Overview

- Mainstream browsers are implementing countermeasures against third-party cookie-based crosssite tracking.
- Trackers are expected to migrate to browser fingerprinting, which does not require cookies, to track users.
- Existing countermeasures against fingerprinting limit website functionality, cause website breakage, and are not scalable.
- We propose FP-Inspector, a syntactic-semantic machine learning approach that detects browser fingerprinting.
- FP-Inspector detects 26% more scripts than the state-of-the-practice with an accuracy of 99%.

Fingerprinting the Fingerprinters: Learning to Detect Browser Fingerprinting Behaviors

Umar Iqbal
University of Iowa

Steven Englehardt
Mozilla Corporation

Zubair Shafiq
University of California-Davis

Fingerprinting Behavior

- A syntactic-semantic machine learning approach that detects browser fingerprinting.
- Syntactic-semantic representation to model script's behavior.
- Semantic representation is created through script's execution.
- Machine learning to learn fingerprinting patterns.
- Combination of APIs commonly used for fingerprinting.
- Limited interaction with the webpage.
- Communication with external entities.

FP-Inspector

- A syntactic-semantic machine learning approach that detects browser fingerprinting.
- Syntactic-semantic representation to model script's behavior.
- Semantic representation is created through script's execution.
- Machine learning to learn fingerprinting patterns.
- Combination of APIs commonly used for fingerprinting.
- Limited interaction with the webpage.
- Communication with external entities.

Evaluation

- Syntactic representation: Source as a signal of functionality.
- Semantic representation: Frequency as signal of uniqueness.
- Evaluation: Decision tree-based ensemble.
- Results: 99.9% accuracy.
- 36% more scripts than heuristics.
- Overkill of popular websites.
- 319 / Million: 16%.
- 2,349 / 100K: 19%.
- Websites that use ads for monetization.
- Popular trackers & Ad fraud services.

Syntactic vs. Semantic

- Syntactic representation: AST.
- Semantic representation: Model training.
- Syntactic representation: Detection + Measurements.
- Semantic representation: Model training.
- Evaluation: Decision tree-based ensemble.

Key Takeaways

- FP-Inspector improves the state-of-the-art in browser fingerprinting detection by incorporating syntactic-semantic representation.
- Fingerprinting adoption is on the rise with more than one-third of top 1K popular websites using fingerprinting.
- We open source our implementation and detections so that the community can benefit from them.
- FP-Inspector’s detections are incorporated by popular filter lists, such as EasyPrivacy, Disconnect, and DuckDuckGo.
- More analysis like fingerprinting APIs discovery in the paper.

Contact details

Umar Iqbal
umariqbal.com

Steven Englehardt
steven.englehardt@gmail.com

Zubair Shafiq
zshafiq@cs.ucdavis.edu

Paper
Source & Date
Contact details