Advertising, Analytics and Tracking

Thierry Sans

Disclaimer - Advertisement inside

This presentation mentions several *Google* products

➡ Google and Facebook accounts for 65% of worldwide online ad revenue

January 04, 2017 | source Fortune

Design an ad and give it to me

I have a cool car to sell and
I want people to know about it

Advertiser

I have a cool website, many viewers, and I want to make money out of it

Content Publisher

Put this ad on your webpage!

Content Publisher

Advertiser

The ~~victim~~ target

Hey, that's a cool car!

CSCC09 Winter 2017 Announcements & News Lectures Labs Course Work Resources Piazza Github Feedback

UNIVERSITY OF TORONTO SCARBOROUGH

CSCC09 Programming on the Web

An introduction to software development on the web. Concepts underlying the development of programs that operate on the web. Operational concepts of the internet and the web, static and dynamic client content, dynamically served content, n-tiered architectures, web development processes and security on the web.

Instructor: Thierry Sans

Course Objectives

Web technologies change at a rapid rate. Between the time that I wrote this and the time that you are reading it for the first time, it is likely that some hot new technology will have sprung up. As a software developer/computer scientist, how will you assess the latest technology and decide for yourself or your company whether it will fit in with the libraries and frameworks you are already using or whether it is a passing fad?

Our goal in this course is to learn about the fundamental underlying concepts of the protocols, languages, and architectures that make up web programming. These underlying general concepts will allow us to compare different approaches to building web applications and will give us the tools to evaluate new technologies as they come along.

As an *Advertiser*

Two popular models

On other's websites - click banners

- Pay per click advertising (fix price)

DoubleClick by Google

Google AdSense

On search engine result page - sponsored links

- Buying keywords (bidding price)

Google AdWords

➡ See the *List of advertising networks* (Wikipedia)

As a Content *Publisher*

Ad Serving Services

Embed ads in your webpage/webapp

- The ad network rewards you with cash every time a visitor clicks on an ad on your webpage

Technically speaking

For the web programmer

➡ A javascript snippet (to be inserted in the webpage) that performs ajax requests to the ad networking company (ad is shown in an iframe)

For the visitor

➡ A third party cookie tracking his/her visits through different sites to display more relevant ads

Web Scraping and Click Fraud

Web Scraping

Idea

- A website that will extract, collect and aggregate data from other websites

➡ Spamming search engine (spamdexing)

Goal

- Attract visitors to your website and fool them to click on ads

Click Fraud

Having a bot (a computer program) that automatically clicks on

- ads displayed on your website
  (to increase your earnings)

- ads anywhere on the web but targeting specific ads
  (to increase the expenses of your competitors)

Detecting Click Fraud

➡ For advertising networks, there is a conflict of interest

- Lot of research work to detect click fraud
- Mature technology deployed by ad networks

Case

- ➡ Google Clique by Michael Anthony Bradley (2004)
- ◉ Not detected by Google at first

Click Fraud For Experts

"An Eastern European pack of cyber thieves known as the *Rove* group hijacked at least 4 million computers in over 100 countries to make off with $14 million in "illegitimate income" before they were caught."

"The suspects entered into deals with various internet advertisers in which they would be paid for generating traffic to certain websites or advertisements. But instead of earning the money legitimately, the FBI said the defendants used malware to force infected computers to unwillingly visit the target sites or advertisements"

By *RICHARD ESPOSITO* and *LEE FERRAN* | ABC News – Wed, Nov 9, 2011

Web Analytics

Measuring, Analyzing and Assessing

➡ You want to maximize your revenue from advertisement

- Which website guide the users to your website?
- What are the keywords that they typed in the search engine that guide them to your website?
- What do they do on your website?
- How long do they stay? What pages do they look at?
- Where are they from geographically?

✓ Web Analytics

Two Techniques

- *Log file* analysis (server side)

➡ Server side code analyzing the web server logs

- *Page tagging* analysis (client side)

➡ Javascript code analyzing the user interactions

Web Analytics

- Analytics in-house

✓ can mix *log analysis* and *page tagging*

- Analytics as a service

✓ *page tagging* only

➡ See the *List of web analytics software* (Wikipedia)

Web Tracking

Ad Cookie

- Cookie with unique ID to identify the same user visiting different websites

Let's look at Mozilla *Lightbeam*

http://www.mozilla.org/en-US/lightbeam/

Privacy mode

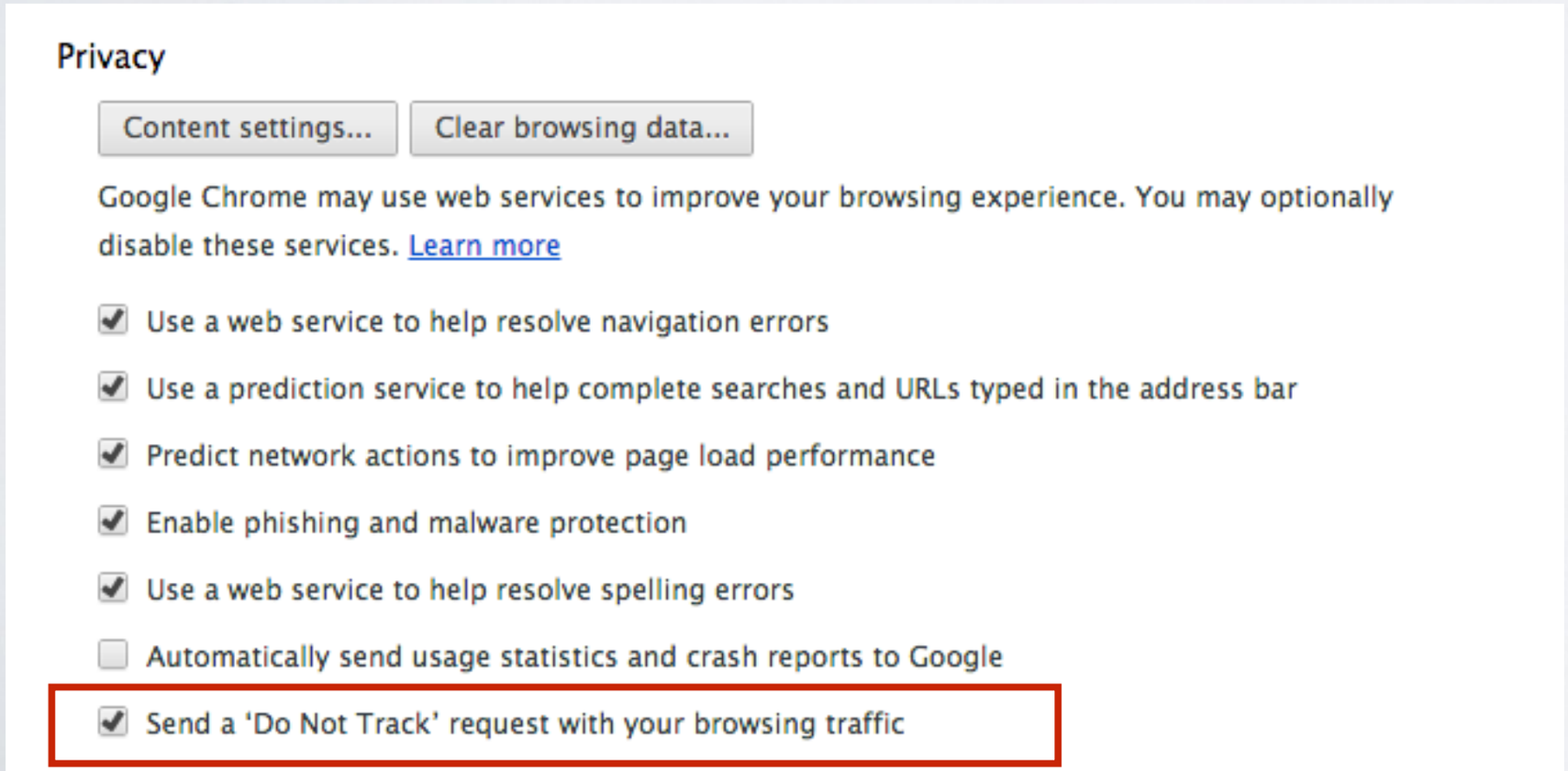
✓ Disable browser data storage

- (frontend) web cache
- HTTP cookies
- HTML5 local storage
- Flash/Silverlight cookies

◉ Does not protect against browser extensions

Do Not track

- HTTP header field (proposed in 2009)
- Website can decide whether or not to honor such a request



Next Generation Tracking

- Google Ad ID

  http://spectrum.ieee.org/tech-talk/telecom/internet/googles-adid-aims-to-replace-cookies-for-tracking-web-users