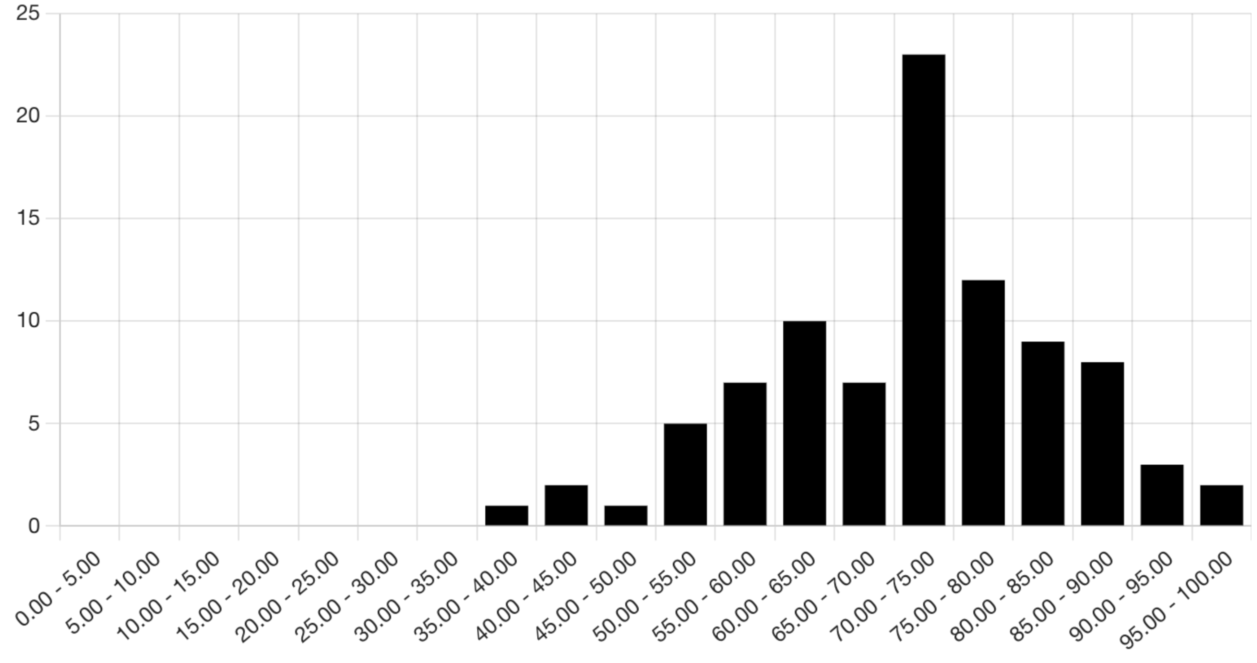
CS-523 Midterm Most Repeated Errors

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 capabilities), 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 capabilities instead of describing an attack).
2. **Keep it short and concise**
   Don’t write too much or give several alternative answers: extending your answer beyond what is asked will not result in bonus points. Moreover, if the additional details written are incorrect it will result in loss of points.
3. **Quoting the lecture slides without proper justification**
   Do not just quote facts from the slides without providing justification about why and how that fact applies to the question. For instance, saying that the

predecessor attack applies in Crowds without mapping the attack to the scenario in the question. Such answers do not allow us to understand whether the concepts in the lectures are well understood.

Question 1

1. **Lack of detail about the professor's capabilities**
   Many answers did not state the capabilities the professor requires to infer that Mario and Luigi are talking to each other. The professor's capabilities determine what she can observe in the network to carry out the analysis. In this case, the professor can see packets incoming and outgoing the Tor network over time so they can correlate flows and know who speaks with whom.

2. **Assuming that the professor is a global adversary**
   The professor's monitoring capabilities are confined to the room where Mario and Luigi are taking the exam. She has no broader visibility and control over a larger scope, such as the entire Tor network. The professor is, thus, a local adversary.

3. **Using Tor means communicating with a fellow student**
   The fact that two students were using Tor does not mean they communicated. For example, Mario and Luigi could be using the Tor browser (which uses the Tor network by default) to access the course slides on Moodle.

4. **Lack of detail on the technique used by the professor to detect the communication**
   Due to their capabilities, the professor can perform a correlation attack to verify that Mario and Luigi communicated via Tor. Answers that did not explain how the professor made the inference (e.g. just stating that the professor monitors traffic) were penalized. We expected answers that explained (as the question asked) how the professor could correlate packets coming from one student and arriving to the other student.

Question 2

1. **Incorrect/implicit assumptions on messages being encrypted**
   Answers that do not explicitly state their assumptions regarding encryption of messages or do not justify their claim about the privacy Crowds provides and how that helps in this question were penalized. For instance, answers that just stated that "Crowds obscure communication" or "Crowds make it harder to detect communication between two parties" without **further** explanation regarding why or how did not receive full points.

2. **Missing justification on how the predecessor attack maps to the question**
   Answers that simply state that the predecessor attack works on Crowds without explaining how this attack would be carried out by the professor in the scenario of the question to infer that Luigi and Mario are communicating did not receive full points.

3. **Underspecified assumptions regarding the configuration of the scheme (Also applies to question 3)**
   Some answers did not specify explicitly either the participants or the envisioned deployment of the mechanism i.e. Crowds, DC Networks, or Mix-Nets. For example, do only Mario and Luigi use the mechanism or are there more participants (more students or the entire class)? is mixing done at nodes inside or outside the LAN in Mix-nets?
   Often, the answers are true only under specific assumptions and are incorrect under other assumptions. Without the answer stating explicitly what was assumed, it is not possible to assess whether the student understood the lecture concepts and whether the answer is correct. Such answers did not receive full points.

Question 3

1. **Quoting slides to state which scheme provides what properties**
   Simply stating the anonymity properties provided by a mechanism without justifying why that property is important to help Luigi and Mario to communicate in the question. For example, some answers stated that “DC network provides sender and receiver anonymity, so it’s a good choice for Mario and Luigi to communicate”. The answers that did not justify why sender and receiver anonymity are sufficient to hide Mario and Luigi’s communication from the professor, given the professor’s capabilities as an adversary (For instance, what prevents the professor from still performing statistical analysis?). Such answers did not receive full points.

2. **Focusing solely on privacy without acknowledging the potential impact on usability and performance**
   Some answers stated that “Mix-nets enable Mario and Luigi to communicate without being detected by the professor, especially if high delays are introduced in the communication that make correlation attacks more difficult”. These answers fail to acknowledge that such delays can make real-time communication difficult, especially during an exam. Such answers did not receive full points. Answers that stated the trade-off explicitly received full points.

3. **See MRE #3 from question 2**

Question 4

1. **Misunderstanding the goal of the attack**
   The question was about how to obtain profiles to send targeted advertisements to Cigros customers. To this end, Cigros needs to find a way to link one or multiple visits to a customer's identity (i.e., de-anonymize the traces). Then, Cigros can match the behaviors seen on the traces to a specific customer's identity and send them ads. One visit is enough to profile customers and send them targeted advertisements. Some answers described an attack linking two or move visits from the same client which is not enough to actually send targeted advertisements. Linking multiple visits (and corresponding traces) by the same customer would allow to create more precise profiles for each customer but if Cigros is not able to get the identity of this customers and de-anonymze the traces then they will not be able to send any targeted advertisements.

2. **Not explaning how the attack links traces to identities**
   Simply stating that Cigros can link the anonymous traces to credit card information does not allow us to assess whether the concepts of the course are understood. The question asks for a *description* of the attack. Describing the attacks means providing details that explain how exactly this link is made: for example, through linking the items on a receipt to the items picked up by an anonymous client on a video trace, or through matching the timestamp on a receipt to when an anonymous client was seen to be paying on their trace.

3. **Stating that credit card information is not enough to de-anonymize the traces:**
   Credit card numbers can indeed be pseudo-identifier, however credit cards also carry the customer's name and therefore their identity! This together with linking the traces with the corresponding credit card is enough to de-anonymize the traces.

4. **Adversary's capabilities not being clearly formulated**
   Capabilities are the information and computing abilities that makes it possible for Cigros to carry an attack. Here, Cigros' capabilities are its ability to access credit card information during payment as well as being able to construct traces from the full coverage of the shop with infra-red cameras. Stating that Cigros can link the anonymous traces to credit card information is not a capability, but rather the attack made possible by Cigros's aforementioned capabilities.

Question 5

1. **Mixing unlinkability and anonymity** (or **pseudonymity)**
   Saying “Cigros cannot link a trace to a customer, hence customers are unlinkable” is wrong. Unlinkability is a property of events, i.e., is Cigros able to link two visits of the same customer together? While anonymity is the property of mapping an event to an identity, i.e., a visit to a customer name.
   Note that there is a link between those two properties: when there is no anonymity, there is no unlinkability. Indeed, if Cigros can de-anonymized visits, then Cigros can link two visits of the same customer together by first de-anonymizing them and match them based on customer’s name.

2. **Claiming unlinkability without a justification**
   If the property was claimed without justification, we did not give points. Customers are likely to buy the same products, or can be recognizable through their gait, or the time of visit, etc... This set of attributes, available to Cigros, makes customers linkable across visits, hence breaking unlinkability.
   Given these, we did not expect answers claiming unlinkability, but we accepted them when it was properly justified and argued for, e.g., if the answer pointed out limitations in accuracy of such linkage attacks.

3. **Adding assumptions outside of the question**
   Some answers relied on assuming Cigros’ knowledge or capabilities that were beyond the scenario specified in the question. For example, assuming that users have a fidelity card, that Cigros has access to biometrics, or that Cigros has access to users' location data. Such extra assumptions, which are not about making concrete unspecified aspects but adding extra information, change the scenario, Thus, the answers -- whether right or wrong – do not answer the exam question.

Question 6

We graded this question according to the **justification.** Both “No”-answers and “Yes”-answers with valid arguments received full points. We list common mistakes separately for each kind of answer in the following. For convenience, we abbreviate Differential Privacy as DP.

1. **Common Mistakes When Arguing “DP does not work”**
   a. **Employees joining and leaving breaks DP.** The set of people changing over time itself does not invalidate the protection given by DP. DP aims, by definition, to ensure that the result does not change depending on whether a user is or not on the dataset. Other changes in the database, such as a change on specialization categories, i.e., a user consults a doctor with a new/unseen

specialization, would break DP, as it would affect the sensitivity computation which would change the parameters of the system.

2. **Common Mistakes When Arguing “DP works”**
    - a. **Considering record-level privacy instead of user-level privacy.** To achieve record-level privacy (i.e., privacy of whether a visit to one specialization happened), removing each record only changes one count of one specialization for the table. Hence, the sensitivity in record-level privacy is one. However, in the question, each user can affect more than one record for more than one specialization. If we calculate the parameters under record-level privacy, once a user changes several records for several specializations, the noise does not give any bound to the information leakage of a user. Therefore, this question requires user-level privacy (i.e., privacy of whether a user is included or not in the dataset), and we must compute the parameters of DP, e.g., the sensitivity, considering how much the presence or absence of a user can change the result.
    - b. **Using the actual values in the database to compute the sensitivity parameter**. DP is designed to protect users in the worst-case scenario. Therefore, one cannot compute its parameters based on the data in the database, which may not contain such worst case (as discussed in DP lecture slide 27). Instead, one has to think about what the worst-case scenario in the application is, and use that as input for the parameter computation. For instance, it is possible that there is no visit to a general practitioner from employee A in the current dataset, however, this does not mean that the sensitivity is 0 for employee A. It is possible that A gets a serious cold and visits the general practitioner many times in a month in the future (as for the explicit upper bound of sensitivity, see the following discussion in point d).
    - c. **Assuming multiple visits to a specialization from an employee is counted as just one visit.** The question clearly specifies that “Every month, Garanta aggregates the records of all employees to compute *the total number of doctor visits* per medical specialization”. Assuming that Garanta only counts how many specialists have been visited is a different problem. Thus, the answers under such assumption -- whether right or wrong – do not answer the exam question.
    - d. **Arguing an upper bound on the sensitivity without reasoning how the upper bound is derived and why it makes sense**. An example answer without reasoning is “the sensitivity is at most 23 per specialization per month”, without adding more details of how this conclusion is reached and whether it is arbitrary or corresponds to real constraints. One example of a well-justified answer is “the upper bound of sensitivity for an employee

per month per specialization is 23=31-8, assuming the person 1) cannot visit a specialization more than once per day in a month with at most 31 days and 2) cannot visit during weekends, which removes 8 days per month".

e. **Defining the privacy budget of each specialization based on a subjective perception of the sensitivity of different specializations**. DP parameters should not be computed based on the perception of the implementor nor the social norm (what most people think at a certain time etc.), because in such case the privacy loss would depend on the specific scenario of a user. For instance, it was almost not sensitive to visit a general practitioner before Covid, however, the visit to GP became super sensitive during Covid, which no one could predict before Covid happened.

f. **Arguing that a person can visit certain specialization more times than others and split the privacy budget unevenly.** This uneven allocation of budget risks putting larger epsilon of some specializations, which results in more privacy loss of arbitrary specializations compared to the ones with smaller epsilon.

g. **Not using sequential composition but parallel composition for each specialization of an employee.** Each user can visit multiple specializations, removing one user can potentially affect the total number of visits for all specializations. The records per user are *not disjoint across specializations,* and hence, we cannot use parallel composition. See Friday Live Exercise on DP Question Waterwolf Part 2 for a similar scenario which results in sequential composition across different websites.

3. **Common Mistakes for Both Answers**
   a. **Arguing that an adversary can increase their knowledge by seeing the results even after applying DP**. A correct implementation of DP, by design of considering the worst-case scenario of a user's change to the result, will not reveal more information to the adversary apart from its background knowledge.
   b. **Incorrect threat model that says Garanta is not trusted or arguing about using input perturbation instead.** An insurance company must have access to accurate information of client medical records, otherwise, insurance services, e.g., reimbursement, cannot happen.

Question 7

1. **Stating Mechanism 1 is unforgeable because it uses a collision-resistant hash function**
   Collision-resistance only states that if the adversary does not have access to the

hash input, it is infeasible for them to find another input that maps to the same hash output. The input to the hash function is an AVS number, the enrolment month, and the current month. In this setting, the adversary can either get access to the hash input easily (e.g., Garanta might know the AVS number, a colleague may see the AVS number from some documents lying around and knows when their colleague started at the company), or it can brute-force the inputs easily (12 possibilities for the month, at most 7 million ≈ 2^{22} possibilities for the AVS number). Hence, collision-resistance of the hash function does not give any guarantee about an adversary not being able to forge a credential.

2. **Stating Mechanism 1 provides verifier unlinkability because it uses a pre-image resistant hash function**
   The verifier has access to the inputs to the hash, and can thus enumerate all possible combinations of inputs, hash them, and compare the result; this allows them to link credentials even if they cannot recover the values from the hashes.

3. **Mistaking the issuer to be the Swiss government**
   The issuer of the credentials in the question scenario is Worried Co. (cf. “Worried Co. issues each employee an anonymous credential per month”); the Swiss government only assigns the AVS number to people but is never involved in the system described in the question.