table III lists the blocklists used to classify third-party domains as advertising and tracking (*A&T*) as well as their descriptions.

TABLE III: Blocklists used for *A&T* identification.

Blocklist	Description
AdGuard DNS [34]	Default blocklist for AdGuard DNS service
StevenBlack [53]	Default blocklist for Pi-hole ad blocking
LanikSJ Admiral [54]	Blocks ad-blocker detectors like Admiral
Anudeep AdServers [55]	Ad server list by AnudeepND (via NextDNS)
EasyList Default [32]	Primary ad blocking list for AdBlock, AdGuard, uBlock Origin
EasyList Privacy [33]	Tracker blocking companion to EasyList
Firebog Prigent [56]	Ad list by Fabrice Prigent
Frogeye [57]	Tracker list by Geoffrey Frogeye

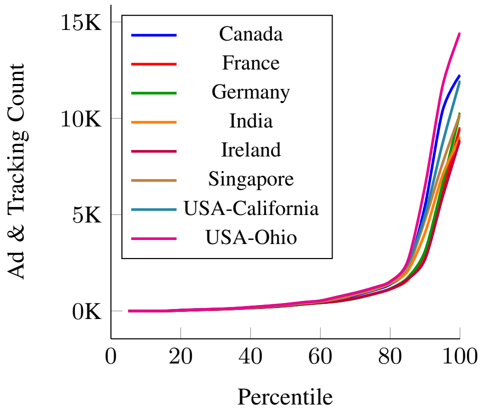


Fig. 8: CDF of 3rd-party *A&T* domains across website percentiles (5-point increments) by location. X=100 represents the 95th-100th percentile. The steep right-to-left decline demonstrates that *A&T* is concentrated in higher percentile websites.

E. GDPR Adequacy Destinations

Under GDPR Article 45, the European Commission has issued adequacy decisions [36] for the following countries and territories, allowing personal data to flow from the EU to them without additional safeguards: Andorra, Argentina, Canada (commercial organizations), the Faroe Islands, Guernsey, Israel, the Isle of Man, Japan, Jersey, New Zealand, the Republic of Korea, Switzerland, the United Kingdom, Uruguay, and the United States of America (under the EU–U.S. Data Privacy Framework).

In our server-IP analysis (§IV-C), we therefore treat data transfers to these jurisdictions as compliant with the GDPR's cross-border transfer restrictions.

F. Tracking Distribution and Category Analysis

Skewed distribution of tracking across sites. Tracking exposure is highly skewed across sites: most sites contact relatively few third-party *A&T* domains, while a small fraction contact dozens or even hundreds. Fig. 8 illustrates this concentration by showing the distribution of third-party *A&T* domains across website percentiles. The top 50% of sites contribute roughly 97% of all observed third-party *A&T* apex domains in our Chrome accept-all configuration. This pattern holds consistently across all vantage points.

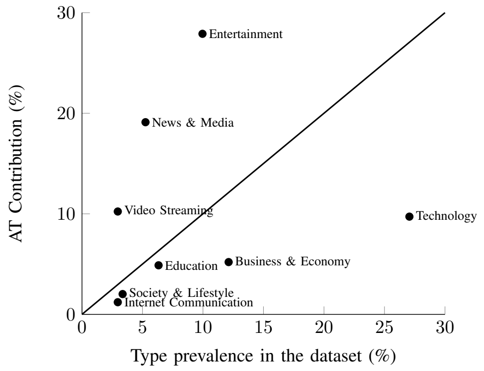


Fig. 9: Category prevalence vs. *A&T* contribution in US-Ohio using Chrome. Categories above the diagonal contribute disproportionately to *A&T*.

Which sites track the most? To understand which website types contribute more to *A&T*, we group sites by category. Fig. 9 shows, for Ohio with Chrome, the relationship between each category's prevalence in our dataset and its contribution to total *A&T* requests. Categories above the diagonal contribute disproportionately high tracking: Entertainment (10% of sites, 28% of *A&T*), News & Media (5% of sites, 19% of *A&T*), and Video Streaming (3% of sites, 10% of *A&T*).

G. Measurement Infrastructure

Our measurement infrastructure consists of distributed crawling nodes and a centralized orchestration layer.

Regional Crawling Nodes. We deploy AWS EC2 instances in eight geographic regions to perform web crawls:

- **North America:** Ohio (us-east-2), California (us-west-1), Canada (ca-central-1)
- **Europe:** Ireland (eu-west-1), Germany (eu-central-1), France (eu-west-3)
- **Asia:** Singapore (ap-southeast-1), India (ap-south-1)

All crawling nodes use the same instance type to ensure measurement consistency with the following specifications:

- **Instance type:** m6a.32xlarge
- **CPU:** 128 vCPUs
- **Memory:** 512 GB RAM
- **Network:** 50 Gbps
- **Disk:** 300 GB gp3 SSD with 20k IOPS and 1 GB/s bandwidth
- **OS:** Ubuntu 22.04 LTS

Each node runs Browsertime to automate browser interactions and collect HAR files that record all network requests.

Central Orchestration Server. We use a Dell PowerEdge R750 server to coordinate crawls across all regions and process the collected data:

- **Model:** Dell PowerEdge R750
- **CPU:** 64 cores (2x Intel Xeon Gold 6326 @ 2.90 GHz)
- **Memory:** 1024 GB RAM

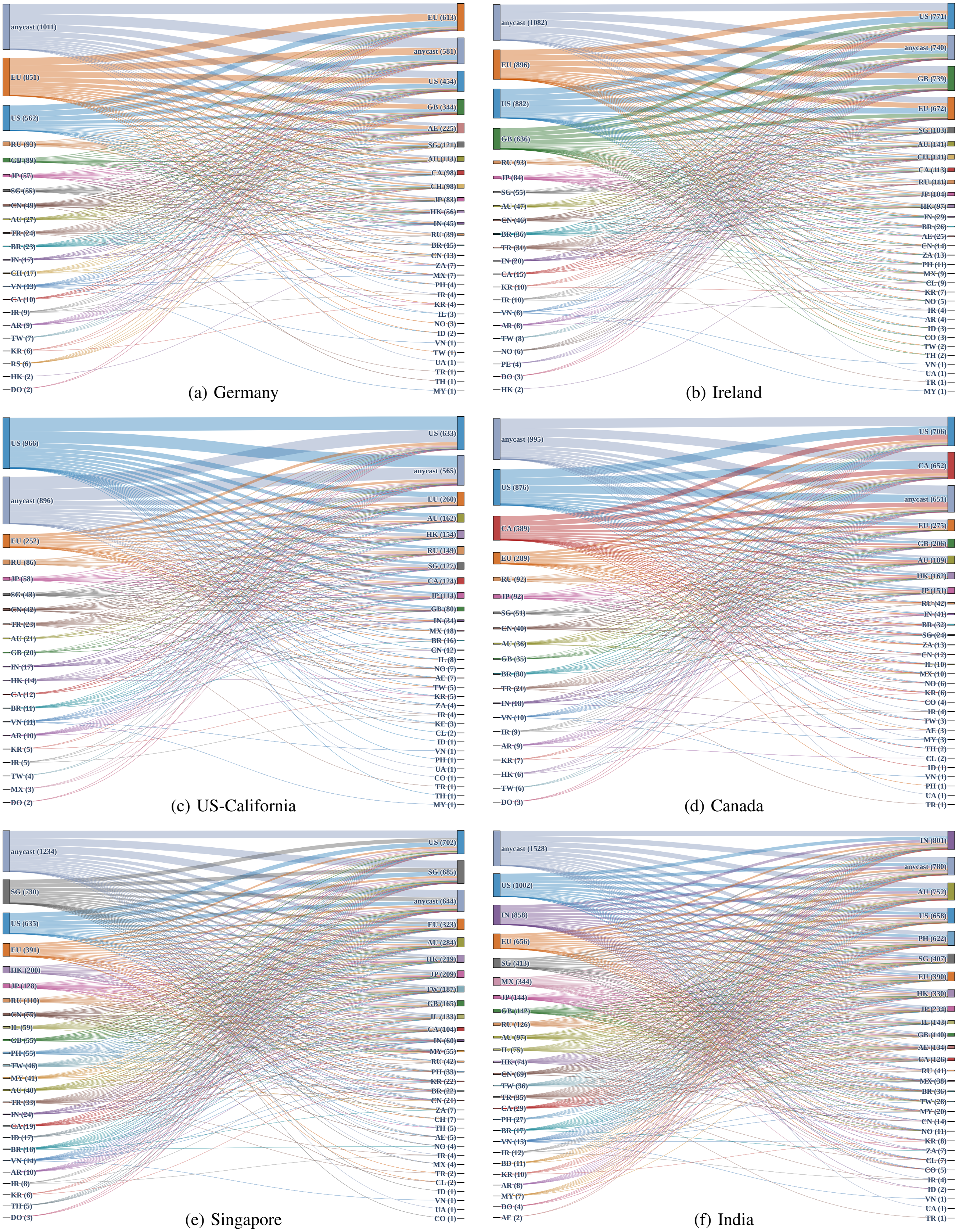
This server schedules crawls, monitors progress across regions, collects HAR files from the crawling nodes, and performs initial aggregation and full analysis.

Invalid Page Classification Server. We use a GPU-equipped Dell PowerEdge R750 server for Llama4-based invalid-page classification:

- **Model:** Dell PowerEdge R750
- **CPU:** 64 cores (2x Intel Xeon Gold 6326 @ 2.90 GHz)
- **Memory:** 1024 GB RAM
- **GPU:** NVIDIA A100 PCIe 80 GB

H. Additional Cross-Border Data Flow Diagrams

Fig. 5 in the main text shows cross-border data flows for Ohio and France. Here we present the corresponding diagrams for the remaining vantage points Fig. 10.



(a) Germany

(b) Ireland

(c) US-California

(d) Canada

(e) Singapore

(f) India

Fig. 10: Cross-border data flows showing first-party (left) and *A&T* (right) server locations based on IP geo-location, across remaining six vantage points: Germany, Ireland, US-California, Canada, Singapore, and India.

I. Detailed Host Location Statistics

Table IV and Table V show respectively the prevalence of cookie banners and *A&T* contribution across host locations and vantage points.

TABLE IV: Share of sites (in %) that display a cookie banner, broken down by hosting jurisdiction (rows) and user vantage point (columns). EU vantages are grouped on the left, non-EU vantages on the right.

	EU vantages			**Non-EU vantages**				
Host	FR	DE	IE	OH	CA	CAD	IN	SG
US(118)	61%	64%	63%	27%	31%	40%	33%	33%
EU(55)	62%	60%	60%	42%	44%	44%	42%	42%
SG(9)	33%	33%	33%	11%	11%	11%	11%	11%
IN(4)	50%	50%	50%	25%	25%	25%	25%	25%
CA(2)	50%	50%	50%	0%	0%	0%	0%	0%
RU(33)	39%	39%	36%	30%	30%	24%	18%	24%
CN(16)	19%	19%	19%	19%	19%	19%	19%	19%

TABLE V: AT contribution % from sites that present cookie banner in different regions

	EU vantages			**Non-EU vantages**				
Host	FR	DE	IE	OH	CA	CAD	IN	SG
US(118)	81%	83%	82%	10%	17%	33%	18%	18%
EU(55)	97%	97%	96%	55%	58%	59%	57%	56%
SG(9)	90%	90%	88%	1%	1%	1%	2%	1%
IN(4)	81%	82%	81%	6%	6%	5%	7%	9%
CA(2)	80%	80%	82%	0%	0%	0%	0%	0%
RU(33)	69%	69%	64%	35%	37%	33%	30%	34%
CN(16)	83%	92%	60%	76%	77%	89%	70%	75%

J. Concentration and Regional Clustering of Heavy Trackers

We can also view this skewness from the perspective of *who* contributes most of the *A&T* volume. Given this heavy concentration, we next ask whether the same sites dominate everywhere or whether the identity of heavy trackers changes by region. For each location, we take the top 10% of sites by *A&T* contribution and measure the overlap between these sets across locations and show it in Table VI. We focus on the top 10% because this subset captures the heaviest contributors while still leaving at least dozens of sites per location, which stabilizes overlap estimates; we observed qualitatively similar regional clustering when using other top-X% thresholds. We observe substantial overlap overall, suggesting a stable core of tracking-intensive sites that appear near the top of the ranking in many regions, but also clear regional clustering among the very heaviest contributors. Our analysis shows that EU vantage points (FR, DE, IE) have a similarity of 0.9-0.92 among each other, while India and Singapore show 0.82 similarity with each other, which is higher than either EU or North America. From North American vantage points, similarity is higher within North America than across regions; Asia also shows strong internal similarity. This pattern is consistent with a picture in which a common global set of large sites dominates tracking, but their relative intensity and ranking vary by region, indicating location-aware advertising and analytics deployments.

TABLE VI: Jaccard similarity of top 10% websites contributing to third-party *A&T* across locations. High similarity within regional groups (NA, EU, Asia) indicates regional clustering of high-tracking websites.

	North America			**Asia**		**EU**		
	CAD	OH	CA	IN	SG	FR	DE	IE
CAD	1.00	0.80	0.74	0.74	0.78	0.59	0.59	0.59
OH	0.80	1.00	0.80	0.74	0.76	0.57	0.57	0.56
CA	0.74	0.80	1.00	0.80	0.80	0.68	0.70	0.66
IN	0.74	0.74	0.80	1.00	0.87	0.70	0.66	0.68
SG	0.78	0.76	0.80	0.87	1.00	0.74	0.72	0.72
FR	0.59	0.57	0.68	0.70	0.74	1.00	0.90	0.92
DE	0.59	0.57	0.70	0.66	0.72	0.90	1.00	0.90
IE	0.59	0.56	0.66	0.68	0.72	0.92	0.90	1.00

K. Chronological Summary of Related Work

Table VII summarizes closely related studies along common study-design axes. For each work, we report its primary objective, the number of browsers (or measurement client configurations), the *browsing/measurement locations* from which measurements were conducted (i.e., vantage points), and the reported scale in number of sites (or the closest equivalent when the study uses a different primary unit).

TABLE VII: Regulatory and measurement-focused related work summarized along study-design axes. “Geographic scope” refers to *browsing/measurement locations* (vantage points) used to collect data. “# Sites” is reported as the number of sites (or the closest equivalent when the study uses observational user data).

Prior study	**Year**	**Focus/Objective**	**Browsers**	**Geographic scope**	**# Sites**
Iordanou et al. [20]	2018	Cross-border tracking endpoints for EU users by mapping where third-party tracking communications terminate (destination infrastructure).	1 (Chrome)	Observational users (multi-country; EU28 subset analyzed; not controlled vantages).	5,693
Sørensen et al. [45]	2019	Third-party presence before vs. after GDPR enforcement using longitudinal crawling to quantify changes in third-party inclusion over time.	1 (Firefox)	Single EU-based browsing location (crawling VM in EU; not location-varied).	1,250
Sanchez-Rola et al. [46]	2019	Tracking persistence after opt-out attempts: contrasts user-facing opt-out/consent choices with observed cookies and tracking activity.	1 (Chrome)	3 browsing locations (Spain; France; Ireland).	2,000
Vallina et al. [49]	2019	Privacy practices and tracking in the adult-web ecosystem under GDPR, including measurement of tracking technologies and compliance signals.	2 (Firefox and Chrome)	Spain (physical) + VPN vantages in other EU member states + SG/IN/RU/US/UK.	6,843
Urban et al. [1]	2020	Third-party dynamics “in the field”: measures how third parties appear beyond landing pages and characterizes embedding patterns at scale.	1	3 browsing locations (Europe/DE; North America/US; Asia/JP).	10,000
Liu et al. [47]	2024	Consent/CMP auditing under GDPR vs. CCPA contexts: tests whether opt-out/consent choices propagate to downstream advertising behavior.	1 (Firefox)	2 browsing locations (EU-/Frankfurt; US/Northern California).	352
Stafeev et al. [51]	2024	Crawling methodology and measurement design space (SoK): evaluates how crawler strategy affects coverage and conclusions in web measurements.	1	Not location-focused (no explicit browsing-location variation emphasized).	2,000
Hantke et al. [52]	2025	Web measurement accuracy and reproducibility: proposes recording/archiving and replay to support reproducible web archive construction and measurement fidelity.	1	Not location-focused (tooling contribution; browsing location not a primary axis).	10,000
Hausladen et al. [48]	2025	CCPA/GPC opt-out compliance at scale and over time: evaluates whether sites honor Global Privacy Control signals in practice.	1 (Firefox)	Single browsing location (California via VPN; not location-varied).	11,708
Singh et al. [50]	2025	Tracker exposure and related data flows in under-measured regions (Global South) using distributed, volunteer-based measurements across many countries.	1 (Chrome)	23 browsing locations (countries) across Africa/Asia/Europe/N. America/Oceania/S. America.	≈100 per country
RegTrack	**2025**	Cross-jurisdictional tracking under browser choice, user location, hosting jurisdiction, and consent state within a controlled factorial measurement design.	**4**	**8 browsing locations (NA/EU/Asia).**	**743**