# Online Advertising and Tracking SaTC 2019 PI meeting breakout group report Co-leads: Athina Markoupolu, Zhiyun Qian, Zubair Shafiq

### 1. Problem/Domain Summary

Online advertising relies on a sophisticated third-party tracking ecosystem to track users for behavioral targeting. Online tracking uses non-transparent and intrusive mechanisms such cookies and browser fingerprinting that do not provide any meaningful control to users. Online advertising and tracking poses serious concerns for individuals as well as our society at large.

From the perspective of individual users, online behavioral advertising is:

- privacy invasive (e.g. re-targeted ads are eerily creepy).
- prone to exploitation by hackers (e.g. malvertising is a common occurrence on popular websites and puts millions of users at risk).
- negatively impacts web performance (e.g. inclusion of many trackers slows page loads).

On a societal level, online behavioral advertising has serious implications for:

- Surveillance and censorship (e.g. nation states piggyback on tracking identifiers used for online advertising).
- Fairness (e.g. advertisers misuse behavioral targeting capabilities provided by online advertising for discrimination on the basis of gender, race, or national origin).
- Data economics (e.g. data brokers unethically profit from selling people's information without their knowledge and consent).

The state-of-the-art in online behavioral advertising can be broadly into the following three categories.

- **Transparency:** Researchers have conducted large-scale measurement studies of online advertising and tracking on the web,<sup>1</sup> mobile,<sup>2</sup> and more recently IoT.<sup>3</sup> These studies have uncovered a variety of tracking techniques used by online behavioral advertising.
- Control: Researchers have developed tools for anti-tracking and ad-blocking.<sup>4</sup> Many of these tools and techniques are being implemented in browsers such as Safari, Firefox, Chrome, Brave, Opera, Cliqz, and Edge.
- Policy: Researchers have studied the impact of recent and upcoming privacy regulations such as GDPR in Europe and CCPA in California.<sup>5</sup>

<sup>&</sup>lt;sup>1</sup> Online Tracking: A 1-million-site Measurement and Analysis, ACM CCS 2016

<sup>&</sup>lt;sup>2</sup> Apps, Trackers, Privacy, and Regulators A Global Study of the Mobile Tracking Ecosystem, NDSS 2018

<sup>&</sup>lt;sup>3</sup> Watching You Watch: The Tracking Ecosystem of Over-the-Top TV Streaming Devices, ACM CCS 2019

<sup>&</sup>lt;sup>4</sup> AdGraph: A Graph-Based Approach to Ad and Tracker Blocking, IEEE S&P 2020

NoMoAds: Effective and Efficient Cross-App Mobile Ad-Blocking, PETS 2018

<sup>&</sup>lt;sup>5</sup> We Value Your Privacy... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy, NDSS 2019

## 2. Key Research Challenges

There are significant technical challenges that need to be addressed to advance transparency, control, and policy issues in online behavioral advertising.

Transparency: There is a lack of suitable measurement tools to handle the scale and diversity.

- There are millions of websites and apps that need to be instrumented, measured, and analyzed longitudinally to keep up with the ever evolving advertising and tracking ecosystems.
- There is also a heterogeneous mix of platforms on the web, mobile, and IoT. On one hand, there are at least three different engines (Gecko, Blink, Webkit) that are used in popular web browsers. The instrumentation design and implementation does not translate well across different browser engines. On the other hand, IoT ecosystem is very fragmented with tens of different platforms, most of which are not readily amenable to instrumentation.

Control: Anti-tracking and ad-blocking research has to deal with the inevitable arms race.

- It is challenging to automate detection of tracking and advertising, which currently relies on informal crowdsourcing. Researchers are looking into AI/ML based techniques to automated detection of tracking and advertising.
- It is also challenging to ensure that detection techniques are not susceptible to simple evasion. Existing anti-tracking and ad-blocking techniques use simple patterns in URLs and DOM elements that can be trivially manipulated to evade detection. AI/ML based techniques are also susceptible to adversarial inputs.

Policy: New tools are needed to inform policymakers.

- Measurement research on online advertising and tracking is needed to inform new privacy legislation initiatives across the globe.
- New measurement tools and techniques are needed to audit compliance of existing regulations.

# 3. Potential Approaches

Recent advances in theory and systems present new opportunities to address these challenges.

- **Systems:** New tools and techniques are being developed for information flow tracking. Recent advances in web browser and operating system instrumentation allow the needed visibility for efficient online and offline (post-hoc) detection of privacy leaks.
- **Theory:** Recent advances in differential privacy and private learning provide opportunities to re-architecture online behavioral advertising ecosystem in a privacy-preserving manner.
- It is challenging to bridge this gap between theory and systems.

Research should also explore alternate monetization models (i.e. beyond advertising). For example, blockchain-based micropayment based systems can be investigated to fund online content and services. There could also be interesting applications of game theory to figure out how much should users pay for online content and services or users should be paid for their data and attention.

It is important that research efforts to address these issues in online behavioral advertising leverage a multi-disciplinary perspective that not only includes computing but also economics, law, and journalism.

For broader impacts, it is also important to increase public awareness of the privacy issues associated with online behavioral advertising. The research should also inform regulators (e.g. FTC, FCC) and privacy advocacy organizations (e.g. EFF, ACLU, Future of Privacy).

# 4. Long-Term (> 10 years) Significance

The online behavioral advertising ecosystem is responsible for hundreds of billions in advertising revenues worldwide every year. The high financial stakes mean that the tussle is expected to sustain or even escalate in the next 10 years because of the following reasons.

The arms race aspect between ad/tracker blocking and counter-blocking is expected to continue in the short term. While more robust anti-tracking and ad-blocking methods are being developed, it is becoming increasingly challenging to implement these approaches on web browsers and mobile/IoT operating systems using standardized extensibility features such as extension APIs. While the arms race is expected to evolve on several fronts, it is important to continue this "cat and mouse" research to keep moving the equilibrium in the right direction.

The need for online advertising will not go away because it is important for marketers to legitimately advertise new products and services to users. Existing research has investigated the gap between effectiveness of behavioral and contextual targeting. Future research on privacy-preserving advertising is needed to fill the marketing need effectively while preserving privacy.

Online tracking is expected to remain a problem even if the issues with online behavioral advertising are addressed. This is because the tracking data is very valuable and can be exploited for various nefarious reasons such as surveillance and discrimination. For example, there are serious concerns about how tracking information can be exploited to abuse at-risk populations such as kids, elderly, and intimate partners.

#### 5. Other Important Aspects of This Topic (specify)

Meaningful efforts to address these issues need to adopt a cross-disciplinary perspective. The long-term solutions would need to **bridge the technology-policy gap**.