COM 402 exercises 2023, session 11: Privacy: definitions and properties

Exercise 11.1

- Are the following statements true or false? Justify.
  1. It is possible to deploy surveillance only on end-users of systems.
  2. Privacy as control ensures that only the minimal amount of information is provided to the service.
  3. To provide users with anonymity when accessing a web all accesses from one user must be unlinkable.
  4. Fine-grained accountability and auditability make it difficult to implement systems with strong privacy protection.

Exercise 11.2

- Consider a privacy-preserving forum to ask questions in the class. To provide privacy, when a student posts a question, instead of publishing the student's name, it chooses uniformly at random another name in the class that starts by the same letter. Discuss what is the privacy this mechanism gives in terms of error the professor for the following students. Who has more protection?
  - Charlie, who is in the class with Celia, Carla, Constantin, and Colin.
  - Louisa, who is in the class with Lorenz, Lex, and other two Louisas.

Exercise 11.3

- Aggregation is a privacy-protection technique consisting in regrouping data before processing (more on this in the machine learning lectures). Discuss what kind of privacy is this from the point of view of the paradigms (confidentiality, control, practice) or adversary (social, institutional, anti-surveillance) when:
  1. the aggregation is made locally by the user before releasing her data.
  2. the aggregation by all users is made by a third party.

Solutions to the Exercises

Solution 11.5

If these actions are done in the order said in the exercise, none of them provides plausible deniability: the adversary always knows that the first is real and the others are fake.

If the actions are randomized, only the first provides plausible deniability. Given the 4 searches, as any word in the dictionary could be selected, the adversary cannot identify the real one. For the second option, the adversary knows that words in the pre-curated list are likely not the ones searched for real. For the third option, there is no plausible deniability here: the search is always the one that is made by the user. The synonym may provide some privacy, but essentially is almost the same as the real search.

Note that over time, if a user repeats searches, or searches several words on the same topic, those may be recognizable as they appear more often than the randomly searched terms.

Solution 11.4

If these actions are done in the order said in the exercise, none of them provides plausible deniability: the adversary always knows that the first is real and the others are fake.

If the actions are randomized, only the first provides plausible deniability. Given the 4 searches, as any word in the dictionary could be selected, the adversary cannot identify the real one. For the second option, the adversary knows that words in the pre-curated list are likely not the ones searched for real. For the third option, there is no plausible deniability here: the search is always the one that is made by the user. The synonym may provide some privacy, but essentially is almost the same as the real search.

Note that over time, if a user repeats searches, or searches several words on the same topic, those may be recognizable as they appear more often than the randomly searched terms.

Solution 11.1

1. False. Developers and CEOs of companies, government employees, and in general everyone is at the end of the day an end-user. Once the surveillance infrastructure is deployed, everyone will be under surveillance.
2. False. The paradigm of privacy as control does not really focus on quantity. It focuses on the user knowing how the information is going to be used, but not on minimizing the amount of information disclosed.
3. True. If accesses by a user are linkable, even if we do not know the identity, these accesses become a pseudonym. We cannot anymore say that it is truly anonymous. Thus, in general, unlinkability is needed for anonymity.
4. True. Accountability and auditability mainly rely on logging actions. These logs typically record all actions in the system, becoming an extra source of information that can be used to infer private information about users.

Solution 11.2

In both cases the students enjoy anonymity among the other 5 students. The professor has 1/5 probability of guessing correctly, and 4/5 of making an error. In the case of Louisa, the professor succeeds in guessing the correct name 3/5 of the time, but he still cannot know with certainty which of the two Louisas wrote the question. Both students have the same protection.

Solution 11.3

We have two cases here:

1. If aggregation is local, from the point of view of the paradigms, aggregation can be seen as an obfuscation mechanism that aims at achieving *privacy as confidentiality*. The idea is to not give the adversary any information about individuals. As such, we can also categorize it as *anti-surveillance* privacy.
2. When aggregation is on a third party, then, with respect to this party that sees all the data the protection is *privacy as control*: we give the data to this party to only perform the aggregation, and only share with other parties the aggregated value. We could still say this is an *anti-surveillance* privacy mechanism from the point of view of the final entity that receives the data, but with respect to the aggregator we would be under *institutional* privacy assuming that this aggregator is semi-trusted and will do what is agreed upon and nothing else.