

CS 5436 / INFO 5303 Fall 2024

Homework 1

Due: **September 24**, 11:59p ET

This is an **INDIVIDUAL** assignment.

You may discuss, but each student must submit their own work.

White space after each problem indicates the approximate length of the expected answer.

REQUIRED (40 points)

Problem 1 (6 points)

The same-origin policy in Web browsers is supposed to prevent information flows between Web content from different origins.

(a) Why and how does it allow third-party trackers to learn about first-party URLs visited by the browser?

(b) How can the same-origin policy be strengthened to prevent tracking via third-party cookies? Mention any concrete examples that we discussed in class. (Note: Google's Privacy Sandbox does not involve a redesign of the same-origin policy).

(c) What forms of tracking would still work even if the browser completely blocks all information flows between Web origins?

Problem 2 (2 points)

In some cases, sharing personally identifiable information in a raw form can be illegal, but sharing hashed versions is allowed. Discuss the differences between raw and hashed identifiers, including the related privacy implications.

Problem 3 (6 points)

In this problem, you will investigate and evaluate several privacy-enhancing technologies.

(a) Describe what Chrome's Incognito browsing mode provides. What types of tracking techniques are blocked in this mode? What types of tracking still work?

(b) [Cover Your Tracks](#) is a service that allows you to measure the uniqueness of browser fingerprints. What parts of your browser fingerprint provide the most identifying information (please describe your configuration: OS, browser make and model, etc.)? What can you do to reduce the identifiability of your fingerprint?

(c) Recent research has shown that browser fingerprints change frequently from normal use. Suppose that, on average, the fingerprint of your browser changes every 5 days. To what extent do these changes protect you from tracking, and to what extent do they not?

Problem 4 (2 points)

Compare and contrast Do Not Track and Global Privacy Control. What are the differences in enforcement? Do you think GPC will be more successful than DNT and, if so, why?

Problem 5 (2 points)

Compare and contrast Mozilla's Total Cookie Protection and Google's Chrome privacy sandbox. Which provides stronger privacy protection? Why?

Problem 6 (6 points)

(a) How do the main tracking mechanisms differ between Web and mobile?

(b) Why is it easier to identify mobile users?

(c) What are other key differences between Web and mobile tracking?

Problem 7 (5 points)

We discussed two categories of privacy definitions: control and secrecy.

(a) Cite two privacy theorists in each of these categories.

(b) Select one from each. Summarize their definitions and explain why they belong in respective categories

(c) How would each of the two theories explain why certain online behavioral tracking violate privacy.

Problem 8 (3 points)

(a) According to contextual integrity (CI), a right to privacy is a right to _____?

(b) In terms of CI, defining privacy as a right to control information about oneself is right/wrong? Explain.

Problem 9 (6 points)

Contextual informational (privacy) norms/rules have five parameters.

(a) What does this mean?

(b) List the parameters, briefly describe each, and illustrate with an example.

(c) What does "acting in capacity" mean?

(d) From your home background (culture) state a well formed contextual informational rule/norm that you think might surprise some of your classmates. NOTE: These need not be legal rules.

(e - 2pt bonus) In 2009, Amazon deleted copies of George Orwell's 1984 from customers' Kindle devices. (Yes, they refunded the 99 cents that customers had paid.) There are many perspectives one could take on this action, but how might you use CI as a lens through which to see the privacy aspect?

Problem 10 (2 points)

In analyzing whether a system or a practice violates privacy, contextual integrity (CI) draws attention to factors that may be overlooked by other definitions. Recall the weather, family-tracking, and cheap-gas mobile apps that share users' driving information with data brokers, who then derive an individualized driving score for each user and sell these scores to auto insurance companies.

Use the CI framework to reveal all features of this case that are relevant to the question of whether these apps violate users' privacy. (Note: Your judgment on whether they do or do not is a separate question which we will ask in a later homework.)

PICK YOUR OWN – should add up to 20 points

For each problem you pick, do the entire problem (i.e., you cannot choose-and-mix subproblems)

Problem CS1 (5 points)

How do ad-blocking tools like Adblock Plus, Ghostery, and Disconnect mitigate propagation of personal information to advertisers via RTB?

Problem CS2 (15 points)

[OpenWPM](#) is an open-source Web privacy measurement framework. Use it to analyze 100 popular websites from diverse categories (explain your algorithm for selecting these sites).

(a - 3 points) What are the most common trackers on your chosen sites?

(b - 6 points) How many of the trackers include persistent identifiers? What do you consider a persistent identifier? Identify fingerprinting code (you can use this [link](#) for the list of probable fingerprinting commands – try `recursive_dump_page_source` method).

(c - 6 points) Count the flows of information across different contexts. To do this, assign each site to a particular social context (eg, news, entertainment, etc.) – it's okay to use an automated assignment algorithm. There exists an information flow between context A and context B if the same tracker appears in a page that belongs to context A and a page that belongs to context B.

You can upload a documented Jupyter notebook or report in PDF.

Useful links and helpful hints:

- [OpenWPM tutorial](#)
- [OpenWPM configuration settings available](#)
- You can create your own function in custom_command.py (there's already an example function there)
 - Might be useful: [Selenium with Python](#) for programmatically accessing webpage elements (this is what the LinkCountingCommand() there does)
- openwpm/command_sequence.py might also have some useful stuff
 - command_sequence.recursive_dump_page_source() will dump the page source to the sources directory
- crawl-data.sqlite contains some information retrieved from the web crawl
 - [This link](#) explains some of the tables it stores

Problem INFO1 (7 points)

Briefly summarize the three connections between privacy and autonomy discussed in PIC. Make sure you provide a definition of autonomy that serves as the basis for your answer

Problem INFO2 (2 points)

"Privacy is a right, not a preference!" is a slogan popularized by the Electronic Privacy Information Center. What's the big deal?

Problem INFO3 (1 point)

"All the parameters matter!" Explain.

Problem INFO4 (3 points)

(a) What is meant by "privacy harms"?

(b) List and briefly describe two, making sure you explain why they are privacy harms.

Problem INFO5 (1 point)

Tom Gerety was concerned that privacy “expands like a gas to fill up available space.”
What was his (and others’) approach to addressing this concern?

Problem INFO6 (6 points)

“If you’ve done nothing wrong (got nothing to hide,) you don’t need privacy!”

(a) Expand on this claim, making the most plausible case in its favor. Use examples.

(c) How might you draw on the definition of Contextual Integrity to push back? Feel free to make up your own examples or draw any from lectures that are useful for your answer.