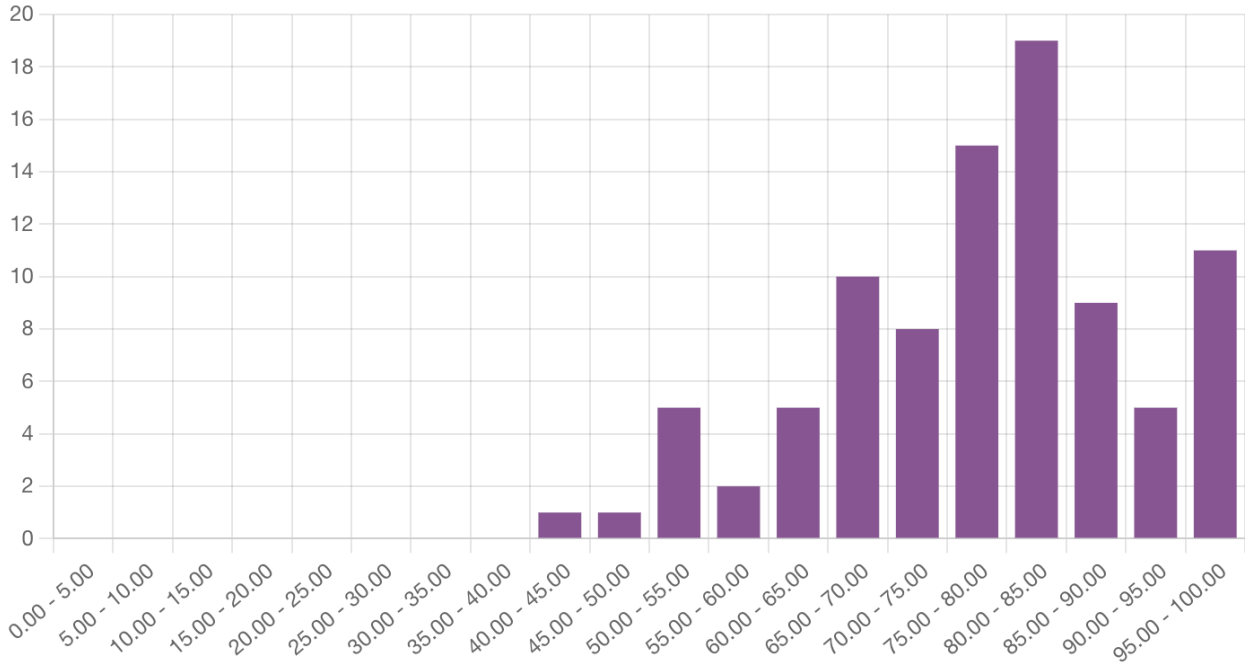
CS-523 Final - Most Repeated Errors

Spring 2024

Grade Distribution



General Advice

1. **Answer (all parts of) the question**
   Many points are lost because either:
   a. Some parts of a question are not answered (e.g., not stating threat model or compare utility of different implementations), or
   b. The answer ignores constraints in the question (e.g., answering about a different adversary model than indicated in the question) OR the answer completely misunderstands the question (e.g., stating attack instead of privacy concern).

Question 1

1. **Doing division on local shares in SMC**

Some answers use addictive secret share, let each party do division on local shares, then aggregate the local division result to get the final percentage. This way of computation does not result for the correct percentage. E.g., if there are three parties who would like to compute b/c with secret values b = 6 and c=6. [b] = [1, 2, 3]; [c] = [1, 2, 3]. The result will be wrong if the parties do division from local shares and add them up at the end, i.e., [6/6] != [1/1] + [2/2] + [3/3].

2. **Not dividing the number count by the total number of students**
   This question asks for a percentage, not the number of students who come to the university by bike.

Question 2

1. **Presenting a trivial "attack"**
   Some answers present a variation on the following "attack": the university forces all students to come to the university by bike. This is not an attack, and the output of the computation corresponds to the physical reality.

2. **Not specifying the threat model and capabilities of the university**

When outlining an attack, please specify under which threat model you are operating (e.g., the university colludes with up to *t* students, or the university is fully malicious and tampers with the computation), as well as its capabilities.

Question 3

1. **Assuming there is a proof of correct FHE computation**
   The question specifically mentions that the proof only guarantees that all registered accounts in the system were used; in particular, there is no mention of a proof of correct FHE computation, and a malicious cloud would still be able to tamper with the computation and alter the output.

Question 4

1. **Lack of detail about how Tor can still work**
   Many answers did not indicate that DNS queries were redirected via Tor. If DNS requests are not routed via Tor, they can still be blocked by the government.

2. **Assuming that the Government is a global adversary**
   The government's monitoring capabilities are confined to the country where there is the conflict zone. It has no broader visibility and control over a larger scope, such as the entire Tor network. The government is, thus, a (powerful) local adversary.

3. **Fail to connect to the internet, then use Tor**
   The fact that someone uses Tor does not mean that they have been to a specific website. Indeed, depending on the number of people using Tor in the country, the anonymity set may vary. If a user has previously made a blocked connection to the website, that doesn't mean they'll do it again later via Tor. This may be a valid argument, but it is much weaker than network analysis.

4. **Lack of detail on the technique used by the government to detect the communication**
   Due to its capabilities, the government can perform a correlation attack to verify that Cameron sent the files using OCT's employees' website via Tor. We expected answers that explained (as the question asked) how the government could correlate packets coming from one student and arriving to the other student.

5. **Provide an alternative means**
   To evade censorship: do not imitate, be!

Question 5

1. **Concluding that the protection described in the question is enough to prevent both stateful and stateless tracking, so Cameron is safe from being tracked on the web**
   The defenses mentioned in the question mainly prevent stateful tracking, stateless tracking (e.g. browser fingerprinting) is still possible. In fact, using a new privacy-preserving browser adds a lot of uniqueness to be exploited from browser fingerprints.

2. **Not providing enough detail on a fingerprinting-based attack**
   Stating that it's possible to track using browser fingerprinting but not elaborating for example on what information or features enable the fingerprinting to successfully track Cameron. This resulted in partial points.

3. **Using personal information such as name/email address to track Cameron**
   Not providing enough information about the assumptions placed on the entities that would need to collaborate to track the user across different websites (e.g. government, different website owners), led to partial reduction of points.

Question 6

1. **The privacy concern presented is not about the users of the app**
   We asked to provide a privacy concern that arises for users of the app. Formulating a privacy concern for employees or YourRecipe was not a valid

answer.

2. **Incomplete or missing threat model**
   When formulating a threat model, it is important to formulate the capabilities and goals of the adversary. Just stating “YourRecipe gets the images”, is not enough. You must describe under what precise assumptions the privacy concern materializes. For example: “Under the assumption that the users send a picture of their fridge to a YourRecipe server that uses the model to infer the list of ingredients, and that the server is honest-but-curious, the privacy concern materializes”.

3. **Formulating an attack instead of a privacy concern**
   Explaining an attack without explaining what the adversary is able to do with the information inferred from the attack is not a valid answer. For example, stating that the adversary can perform a MIA is not enough. We ask for a privacy concern for the users, which can materialize through an attack, but they are not equivalent.

Question 7

1. **Stating that gradients do not leak any information**
   Gradients leak a lot of information about the users’ data in implementation B. Gradient inversion attacks are possible and can reconstruct the batch used by the users.

2. **Not mentioning utility**
   We asked you to detail the impact of both implementations on utility for the users, not just how well it protects against the privacy concern.

3. **Not being clear on the threat model used**
   We asked you to be explicit about the threat model used for your comparison. Just stating “the adversary” is not enough, especially when multiple threat models were mentioned in the previous question.

4. **Saying “the attack/adversary does not work” instead of discussing the privacy concern**
   Indeed, to say that there is no concern, you must show that *no* attack/adversary would work, not just one you choose.

Question 8

1. **Stating that DP does not change how much gradients leak**
   Instance-level DP does protect against reconstruction attacks, reconstructing exact images is made much harder. If the privacy concern stated in the first part involves the reconstruction of exact images, then DP does provide better privacy guarantees.

2. **Stating that DP protects against property inference attacks without justification**
   Users' batches are likely highly correlated data (which is reasonable from the same fridge from the same user), hence the gradients are still vulnerable to property inference attacks even with instance-level DP. So, unless the answer assumed non-correlation, an adversary would still be able to infer some general trends of the user's data: for example, if they don't have certain types of food, or if they have medicine in their fridge.