Harpo: Learning to Subvert Online Behavioral Advertising

Jiang Zhang[1], Konstantinos Psounis[1], Muhammad Haroon[2], Zubair Shafiq[2]
[1] University of Southern California [2]University of California, Davis

Introduction

Privacy-invasive tracking techniques for user profiling and subsequent ad targeting

A real privacy threat

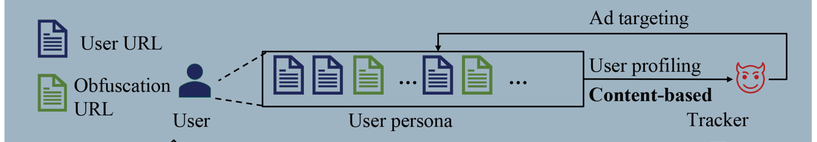
Existing privacy-enhancing solutions

Privacy-by-design
- Users opt in/out of tracking
- Mainly for iOS

Blocking
- Defensive
- Can be circumvented
- Kill advertising ecosystem

Obfuscation
- Offensive
- More ecosystem-friendly
- **But not principled/stealthy**

Threat Model



User:
- Routinely browse the web while misleading the tracker via obfuscator

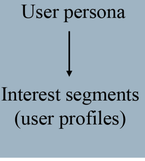
The obfuscator should be:
- Seamless
- Principled
- Stealthy
- Low overhead

Tracker:
- A third-party to provide advertising & tracking services

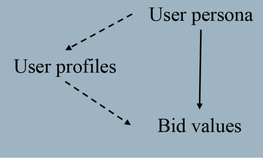
Main assumptions for tracker:
- Complete coverage of a user's browsing history
- Substantial computational resources to train ML models for tracking

Two real-world tracker models:

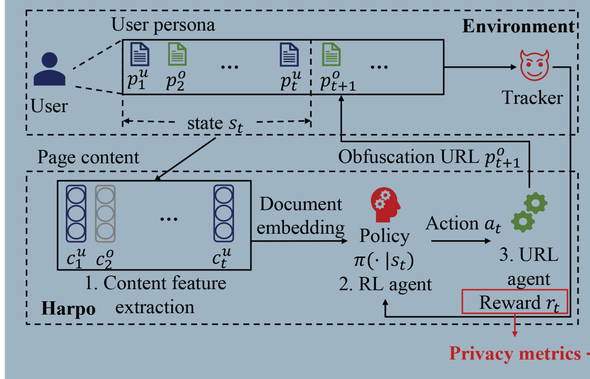
- **User profiling model:**



- **Ad targeting model:**



System Design



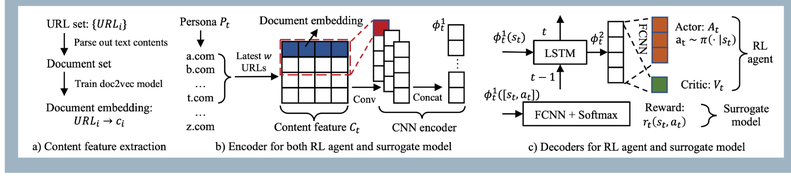
L_1: Percentile of false interest segments after obfuscation

L_2: Difference between segments before and after obfuscation

L_3: Percentile of increased high bids caused by obfuscation

L_4: Average ratio of bid values after obfuscation over before

Content Feature Extraction	Doc2vec embedding
RL Agent	Modeled via Conv + LSTM Trained using A2C (Advantage Actor and Critic)
Surrogate Model	Replicate real-world tracker models **Why?** Virtual environment for efficiently training RL
URL Agent	Harpo browser extension



Summary

Contributions:
- Propose HARPO, a principled RL-based approach to obfuscate a user's browsing history.
- Develop surrogate ML models to train HARPO's RL agent with limited or no black-box access to real-world tracker models.
- Demonstrate the success of HARPO against real-world user profiling and ad targeting models in terms of privacy, overhead, and stealthiness.

Harpo's Artifacts: https://github.com/bitzj2015/Harpo-Artifacts
What's next? A utility-preserving obfuscation approach (https://arxiv.org/abs/2210.08136)!

Evaluation

Privacy:

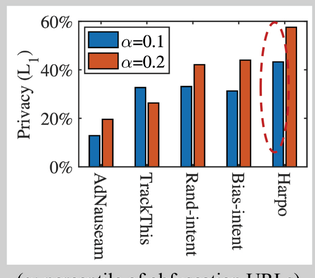
Approaches	Percentile of false segments	# of different segments	Increase percentile of high bids	Ratio of bid values
	Metrics			
	L_1	L_2	L_3	L_4
Control	0.00%	0.00	0.00%	1.00
AdNauseam	12.85%	1.53	2.70%	1.21
TrackThis	32.67%	2.81	-1.50%	0.89
Rand-intent	33.10%	3.18	8.40%	1.69
Bias-intent	31.27%	3.19	10.30%	2.07
Harpo	**43.24%**	**5.22**	**43.30%**	**6.28**

Against user profiling
Up to 3x

Against ad targeting
Up to 16x

Overhead:

Same/better privacy with 2x less obfuscation overhead

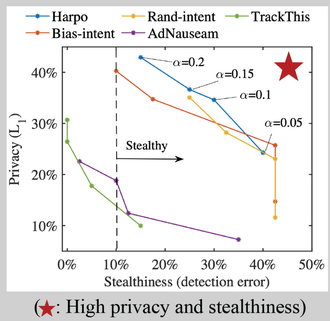


(α: percentile of obfuscation URLs)

Minimal impact on overall user experience
- Increased page load time: 0.2 sec
- Increased CPU usage: 5.3%
- Increased memory usage: 3.9%

Stealthiness:

Better privacy and stealthiness tradeoff



(★: High privacy and stealthiness)

Note that tracker will run fraud detection to detect the usage of Harpo.

ProperData Frontier 2nd Annual Symposium
Dec 1-2, 2022 | Irvine, California

SaTC Frontiers: Collaborative: *1956393* [1], *1955227* [2], *2103439* [3], *1956435* [4]
Institutions: UC Irvine [1], Northeastern Univ. [2], UC Davis [3], USC [4]
Lead PIs: A. Markopoulou [1], D. Choffnes [2], Z. Shafiq [3], K. Psounis [4]