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

COMMON FOR CS AND INFO

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 (8 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 3 (2 points)
TraceTogether is Singapore’s proximity-tracing system, consisting of a mobile app and a physical Bluetooth token. TraceTogether is mandatory in most public places in Singapore, including workplaces, schools, shops, restaurants, etc. Its primary purpose is 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, i.e., updating the keys for each message? Name and explain each property. Be very 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 (6 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)
According to CI, a right to privacy is a right to _____________?

Problem 8 (2 points)
Explain in terms of Contextual Integrity the common definition of privacy as a right to control information about oneself.

Problem 9 (2 points)
“All the parameters matter!” Explain.

Problem 10 (8 points)
Using the privacy policy discussed in your breakout group (Jamboard Exercise 1, February 22) answer the following questions. NOTE: you may copy directly from your group’s jamboard, or you may augment these answers with language of your own (please indicate where you’ve done this) or some combination.

(a) Describe your experience locating or navigating to the Privacy Policy on the website. Briefly describe the website, i.e. services it offers. What information is collected through the website or app? How is the information collected? What information is collected about site visitors from third parties? Are these third parties specified? (e.g. data brokers, Google, Facebook, Government Agencies)

(b) Does the site mention conditions under which information is collected from users, collected from third parties. (i.e. transmission principles) How long is the information held; is it ever deleted/destroyed? Can the privacy policy change? How are visitors notified of changes?

(c) 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).

(d) Using the category of vague terms on page 7 (Reidenberg) annotate 1 sentence from the privacy policy to highlight condition, generalization, modality, and numeric quantifier.

Problem 11 (3 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 12 (6 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 13 (7 points)

(a) Describe the case your group discussed in Class in Jamboard Exercise 2, Feb 24?

(b) Now, choose one of the other cases and perform the same analysis, i.e. addressing the sub-questions on the Jamboard, questions below).

1. Point to two instances in which novel flows brought about by the system in question would disrupt existing data flows.

2. Draw on the 3-layers of evaluation to make an ethical argument favoring the entrenched (or existing) versus novel (i.e. the flow induced by the technology) in each of the flow discrepancies your group has identified. (It's ok to make some assumptions in order to build your arguments. Be sure to reveal these assumptions.)
Problem CS1 (30 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 SecureBot 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. SecureBot 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 SecureBot 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 (7 points)

(a) 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. Do you agree or disagree, explain.

(b) As discussed in class, in 2021, citing privacy, Apple made changes to the advertising environment on iOS 14.5. How might this change affect privacy? (Hint: Use CI)

Problem INFO2 (9 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 this statement drawn from the Comcast privacy policy vague.

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

Problem INFO3 (4 points)

Using the CI Analysis Method (Shvartzsnaider) annotate 1 statement from a privacy policy following the model below:

We [Facebook]_{recipient} also collect contact information attribute that you_{sender} provide if you upload, sync or import this information (such as an address book) from a device. \(TP\)
Problem INFO4 (10 points)
Based on Prof. McGuigan’s says “lecture answer the following questions:

(a) In the context of advertising, what is attribution, and why is it important for advertisers?

(b) Programmatic advertising serves ads based on individual users or what is known about the individual users not based on the website. (For example, seeing an ad for Nike Shoes on the NYT). If precise attribution means making claims about consumer behavior across websites, explain how this challenges Contextual Integrity (Drawing on your own experience with online ads, use two concrete examples to highlight this).