Online tracking: Threat Detection, Measurement and Response

Project Goal: A measurement-based solution to detect, quantify, classify and mitigate web privacy and security threats. The project has 4 components, some near completion and some in early development

1. Platform for web privacy & security studies
(a) Stable platform -- Nearly feature complete
→ *Realistic user simulation and login spoofing*
→ *Thorough instrumentation of HTTP data, browser storage, page content, and Javascript*
→ *Dynamic analysis of tracking scripts*
(b) Strong community involvement with *9 published studies by 5 research groups*
→ *450 stars, 80 forks and 13 contributors on Github*

2. A web privacy census
(a) Monthly, 1-million-site measurements since January 2016
→ *Results published and data released publicly*
→ *Additional data shared with journalists*
(b) Public data access platform & analysis library
→ *Under development, early version in testing*

3. Automated generation of "threat profiles"
(a) Resource load attribution implemented
→ *Provides an understanding of which third parties responsible for resource loading*
(b) Preliminary work to classify capability of third-parties to access private user information

4. Machine learning based tracker detection
(a) Promising early results which detect tracking scripts with a cross-validation accuracy > 99%
→ *Measurement data used as ground truth*
(b) Preliminary work to get fingerprinting script lists into browsers and privacy tools

Publications. Englehardt, Steven, and Arvind Narayanan. "Online tracking: A 1-million-site measurement and analysis." *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016.
Su, Jessica, Ansh Shukla, Sharad Goel, and Arvind Narayanan. "De-anonymizing Web Browsing Data with Social Networks." *Proceedings of the 2017 World Wide Web Conference*, 2017.