CS 5436 / INFO 5303 Fall 2024

Homework 2

Due: October 22, 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 (4 points)
Location-tracking SDKs combine location information with information about the user (including demographic information, preferences, and behavioral data) and tag the resulting profile with an "advertising ID."

(a) For each of {location, user information, advertising ID}, explain where an SDK can get it from and what permissions it needs.

(b) If the same SDK is incorporated into multiple apps, each instance of this SDK generates a separate profile. How are these profiles linked into a single profile?

Problem 2 (6 points)
(a) TLS was designed as an end-to-end secure transport protocol. Why not use it for end-to-end secure messaging, i.e., why do we need special protocols?

(b) Signal provides end-to-end encryption for messages; Telegram does not by default. Why might the operator of the service prefer that messages are end-to-end encrypted?

(c) Imagine a new law that requires communications providers to "escrow" all message decryption keys by encrypting them under the law enforcement agencies' public keys. This way, law enforcement agencies can decrypt any message, if needed (e.g., upon presenting a warrant or court order). What

important security properties of end-to-end secure messaging protocols would be weakened or restricted by such a requirement?

(d) Which security properties of Signal fundamentally rely on ratcheting, ie, updating the keys for each message? Name and explain each property. Be specific about the attack each property considers and what it prevents the adversary from learning.

Problem 3 (1 point)

How can metadata (in particular, the IP address) be linked to someone's identity?

Problem 4 (4 points)
Install and play with the Tor browser:

https://www.torproject.org/download/

Read the self-written overview of Tor and the privacy properties it provides here:

https://www.torproject.org/about/overview.html.en

(a) What privacy properties does the Tor browser provide? Against what type of adversaries? Specify any important caveats in Tor's privacy guarantees.

(b) Access an onion (hidden) service. Describe what this service does (include a few screenshots). Why do you think they operate it as an onion service and not as a conventional Web service?

Problem 5 (6 points)

(a) A motorist watches a pedestrian crossing the road. Does this violate the pedestrian's privacy? Explain your answer.

(b) PIC Chapter 6, introduces the "problem of privacy in public."
(i) Based on PIC Chapter 6, briefly explain what it is and why the problem exists.

(ii) Explain why the pedestrian case in (a) isn't a counterexample.

(c) Explain carefully and briefly why the problem of privacy in public is not a problem for contextual integrity.

Problem 6 (3 points)
The privacy paradox serves as a powerful argument.
(a) For whom? Explain your answer

(b) When critics attempt to defuse it, they offer several rebuttals. List and explain two.

Problem 7 (1 points)
"Stick with the old or go with the flow!"
If CI did not include a way to evaluate new data practices brought about by digital technology and platforms and, where relevant, to compare these with pre-existing data flow norms, it would be considered conservative. In a few sentences, explain why.

Problem 8 (3 points)
Norm/Rule for job interviews: Interviewers are not allowed to ask candidates about their religious practices.

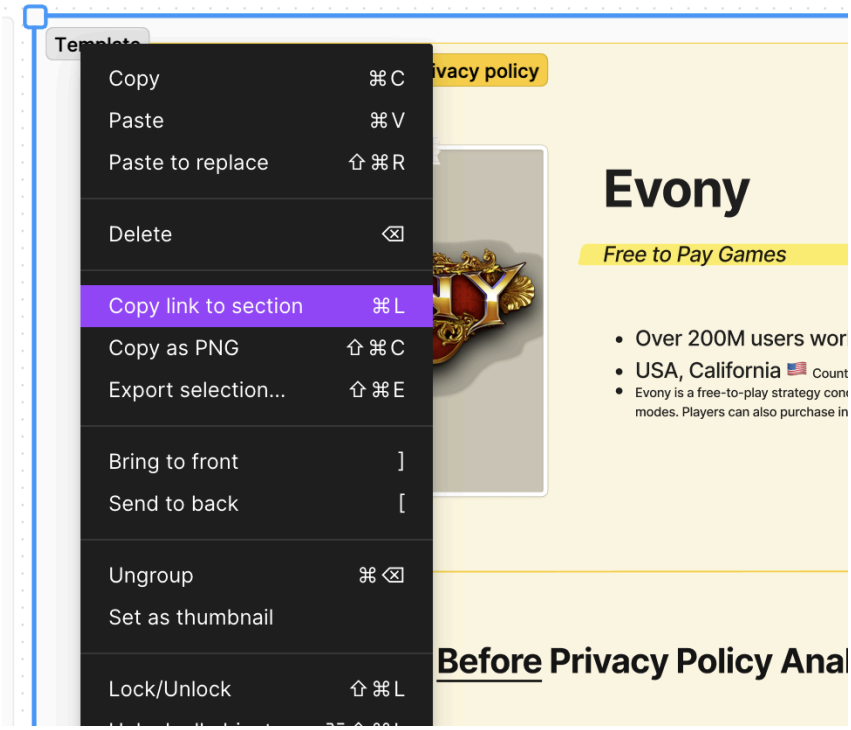
(a) Express this as a CI-Tuple

(b) How might you argue **either** in favor, **or** against this rule

(c) Why is this banner funny and inappropriate at the same time? (bonus +1)

Problem 9 (12 points)
Drawing from your breakout group work, please answer the following questions. NOTE: you may copy directly from your group's output, or you may augment these answers with language of your own. (Please indicate where you've done this, or where you've combined the two.). You may answer the sub-questions in bullet point form.

(a) Policy Overview. Please provide a direct link to your filled-in Figma template. List the company, and Privacy Policy URL, and briefly describe the nature of the website or app (i.e. services or information it offers.) What information is collected by the website or app?



(b) Collection Practices. How is the information collected? Does the policy say anything about the data it infers?

(c) Third Party Practices. Does the site mention conditions under which information is collected from third parties. (i.e. transmission principles)? What information is collected about site visitors from third parties? (List the information or none or doesn't say)

(d) Sharing Practices. Does the site disclose information to third parties? (Y/N) What types of information is shared with these (and unspecified) third parties? With whom does the site/app share the data it collects? Be complete and specific.

(e) Data Retention. Can users easily delete their data? How? How long is the information held; is it ever deleted/destroyed?

(f) Policy Scope. Can the privacy policy change? How are visitors notified of changes? Does the policy grant rights for specific people based on location or citizenship? E.g. GDPR or California Consumer Privacy Act

(g) Consent Process. How is consent obtained? Is it renewed? How can it be revoked?

(h) Transparency. Does the policy use visual aid to be more explainable? Does the policy offer the user help to understand what their specific data may be? E.g. forward to websites with profile information, or simple takeout?

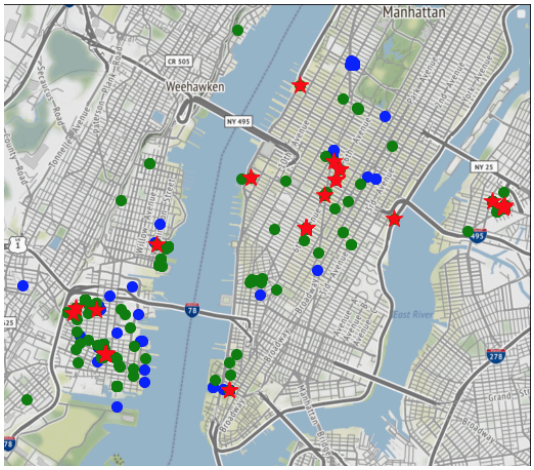
(i) Use Practices. How is the data used (2-3 instances is ok). Does the policy say whether your data is used for training AI? (Y/N)

(j) Trust. Did your trust in the company change after doing the privacy policy analysis? (Imagine that the service is of interest to you:) Would you change your behavior towards it in any way?
(k) Bonus: In your analysis, did you uncover *any* ***strange*** privacy practices? Please provide the exact wording of the policy here (one example). Why did you choose to highlight this practice, and how does it relate to the privacy theories and practices we explored in our class discussions?).

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 (10 points)
Imagine a dataset with anonymized location traces of millions of phones. In this problem, you will investigate how one might use such a dataset to infer the social relationship graph (parent-child, friendships, romantic relationships, etc.).

Use the NYC data in the Foursquare dataset (kaggle) to infer which users may know each other. Assume that two users know each other if they checked in at the same location within 1 hour of each other at least 10 times. Provide this list. Plot the top 5 user pairs on the location map and mark their possible interactions (use geopandas and contextily).



Problem CS2 (10 points)
SecureDrop is an open-source whistleblower platform used by many freedom-of-speech and news organizations.

1. To set up SecureDrop (branch release/2.10.0), simply use `make dev` that builds a Docker container. There are only two interfaces available: source and journalist. Use SecureDrop to send messages back and forth (note that messages from the source will be encrypted).

2. Deploy SecureDrop as two onion services, for the source and the journalist, respectively. Verify that these onion services are accessible outside of your machine over Tor.

3. If the journalist or source is not using Tor with the “Safest mode” enabled, prevent them from submitting forms. Display an alarm and show their visitorID and info about their hardware, browser, and timezone from this fingerprinting library to show how easy they are to track.

4. SecureDrop relies on a Secure Viewing Station that stores the private key used to encrypt the source’s messages. For the dev version, however, the startup script will use a known key (located here) that can decrypt messages. Your task is to alert both users (source and journalist) if they are using this known key (65A1B5FF195B56353CC63DFFCC40EF1228271441) for encryption. If the server uses this key, decrypt all of the source’s messages on the journalist panel (you can ignore files).

Write a report that describes:

(1) The measures that SecureDrop takes to protect privacy of the source (look up the full architecture of the platform, not just the dev Docker version).

(2) Why does the journalist need a Secure Viewing Station to decrypt messages?

(3) Configuration changes to deploy SecureDrop as an onion service.

(4) Screenshots of tasks 3 and 4.

Submit the diff, i.e. git diff origin/release/2.10.0 > changes.patch, with implemented 3 and 4.

Problem INFO1 (6 points)

The city of Metropolis is installing license plate readers at all intersections. This system will allow them to learn about their residents and raise funds by selling the data to local businesses and location data brokers. As a bonus, they’re convinced that residents would drive more safely. Supporters say, “Public is public!” but to address privacy concerns, they would post clear notices.

(a) Explain the logic of supporters’ of the system.

(b) Think of two ways you could would utilize CI to challenge the logic?

(c) Choose two disruptions in data flow that the system would enable. Drawing on the 3-layered evaluation of privacy as contextual integrity, choose a side. Either support the entrenched (or existing) flows against those enabled by the system, OR support of the novel (disruptive) flows that would be enabled by the the system. (*It's ok to introduce some assumptions in order to build your arguments. Be sure to reveal these assumptions.*

Problem INFO2 (4 points)

Some people argue that based on the private/public dichotomy, transit apps like MTA are entitled to share data with OpenAI, revealing, for example, your location in real time and historical travel patterns.

(a) How might they frame their arguments? (Note: there is more than one step to this argument.]

(b) Now use CI to generate a counterargument.

Problem INFO3 (8 points)

Based on these two statements drawn from privacy policies, answer a) and b) below:

Comcast: *"In certain situations, third party service providers may transmit, collect, and store this information on our behalf to provide features of our services."*

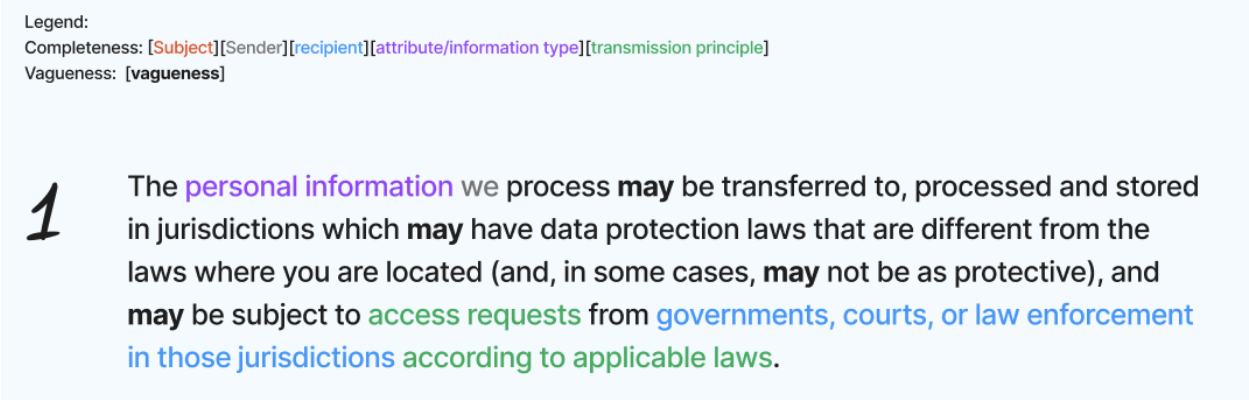
Costco: *"We do not otherwise sell, share, rent or disclose personal information collected from our pharmacy pages or maintained in pharmacist records unless you have authorized such disclosure, or such disclosure is permitted or required by law."*

(a) In what way is the statement drawn from the Comcast privacy policy vague?

(b) Drawing on Contextual Integrity (per Shvartzsnaider, et. al.), compare the completeness of the two statements.

(c) Based on Reidenberg's discussion of *ambiguity*, identify 5 ambiguous terms from the privacy policy you worked on. Explain their ambiguity, what misunderstandings might stakeholders have? (Hint: Use Table 2 and Appendix Table A3 from Reidenberg).

(d) Using the CI Analysis Method (Shvartzsnaider, et. al.) annotate 5 sentences from your group's privacy policy following the model below. You may reuse the same information flow you annotated in class and those colors.:



(e) In the language of CI, Privacy Policies describe a company's data practices and associated data flows. Identify and highlight two such descriptions from your privacy policy that fail to specify a value for one of the CI parameters, e.g. recipient, sender, and/or transmission principle. (can be from d, but explain)

Problem INFO4 (2 points)

NYC restaurants are in a dispute with Doordash. They say they're entitled to personal information (e.g. Name, address, email, phone number …) about customers who place orders from their restaurants through Doordash. Citing privacy as one consideration, Doordash says, No! Pick a side and explain how CI might support it. Hint: the parameters and 3-layered analysis. [Note: this is a real, ongoing dispute, though simplified. You can assume there is no final resolution.]