Homework 2

Due: March 21, 11:59p ET

This is an INDIVIDUAL assignment. You may discuss but each student must submit their own work.

REQUIRED (40 points)

Problem 1 (2 points)
Describe all methods for cross-device tracking mentioned in the Forbrukerrådet report. For each method, list all identifiers used for linkage.

Problem 2 (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 3 (2 points)
TraceTogether is Singapore’s proximity-tracing system, consisting of a mobile app and a physical Bluetooth token. Until April 2022, TraceTogether was mandatory in most public places in Singapore, including workplaces, schools, shops, restaurants, etc. Its primary purpose was to identify users who were in physical proximity to those with known Covid infections, but individual location information collected by TraceTogether is also available to the police and can be used for criminal investigations.

Analyze this situation using the contextual integrity framework.

Problem 4 (8 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) 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?

(c) 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.

(d) Why does Signal need two different ratcheting mechanisms?

Problem 5 (2 points)
Almost all countries authorize law enforcement officers to access data on suspects’ devices using a search warrant, subpoena, or court order. Modern devices, including laptops and mobile phones, encrypt data “at rest.” Describe technical and legal issues involved in accessing such encrypted data.

Problem 6 (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 7 (2 points)

(a) According to Cl, a right to privacy is a right to ____________?

(b) Explain in terms of Contextual Integrity the common definition of privacy as a right to control information about oneself.

Problem 8 (1 points)

“All the parameters matter!” Explain.

Problem 9 (8 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.

If you did not attend class the day of the privacy policy activity, email klp98@cornell.edu to get a privacy policy assigned to you.

(a) List the 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? How is the information collected? What information is collected from third parties? How precisely are these third parties described? (e.g. data brokers, Google, Facebook, Government Agencies. or simply “third parties.”)

(b) Does the site mention conditions under which (how) information is collected from users, collected about users from third parties and illustrate for each of the two how one can see them as transmission principles.

(c) Does the site specify the types of information it shares with third parties? Please list some. Does the site specify terms under which it is shared (i.e. transmission principles.) List one or two.

(d) Does the site collect location information? If yes, in what form? Does it share location data with third parties and if so, does it list stipulations. List a few.

(e) How long is the information held; is it ever deleted/destroyed? Can the privacy policy change? How are visitors notified of changes? How many words is your privacy policy? Does the site respect DNT signals?

Problem 10 (2 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 11 (4 points)

Norm/Rule for job interviews: Interviewers are not allowed to ask candidates about their religious practices.

(a) Express as a CI-Tuple

(b) How might you argue either in favor, or against this rule
(c) Why is this funny? (bonus +1)

Problem 12 (3 points)

(a) Joe Turow mentioned “hidden curriculum” a few times. What did he mean by the hidden curriculum of marketing?
(b) What is attribution? Why is it challenging?
(c) Describe two ways voice patterns may be used and in each case, why this might be a problem

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 Foursquare dataset (data, readme) 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/2.2.0`), 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. Show an alarm and put visitorID and info about their hardware/browser/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/2.2.0 > changes.patch`, with implemented 3 and 4.

**Problem INFO1 (6 points)**

Choose one of these Cases and answer the questions below:

Case 1: Parents purchase and install Social Mood Panda on their kids mobile phones to alert them when their kids are unhappy.

or

Case 2: The city of Metropolis is installing license plate readers at all intersections. This 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. To address privacy concerns, they would post clear notices.

(a) Point to two instances in which data flows enabled by the system in question would disrupt typical data flows among the various parties. (Make sure you specify all the parameters.)

(b) Drawing on the 3-layers of evaluation of contextual integrity, construct an argument favoring either the entrenched (or existing) flows versus novel (i.e. the flow induced by the technology) in each of the two flow discrepancies you have identified. *(It's ok to make 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 third parties revealing, for example, your location in real time.

(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 (2 points)**

Using CI, construct an argument either supporting or opposing one of the information flows allowed in the privacy policy you’ve discussed in Problem 9. Provide a couple illustrations, or concrete instances of the flow you have identified.

**Problem INFO4 (7 points)**

(a) What do Reidenberg et al. mean by a privacy policy that is ambiguous through incompleteness.

Based on these two statements drawn from privacy policies, answer b) and c) 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.”

(b) In what way is the statement drawn from the Comcast privacy policy vague.

(c) Drawing on Contextual Integrity, compare the completeness of the two statements.

(d) Based on Reidenberg’s discussion of ambiguity, identify 3-5 ambiguous terms from the given privacy policies. Explain their ambiguity, what misunderstandings might stakeholders have? (Hint: Use Table 2 and Appendix Table A3 from Reidenberg).

(e) Using the CI Analysis Method (Shvartzsnider) annotate 1 sentence from your group’s privacy policy following the model below:

We [Facebook]_{recipient} also collect contact information that you provide if you upload, sync or import this information (such as an address book) from a device.\textsuperscript{TP}

(f) In the language of CI, Privacy Policies describe a company’s data practices and associated data flows. Identify and highlight two such descriptions that fail to specify a value for one of the CI parameters, e.g. recipient, sender, and/or transmission principle?

Problem INFO5 (1 point)
Per a robust finding in Turow et. al’s surveys, most people believe that if a company has a privacy policy, it won’t share information it collects with 3rd parties. Why is this relevant for the existing regime of privacy regulation?