NC STATE UNIVERSITY

CSC 495.002 – Lecture 5
Web/Social Networks Privacy: Targeted Advertising

Dr. Özgür Kafalı

North Carolina State University
Department of Computer Science

Fall 2017

NC STATE UNIVERSITY

Violations and Regret

- Violation: Reality does not meet user expectation about privacy
- How to detect and predict violations
- Regret: Later become unhappy about negative consequences of sharing behavior
- Common regret scenarios
- How to prevent regrettable actions

Problem Definition

- FTC defines “online behavioral advertising” (OBA) as:
- “The practice of tracking an individual’s online activities in order to deliver advertising tailored to the individual’s interests”

- Is it only online activities? Location tracking (which physical stores you have visited)
- Is it only individual? Aggregation of interests, trends

FTC: The United States Federal Trade Commission

OBA Terminology

- Advertiser: A party with an online ad willing to embed the ad in websites (with payment)

- Publisher: A party with a website willing to place ads from advertisers

- Ad-network: A party that collects ads from advertisers and places them on publisher websites (also takes care of payments)

Cookies

- Collect information about your browsing activity
- Content you click on and other actions you take online
- Small files stored on your computer when you visit a website
- What can be inferred from cookies?
  - Age group (e.g., 18–25)
  - Gender (e.g., female)
  - Purchase interests (e.g., shoes)
- Privacy implications

Useful Cookies

- Remember your preferences and settings (e.g., opting in or out of marketing emails)
- Remember whether you filled in a survey (not asked to do it again)
- Remember whether you've been to the site before (first-time user content might differ from a regular user)
- Show "related articles" according to your interests in a news site
- Remember a location you've entered (e.g., for weather forecasts)

K-anonymity and Differential Privacy

- If nothing revealed, then no OBA (but no potential gain either)
- “Sharing” lecture: How much control do you have on what you share?
- “Inference” lecture: What can you infer from the presented information?

Other Types of Advertising Models

- Contextual advertising: Based on the content of the page only
- Demographic targeting: Based on race, age, etc

Collaborative Filtering

- User-based CF: If customers X and Y have a similar transaction history, then recommend items X has bought to Y
- Item-based CF: If item B is often bought by buyers of item A, then recommend B to a new buyer of A
- Recommendations shown to users based on either/both

http://starecat.com/customer-who-bought-this-item-also-bought-shopping-suggestions-at-groceries-drawing/

Recommender Systems

- For user u, find k other similar users, $u_1, \ldots, u_k$
- For each item purchased by one of these k users, count how many times it was purchased and rank them accordingly
- Recommend items to u based on the ranking

- Potential attack: Influence the recommender system using public outputs of recommender system

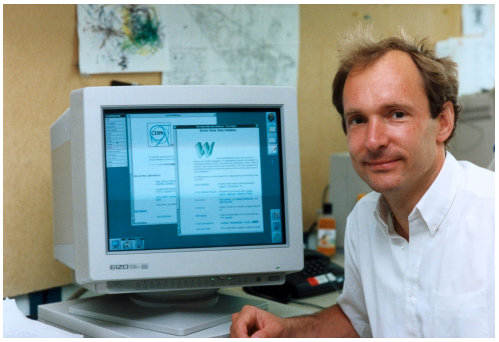
Calandrino et al. "You Might Also Like": Privacy Risks of Collaborative Filtering. IEEE Symposium on Security and Privacy, 2011

Private Browsing

"He's browsing in privacy mode."

http://communicrossings.com/safer-internet-browsing-manage-storage-private-data

World Wide Web



- English scientist Sir Timothy John Berners-Lee
- 1989, while employed at CERN in Switzerland
- To communicate with other research institutions

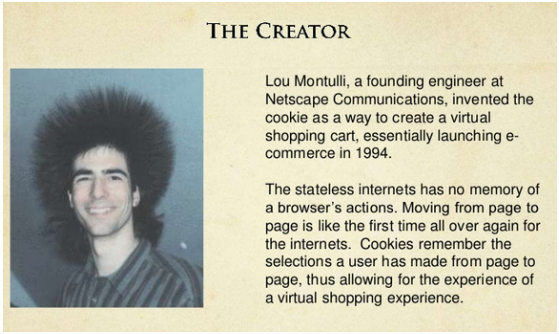
https://en.wikipedia.org/wiki/World_Wide_Web

Lou Montulli

https://en.wikipedia.org/wiki/Lou_Montulli

Cookies

- Why did he invent cookies?
- At the time, there was no way to store information about the state of the page
- Working on an e-commerce solution
- Implement shopping carts



https://www.slideshare.net/moxxycat/cookies-10097074

Cookies for Stateful HTTP

- Reliable mechanism to remember stateful information
  - Give user a better experience for repeated visits
  - Virtual shopping carts
- Record user's browsing activity
  - Past logins
  - Pages visited
  - Information entered into forms such as names or addresses

https://en.wikipedia.org/wiki/HTTP_cookie

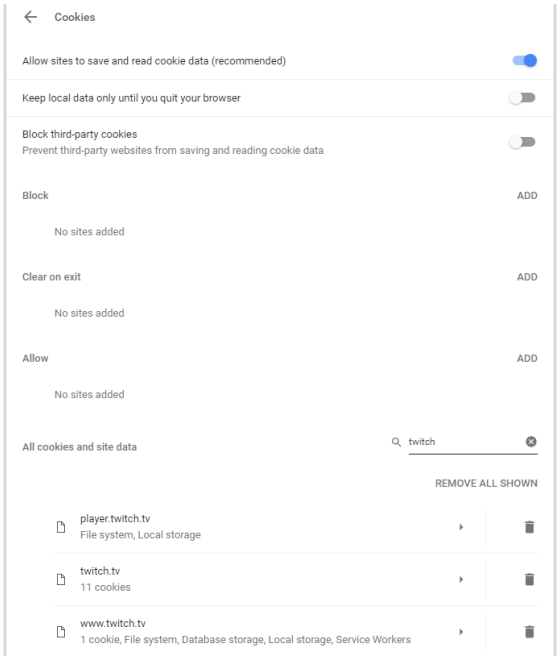
Types of Cookies

- Browser cookies
- Session cookies
- First-party cookies
- Third-party cookies

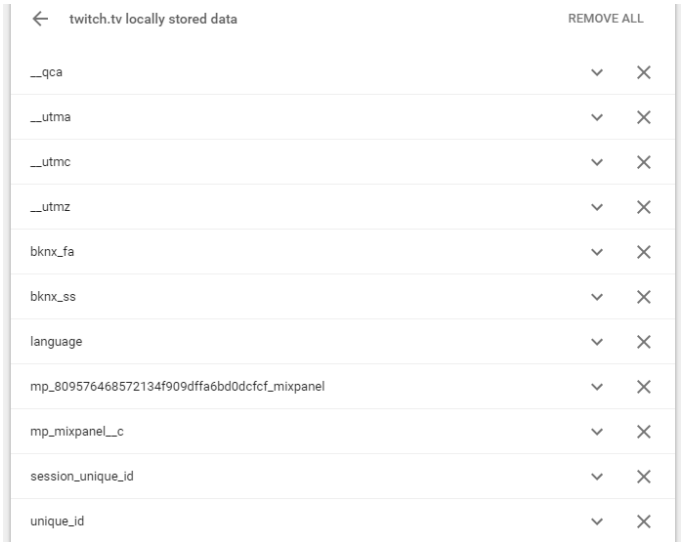
Browser Cookies

- Also known as HTTP cookie, Web cookie, or Internet cookie
- Small piece of data sent from a website
- Stored on the user's computer by the user's web browser
- <Name, Value> pair

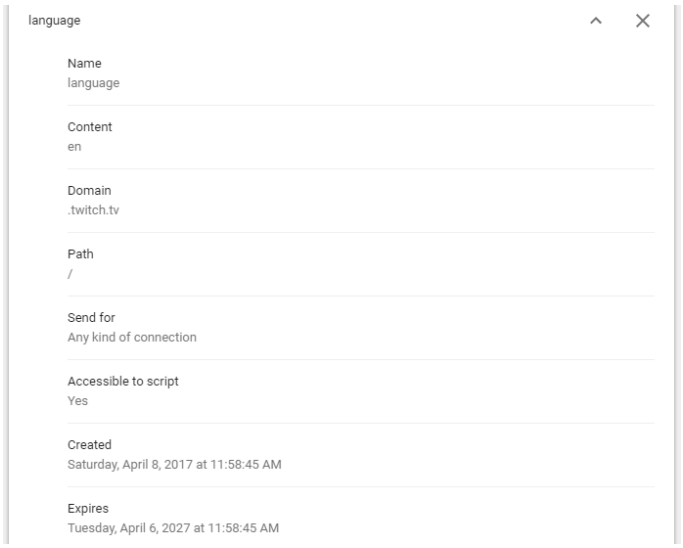
Cookies in Chrome



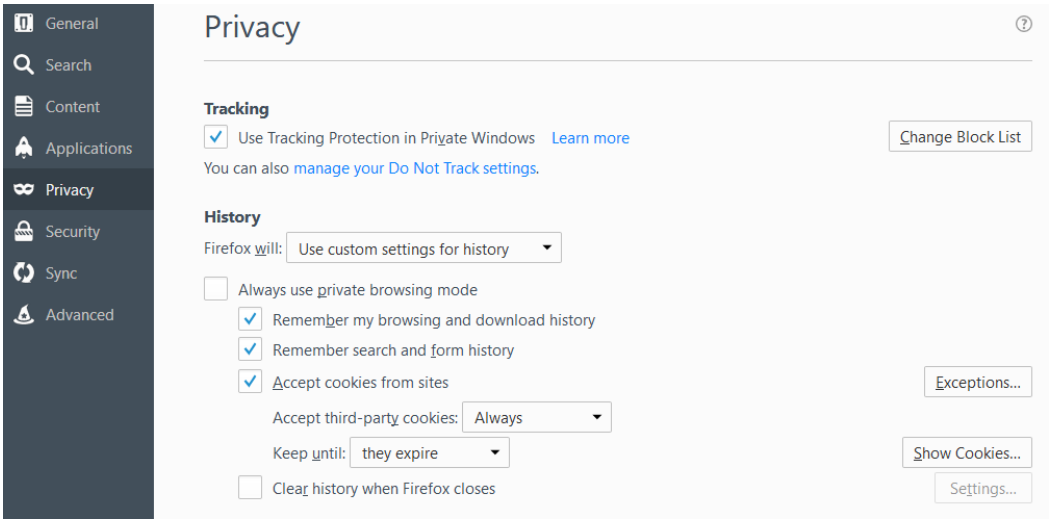
Cookies in Chrome: Twitch



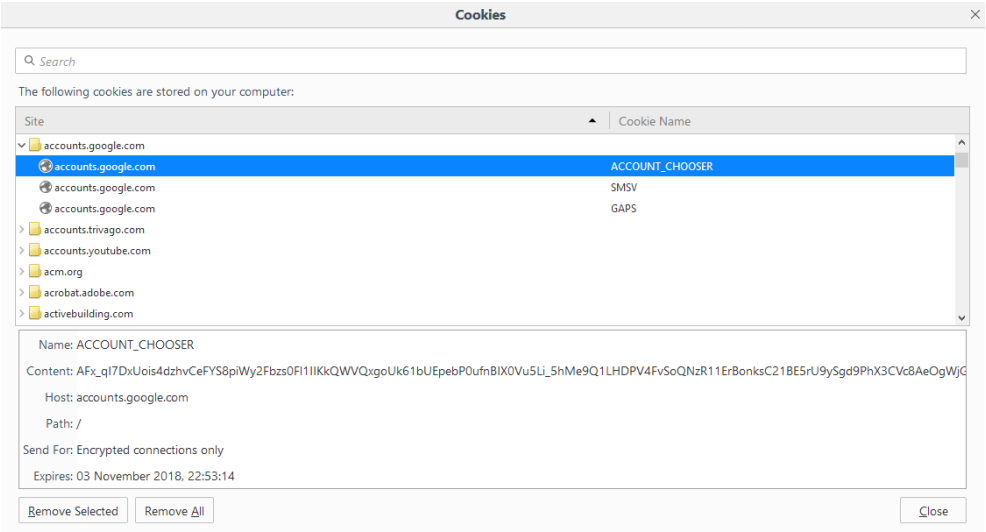
Cookie Attributes



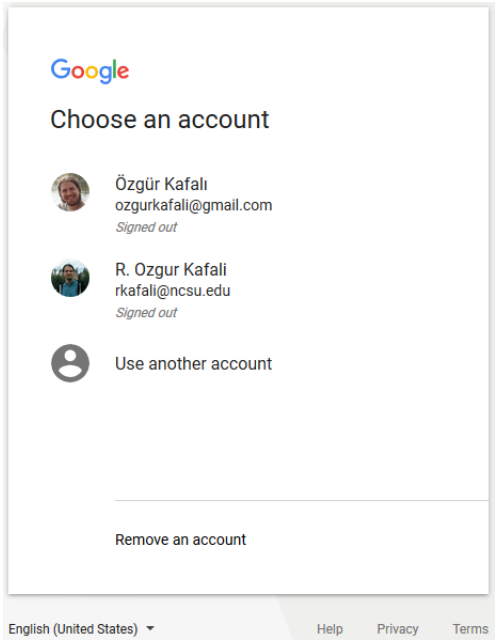
Cookies in Firefox



Cookies in Firefox: Google



Example Cookie: Google Account Chooser

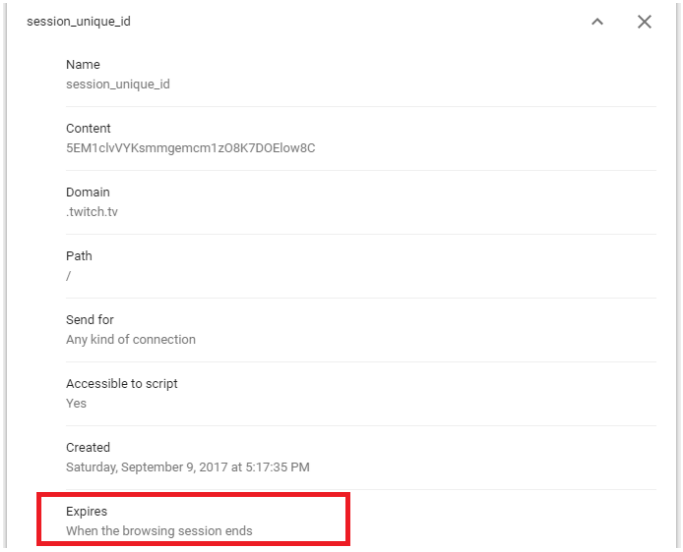


Session Cookies

- Also known as an in-memory cookie or transient cookie
- Exists only while the user navigates the website
- Erased when the user closes the browser
- Do not have a specific expiration date

- Similar to HTTP Session object in Java
- Session cookies do not collect information about the user
- Session identification information that does not personally identify the user

Example Session Cookie



Potential Attacks

- Cross-site request forgery

- Alice logs in to her bank's website (session cookie stored)
- Alice visits Bob's (malicious) site and clicks on an innocent appearing link
- Bob uses Alice's cookie to make a request from Alice's bank
- Request goes through because Alice's session cookie is sent along

First-Party Cookies

- The “domain” attribute of cookie matches domain in the URL
- For example, you visit cnn.com
- And, the “domain” of the stored cookie is “cnn.com”

Third-Party Cookies

- The “domain” attribute of cookie is different from the domain in the URL
- Typically appear when web pages feature content from external websites
- For example, you visit cnn.com
- Cookie from “amazon-adsystem.com” is stored

Cookies for Tracking

- How do advertisers use third-party cookies to track users and show ads?

- Alice visits “www.store.com”
- “www.store.com” is inside “ad.wetrack.com” Ad-network
- “ad.wetrack.com” sets a cookie on “www.store.com”
- Alice visits “www.news.com” (also inside “ad.wetrack.com” Ad-network)
- “ad.wetrack.com” also sets a cookie on “www.news.com”
- Also, “ad.wetrack.com” uses the content of the cookie on “www.store.com” to show Alice adds on “www.news.com”

Exercise: Check Your Cookies

- Check the cookies on your favorite browser
- Search for your favorite website
- Count how many cookies there are?
  - How many first-party cookies?
  - How many third-party cookies?
- Anything unusual?

Survey Results

- About 30% of users clear their 1st party cookies over a period of one month
- On average, 2.5 cookies per computer for Yahoo
- 10% of users disable third-party cookies

https://www.comscore.com/Insights/Presentations-and-Whitepapers/2007/Cookie-Deletion-Whitepaper
http://www.smorgasbork.com/2009/04/29/a-study-of-internet-users-cookie-and-javascript-settings/

Studies

- Look at two studies
  - One mitigation approach against targeted advertising
  - One usability study of tools to limit targeted advertising

Adnostic: Privacy Preserving Targeted Advertising

Adnostic: Privacy Preserving Targeted Advertising*

Vincent Toubiana vincent.toubiana@nyu.edu
Arvind Narayanan relax@stanford.edu
Dan Boneh dabo@cs.stanford.edu
Helen Nissenbaum hfn1@nyu.edu
Solon Barocas solon@nyu.edu

Abstract

Online behavioral advertising (OBA) refers to the practice of tracking users across web sites in order to infer user interests and preferences. These interests and preferences are then used for selecting ads to present to the user. There is great concern that behavioral advertising in its present form infringes on user privacy. The resulting public debate — which includes consumer advocacy organizations, professional associations, and government agencies — is premised on the notion that OBA and privacy are inherently in conflict.

In this paper we propose a practical architecture that enables targeting without compromising user privacy. Behavioral profiling and targeting in our system takes place in the user's browser. We discuss the effectiveness of the system as well as potential social engineering and web-based attacks on the architecture. One complication is billing; ad-networks must bill the correct advertiser without knowing which ad was displayed to the user. We propose an efficient cryptographic billing system that directly solves the problem. We implemented the core targeting system as a Firefox extension and report on its effectiveness.

Toubiana et al. Adnostic: Privacy Preserving Targeted Advertising. Network and Distributed System Security Symposium, 2010

Privacy Preserving Targeted Advertising

- Goal: Support targeted advertising without compromising user privacy (not replace, but complement)
- Idea: Implement OBA as a browser extension
  - Use browser's history
  - Results reside inside browser
  - User information is not leaked to the outside world (only clicked ads are communicated)

What is Tracked?

- Clickstream (all URLs user visited)
- Behavioral profile
  - Intent to purchase (e.g., request quotes, add item to shopping cart)
  - Influence over purchasing habits of others (e.g., time spent on latest news and current trends according to interests)
- Ad impression history (all ads displayed to the user)
- Ad click history (all ads user clicked)

Incentives

- Privacy-conscious publishers
- Low barrier to entry
- Regulatory compliance
- Potentially improved user tracking
- Targeting in private browsing mode
- User control via centralized interface
- Standardized audience segmentation

Implementation Steps

- User profiling: Extract interest categories from visited websites
- Ad network associated with a page sends a list of ads considered appropriate for the page
- Browser decides what to display based on interests
- Similarity measures help match ads (identified with tags) to interests (identified with tags)

Folksonomy

- A corpus of tags
- Users apply public tags to online items
- No hierarchical structure as in a taxonomy or ontology
- Also known as collaborative tagging or social tagging
- “Delicious” website: https://del.icio.us/
- “Steam” game store website: Users tag games

Ads Preference Categories

A Google Ads Preferences categories

Top-level categories

Animals
Arts and Humanities
Automotive
Beauty and Personal Care
Business
Computers and Electronics
Entertainment
Finance and Insurance
Food and Drink
Games
Home and Garden
Industries
Internet
Lifestyles
Local
News and Current Events
Photo and Video
Real Estate
Recreation
Reference
Science
Shopping
Social Networks and Online Communities
Society
Sports
Telecommunications
Travel

Subcategories of *Entertainment*

Entertainment → Celebrities
Entertainment → Clubs and Nightlife
Entertainment → Comics and Animation
Entertainment → Comics and Animation → Anime and Manga
Entertainment → Comics and Animation → Cartoons
Entertainment → Comics and Animation → Comics
Entertainment → Dancing
Entertainment → Entertainment Industry
Entertainment → Fashion and Modeling
Entertainment → Fun and Trivia
Entertainment → Humor and Bizarre
Entertainment → Humor and Bizarre → Bizarre
Entertainment → Humor and Bizarre → Humor
Entertainment → Humor and Bizarre → Paranormal
Entertainment → Movies
Entertainment → Movies → Bollywood and Lollywood
Entertainment → Movies → Horror Films
Entertainment → Movies → Movie Memorabilia
Entertainment → Movies → Movie Rentals and Sales
Entertainment → Movies → Science Fiction and Fantasy Films
Entertainment → Multimedia Content
Entertainment → Multimedia Content → Flash Content
Entertainment → Multimedia Content → Podcasting
Entertainment → Multimedia Content → Video Clips and Movie Downloads

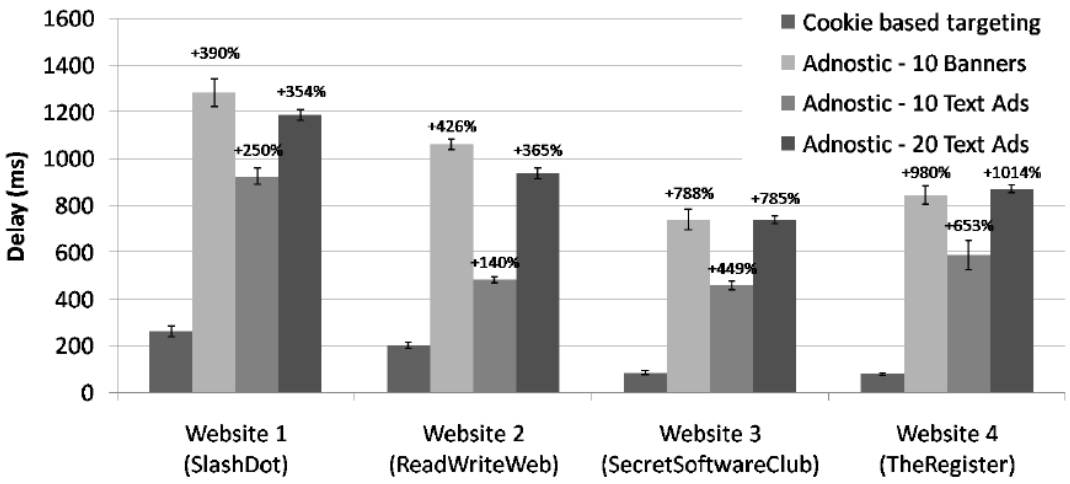
Limitations

- Network latency and bandwidth
- Effectiveness
- Enforcement of non-tracking
- Ad blocking

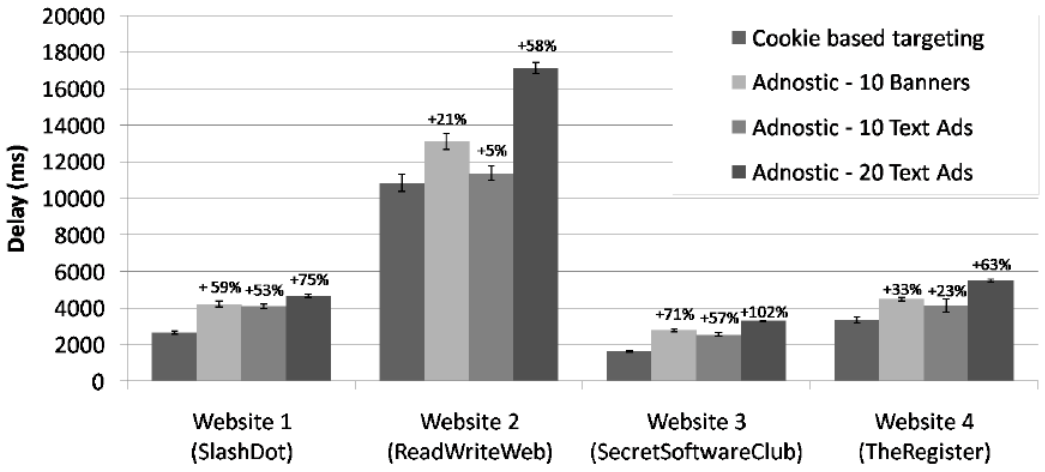
Evaluation

- Consider 4 publishing websites
  - SlashDot: Lightweight site with on average 3 banners
  - ReadWriteWeb: Heavy site with on average 13 banners and external content
  - SecretSoftwareClub: Very lightweight site with text ads
  - TheRegister: Text ads and banners

Average Ad Rendering Time



Average Page Loading Time



Similar Tools

- Privad: https://addons.mozilla.org/en-US/firefox/addon/privad-client/
- TrackMeNot: https://cs.nyu.edu/trackmenot/

Why Johnny Can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising

Why Johnny Can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising

Pedro G. Leon, Blase Ur, Rebecca Balebako, Lorrie Faith Cranor, Richard Shay, and Yang Wang
Carnegie Mellon University, Pittsburgh, PA
{pedrogln,bur,balebako,lorrie,rshay,yangwan1}@cmu.edu

ABSTRACT

We present results of a 45-participant laboratory study investigating the usability of nine tools to limit online behavioral advertising (OBA). We interviewed participants about OBA and recorded their behavior and attitudes as they configured and used a privacy tool, such as a browser plugin that blocks requests to specific URLs, a tool that sets browser cookies indicating a user's preference to opt out of OBA, or the privacy settings built into a web browser. We found serious usability flaws in all tools we tested. Participants found many tools difficult to configure, and tools' default settings were often minimally protective. Ineffective communication, confusing interfaces, and a lack of feedback led many participants to conclude that a tool was blocking OBA when they had not properly configured it to do so. Without being familiar with many advertising companies and tracking technologies, it was difficult for participants to use the tools effectively.

Consumers may control OBA using a number of tools. However, to use these tools successfully, users must be able to install a tool, configure it to match their preferences, and effectively use it. While these tools have the potential to satisfy the concerns of consumers and regulators, there has been little rigorous evaluation of their usability and effectiveness.

In this paper, we present results of an in-depth study investigating the usability of tools that limit OBA. We found serious usability flaws in all nine tools we examined. The online opt-out tools were challenging for users to understand and configure. Users were confused by technical jargon and complicated settings in some tools. Users also struggled to install and configure Tracking Protection Lists (TPLs) and other blacklists to make effective use of blocking tools. They often erroneously concluded the tool was blocking OBA when they had not properly configured it to do so.

Leon et al. Why Johnny Can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising. Conference on Human Factors in Computing Systems, pages 589–598, 2012

Study Overview

- Setting: Laboratory study with 45 participants
- Objective: Test usability of 9 tools to limit OBA
- Tool types:
  - Tools for setting cookies to opt out of OBA (e.g., http://optout.aboutads.info/)
  - Privacy settings of browsers (e.g., Chrome, Firefox)
  - Browser plugins for blocking specific URLs (e.g., Adblock)

Methodology

- Semi-structured interviews with participants to gather
  - Perceptions about OBA
  - Knowledge about OBA
  - Attitude towards OBA
- Configure and use a privacy tool
- Record behavior and attitudes (audio recording and screen capture)
- Work as though they were using their own computer

Wall Street Journal Educational Video on OBA

Wall Street Journal: http://www.wsj.com/video/how-advertisers-use-internet-cookies-to-track-you/92E525EB-9E4A-4399-817D-8C4E6EF68F93.html

Findings

- Serious usability flaws
- Difficult to configure
- Default settings not privacy protective
- Confusing interfaces

Adblocking Tools

Tool	Capabilities	Strengths	Weaknesses
Blocking			
AdBlock Plus	Blocks tracking, blocks ads	Facilitates awareness of trackers when users click icon. Users are guided to pick a filtering list.	Configuration interface confusing, includes jargon. Difficult for participants to find specific trackers to unblock. Difficult for participants to understand differences between filtering lists.
Ghostery	Blocks tracking	Facilitates awareness of trackers through on-screen alerts. Alerts helped resolve broken website elements. Easy installation.	Configuration interface includes jargon. Participants unaware that default settings don't block trackers. Multiple steps required to enable blocking.
IE-TPL	Blocks tracking, enables DNT headers	Easy to install TPLs from provider websites.	Configuration interface confusing. Participants unaware that default settings don't block trackers. Participants did not realize they had to choose a TPL in order to be protected. Even when prompted, participants were unable to choose a TPL using the interface. Difficult to unblock specific trackers.
TACO	Blocks tracking, sets permanent opt-out cookies and blocks third-party cookies	Sets opt-out cookies by default and prevents deletion. Facilitates awareness of trackers from icons and alerts. Suggests workarounds for broken website elements. Provides diverse privacy features.	Large number of privacy features overwhelmed participants. Configuration interface confusing, includes jargon. Initial configuration took a long time. Difficult for participants to find specific trackers to unblock. Participants unaware that default settings don't block trackers. Participants didn't notice workaround suggestions.
Opt-out			
DAA	Sets opt-out cookies for 79 advertising companies	Provides links to more information about each tracker. Easy to select specific trackers.	Initial configuration took a long time. Difficult to navigate to actual opt-out page. Not obvious that opting out of all trackers requires switching out of default tab on opt-out page. Participants incorrectly believed that they were opting out of tracking. Participants did not realize that deleting cookies nullifies opt-outs. Opt-outs sometimes fail. Participants unable to confirm opting out was effective.
Evidon	Sets opt-out cookies for 184 advertising companies and provide links to opt out of 118 additional companies	Provides links to more information about each tracker. Easy to select specific trackers. Provides links to non-standard opt-outs. Provides the most comprehensive list of tracker and advertising opt-outs.	Initial configuration took a long time. Participants incorrectly believed that they were opting out of tracking. Difficult to navigate to actual opt-out page. Participants did not realize that deleting cookies nullifies opt-outs. Difficult for users to complete non-standard opt-outs. Opt-outs sometimes fail. Participants confused by "opt-out request sent" messages with no additional information. Participants unable to confirm opting out was effective.
PrivacyMark	Sets opt-out cookies for 160 advertising companies	One-click opt-out.	Participants did not realize that deleting cookies nullifies opt-outs. Participants unable to confirm opting out was effective. Requires dragging icon to bookmarks toolbar, which participants could not find. Tutorial video states incorrectly that tool will stop tracking. Participants thought clicking icon would delete cookies.
Built-in			
IE-Settings	Blocks specified cookie types	Default settings provide some protection.	Configuration interface confusing, includes jargon. Participants couldn't figure out how to block all third-party cookies.
Firefox	Blocks specified cookie types, DNT	Participants could easily block third-party cookies and enable DNT headers.	Participants didn't know what protection DNT provided.

Design Implications

- Usability issues with blocking content
- Need privacy protection, but don't mess up websites
- No feedback whether the tool is working properly

Facebook Ads

- News article: https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race
- Links are also on the course website

Things to Look For

- Root cause: What went wrong?
- If it was not intentional, what was the original aim?
- Affected parties
- Implications and similar problems
- Mitigation (using methods we have seen): Prevention, detection, recovery

- Take 10 minutes to look at the incident on your own

- Now discuss with your neighbor
- Also take a look at the summary report: https://drive.google.com/file/d/0B3m-I0YVAv0EbEdrS2hiSF9JUWc/view

Verizon and Google Cookies

- Verizon news article: https://www.eff.org/deeplinks/2014/11/verizon-x-uidh
- Google news article: https://www.wired.com/2012/02/google-safari-browser-cookie/
- Links are also on the course website

Things to Look For

- What are the similarities and differences between the two incidents?
- Mitigation (using methods we have seen): Prevention, detection, recovery
- Take 10 minutes to look at the incidents on your own

- Now discuss with your neighbor
- Also take a look at the summary reports
  - Verizon: https://drive.google.com/file/d/0B3m-I0YVAv0EMzFmZFlXaFpZUm8/view
  - Google: https://drive.google.com/file/d/0B3m-I0YVAv0EVVRrVGxxSlVCSUE/view