



OmniCrawl: Comprehensive Measurement of Web Tracking with Real Desktop and Mobile Browsers

Darion Cassel¹, Su-Chin Lin², Alessio Buraggina³, William Wang⁴, Andrew Zhang⁵, Lujo Bauer¹,
Hsu-Chun Hsiao², Limin Jia¹, and Timothy Libert⁶

¹ Carnegie Mellon University

⁴ University of Chicago

² National Taiwan University

⁵ University of Illinois Urbana-Champaign

³ University of Miami

⁶ Google



Carnegie
Mellon
University

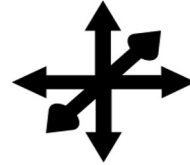


國立臺灣大學
National Taiwan University

Motivation: Mobile Browser Fingerprinting

Phones and computers: Complex configuration of hardware and software

Hardware
Characteristics



Software
Characteristics

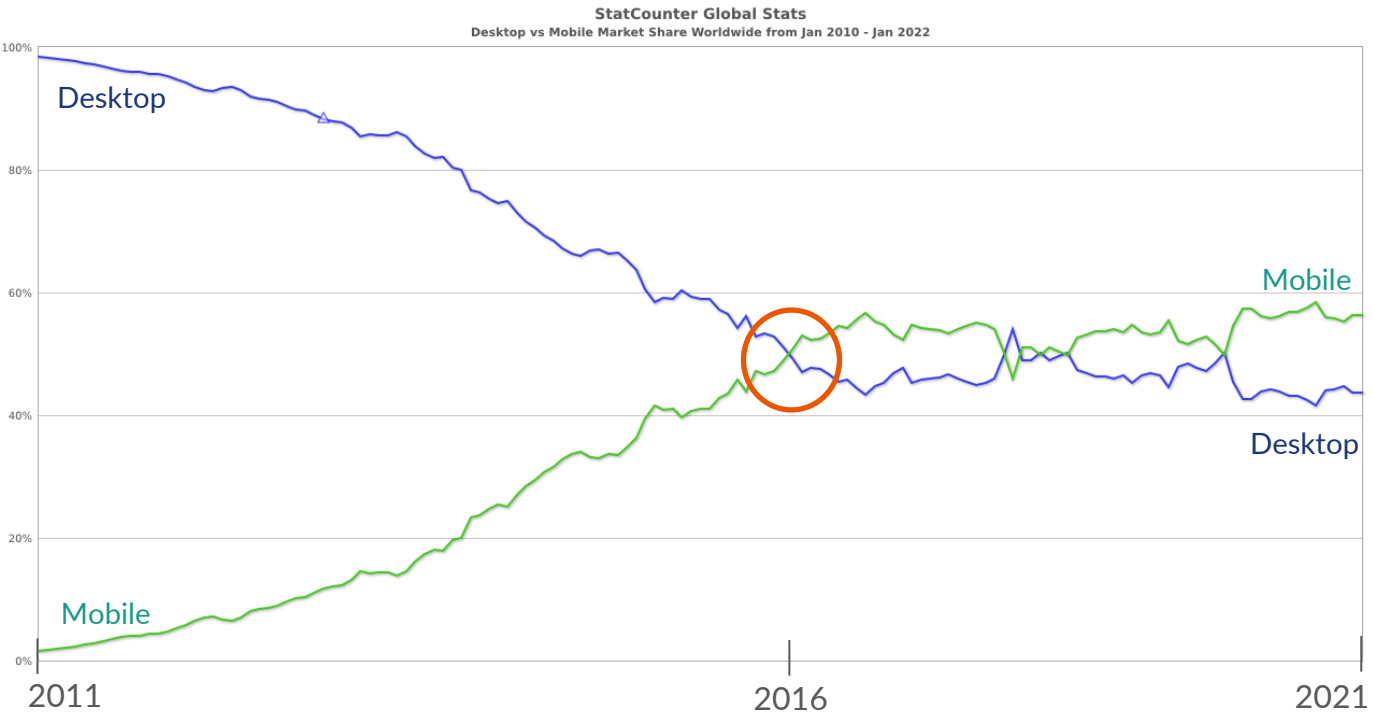


ABCDEFGHIJKLM
NOPQRSTUVWXYZ
abcdefghijklm
nopqrstuvwxyz
1234567890

Browser fingerprinting:

Tracker gathers enough characteristics to *uniquely identify* a browser across site visits

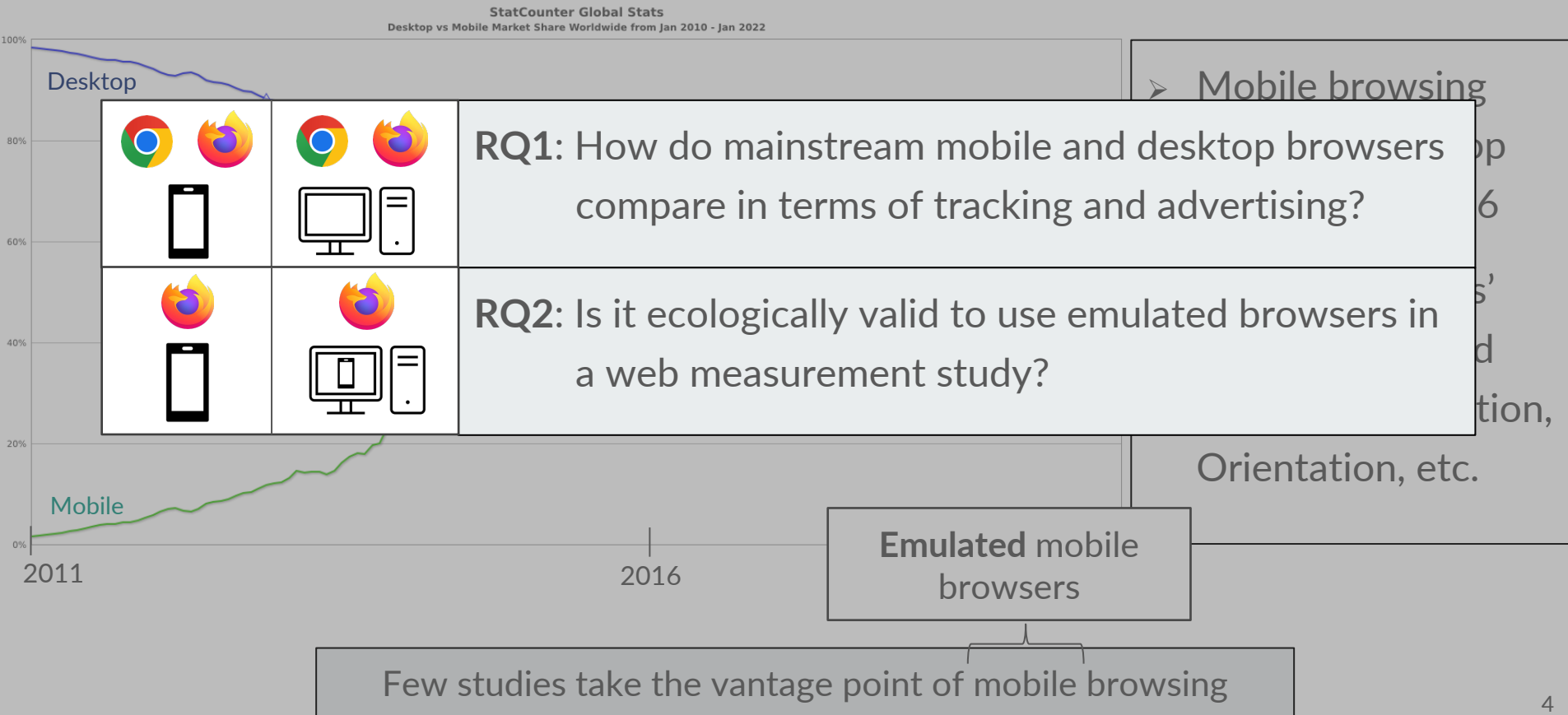
Motivation: Rise of Mobile Browsing



- Mobile browsing **overtook** desktop browsing in 2016
- Mobile browsers' **sensor APIs** used for tracking: Motion, Orientation, etc.

Few studies take the vantage point of mobile browsing

Motivation: Rise of Mobile Browsing



Motivation: Rise of Privacy-focused Browsers

- Many mobile browsers claim to enhance privacy



Firefox Focus:

- *“Block known [...] ad, analytics, and social trackers”*



DuckDuckGo:

- *“stop advertisers from tracking you on the sites you visit”*



Ghostery:

- *“block trackers”*
- *“anonymizes your data to further protect your privacy”*

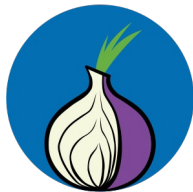
Brave



Firefox Focus



Tor Browser



DuckDuckGo



Ghostery



Motivation: Rise of Privacy-focused Browsers

- Many mobile browsers claim to enhance privacy



Firefox Focus:

	RQ3: How effective are privacy-focused browsers at blocking tracking and advertising?
	RQ4: What are the strengths and weaknesses of individual privacy-focused browsers?

Brave



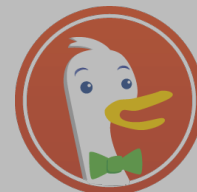
Firefox Focus



Tor Browser



DuckDuckGo

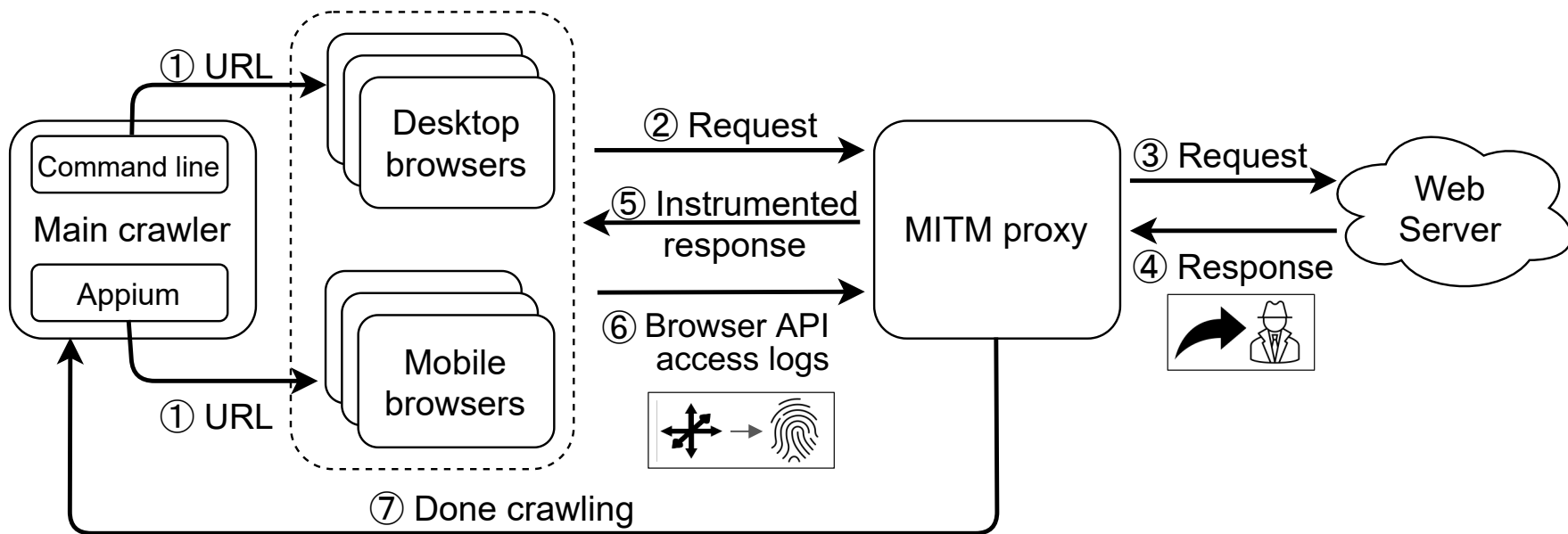


Ghostery



OmniCrawl Infrastructure Design

Goal: Synchronized crawling across multiple desktop and mobile browsers



NA: 1x Windows 10, 1x Ubuntu 18.04

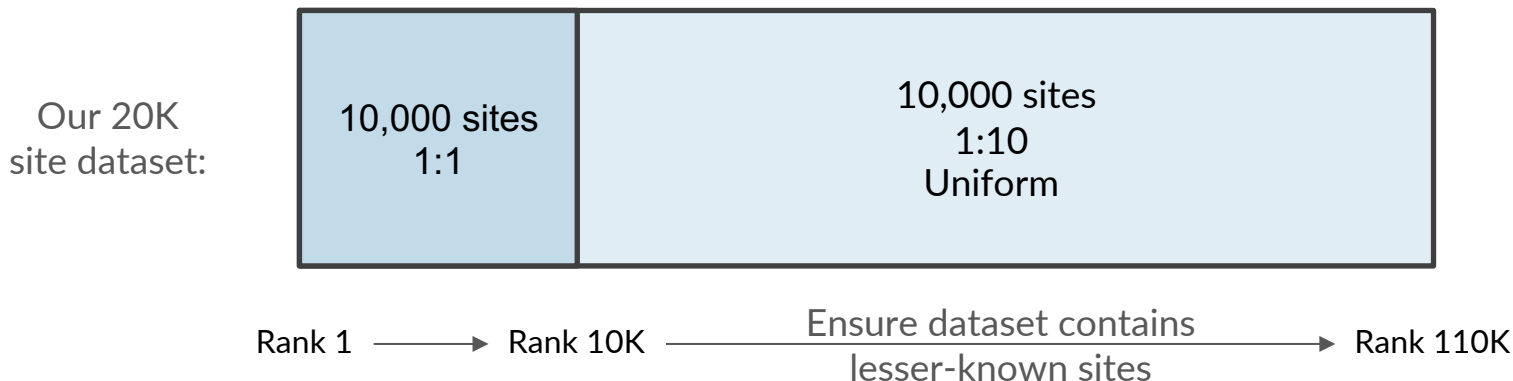


NA: 9x Motorola G5 Plus (Android 8.1)

Dataset Collection Methodology

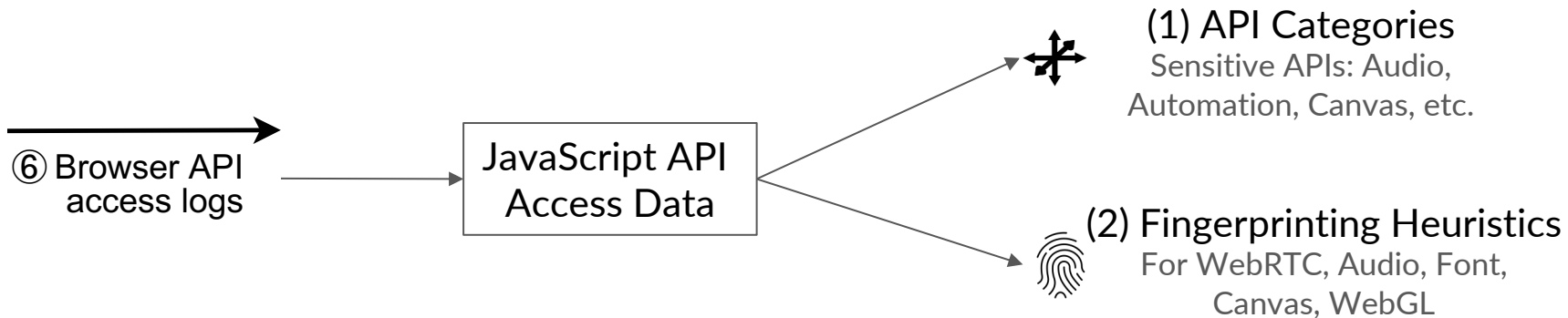
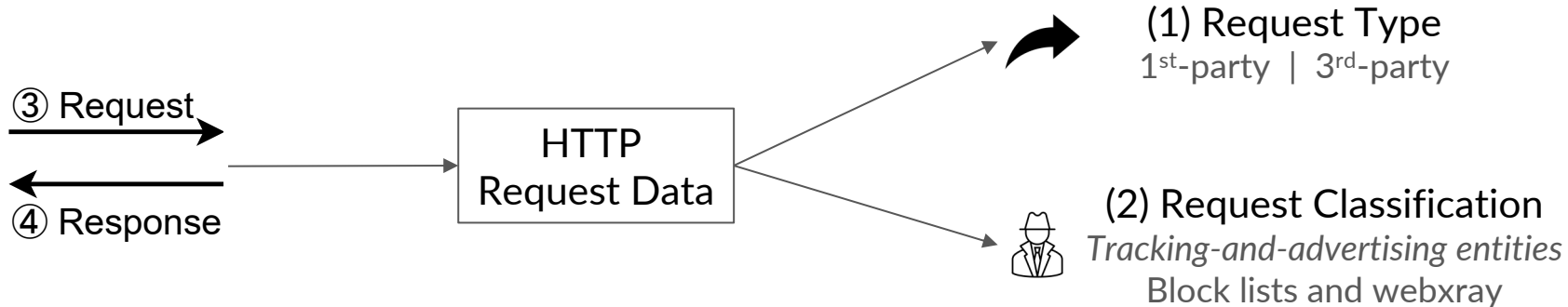
Goal: Crawl popular (high-ranked) *and* lower-ranked websites

Tranco Ranking: Stability and manipulation resistance¹






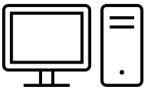
Dataset Analysis Methodology

Goal: Measure third-party tracking-and-advertising and potential fingerprinting



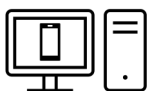
Statistical testing methodology to establish significance

Summary of Key Results – RQ1

 	 	<p>RQ1: How do mainstream mobile and desktop browsers compare in terms of tracking and advertising?</p>
--	--	---

- Mobile-only & desktop-only 3rd-party TA entities → **0.45%** of requests
 - Ecosystems of third-party tracking-and-advertising entities **more homogenous** than previously thought
- Mobile → more accesses to APIs indicative of fingerprinting
 - +54% Screen API
 - +21% WebRTC FP-heuristics

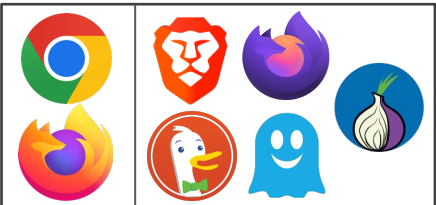
Summary of Key Results – RQ2



RQ2: Is it ecologically valid to use emulated browsers in a web measurement study?






- OpenWPM-Mobile → significantly **more** requests and accesses
 - +6% 3rd-party tracking-and-advertising requests
 - +50% Plugin API accesses
- Emulated browsers **may not be suitable** for measuring third-party requests or browser APIs accesses



Summary of Key Results – RQ3

	<p>RQ3: How effective are privacy-focused browsers at blocking tracking and advertising?</p>
---	--

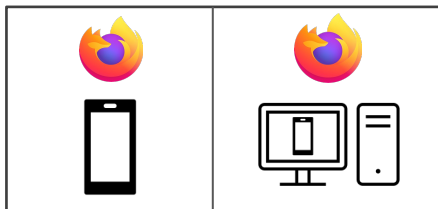
- **Do not** uniformly reduce third-party tracking-and-advertising requests
- Most effective for **highly prevalent** entities:
 1. Google (-60%)
 2. Facebook (-66%)
 3. Adobe Systems (-56%)

Summary of Key Results – RQ4

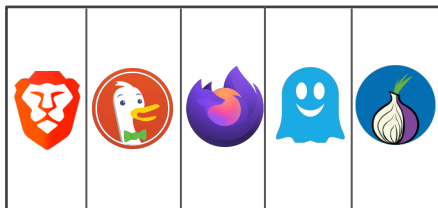
					RQ4: What are the strengths and weaknesses of individual privacy-focused browsers?
---	---	---	---	---	--

- *Generally* effective at reducing tracking-and-advertising... but
 - **Vary widely** in effectiveness
 - e.g. 3TA requests:  13x 
 - API accesses:  -25% group
 - **Deviate** from marketing claims
 - e.g. Tor “tracker blocking” unsupported

Recommendations for Studies and Vendors



- Future web measurement study design:
 - Avoid **emulated** browsers for mobile-specific measurements
 - Modify off-the-shelf browser **drivers** (esp. Selenium)



- Privacy-focused browser vendors:
 - Clarify concrete **implications** of marketing claims
 - Offer more **user control** over blocking comprehensiveness

More details in the paper & artifact available: <https://github.com/omnicrawl/>



OmniCrawl: Comprehensive Measurement of Web Tracking with Real Desktop and Mobile Browsers

Darion Cassel¹, Su-Chin Lin², Alessio Buraggina³, William Wang⁴, Andrew Zhang⁵, Lujo Bauer¹,
Hsu-Chun Hsiao², Limin Jia¹, and Timothy Libert⁶

¹ Carnegie Mellon University

⁴ University of Chicago

² National Taiwan University

⁵ University of Illinois Urbana-Champaign

³ University of Miami

⁶ Google



Carnegie
Mellon
University



國立臺灣大學
National Taiwan University