

CS 5436 / INFO 5303 Spring 2023

Homework 1

Due: **February 23, 11:59p ET**

This is an **INDIVIDUAL** assignment.

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

REQUIRED (40 points)

Problem 1 (3 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 2 (9 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 3 (3 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 4 (3 points)

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

Problem 5 (1 point)

What is the difference between a descriptive and normative conception of privacy? Explain in a few sentences.

Problem 6 (5 points)

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

- (a) Cite two privacy theorists in each of these of each categories.
- (b) Select one from each. Summarize their definitions and explain why they belong in respective categories
- (c) These two theories would offer different reasons why online behavioral tracking violate privacy. Briefly explain.

Problem 7 (5 points)

- (a) What is meant by “privacy harms”?
- (b) List two privacy harms and describe them in a few sentences
- (c) For each harm, describe an example, and explain why they constitute privacy harms (what do they have to do with privacy? why are they harmful?)

Problem 8 (3 points)

Tom Gerety was concerned that privacy “expands like a gas to fill up available space.”

- (a) What was his (and others’) approach to addressing this concern?
- (b) Based on reading PIC 6 discuss one reason why the basis for this approach runs into problems of its own. (Your answers must reference page numbers in the relevant PIC chapters.)

Problem 9 (3 points)

- (a) What is the privacy paradox and why is it referred to as such?
- (b) List two rebuttals to the paradox. (Your answers must reference page numbers in the relevant PIC chapters.)

Problem 10 (5 points)

According to Contextual Integrity, informational (privacy) norms/rules have five parameters.

- (a) What are they? (describe in a few sentences).
- (b) 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.
- (d - 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?

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

[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 Ok 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 (5 points)

Privacy skeptics claim that privacy, in fact, is not particularly important, or worth protecting as privacy advocates claim. Briefly (a few sentences each) describe three arguments that these skeptics offer in support of their claims. (Your answers must reference page numbers in the relevant PIC chapters.)

Problem INFO2 (10 points)

Explain connections (hint: three) between privacy and autonomy. Provide a definition of autonomy and justify your definitions and connections.

Problem INFO3 (10 points)

(a) PIC Chapter 6, introduces the problem of “privacy in public.” Describe it in a short paragraph and explain how it arises from prior definitions.

(b) Explain carefully why the problem of privacy in public is not a problem for the account of privacy as contextual integrity.

Problem INFO4 (8 points)

Someone says to you, “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.

(b) Rebut this argument drawing on any one of the philosophical accounts of privacy discussed in chapter 4 or 4.

(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.