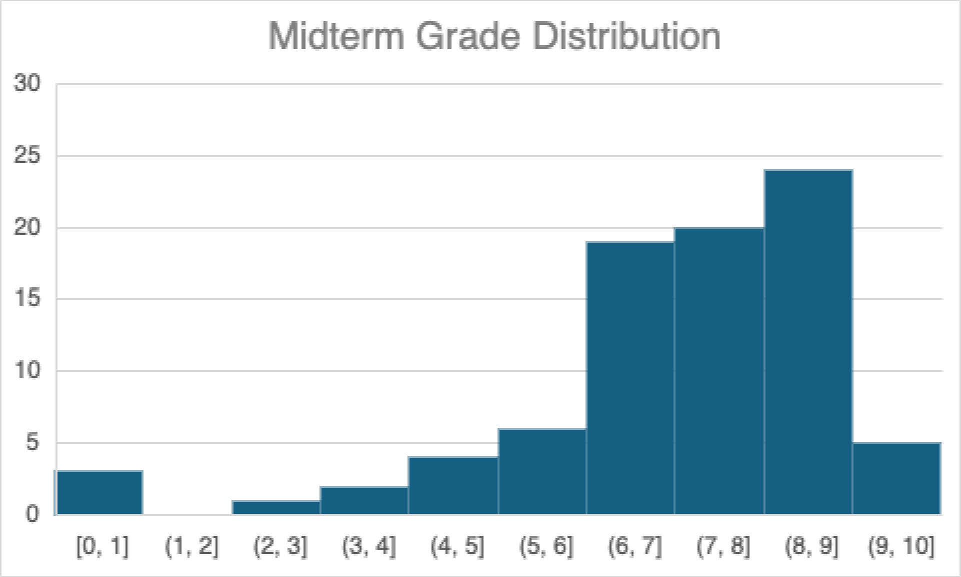
CS-523 Midterm: Most repeated errors

Spring 2025



General Advice

Answer (all parts of) the question. Many points are lost because:

1) Some parts of a question are not answered, e.g.,
   - The answer misses to state a threat model
   - The answer only describes a privacy concern but does not define an attack under which this concern might materialise
   - The answer is missing a justification

2) The answer ignores constraints in the question, e.g., the answer is under a different adversarial model than indicated in the question.
3) The answer completely misunderstands the question, e.g., states an attack instead of privacy concern.

SMC: NoFly

Q1

Use AND gates to check for equality. Many answers used AND gates in their proposed implementation of the circuit f to check for equality between two input values. This is wrong because AND gates cannot provide this functionality. Instead, to check for equality between two values either equality gates or alternatively an XOR gate should be used.

Use BGW circuits to implement boolean comparisons. Some answers proposed a boolean circuit f, and wanted to use BGW to compute this circuit. However, BGW handles arithmetic circuits. Instead, garbled circuits need to be used if you want to support a boolean circuit.

Q2

Assume that an honest-but-curious adversary cannot submit arbitrary inputs. Some answers stated that the desired privacy properties could only be achieved under an honest-but-curious threat model because such an adversary, as opposed to a malicious one, could not input arbitrary values. This is an incorrect justification. While an honest-but-curious adversary is bound to behave honestly during the execution of the protocol, it can supply arbitrary inputs.

Q3

Miss a possible attack. Some answers incorrectly stated that Sauron cannot launch an attack since he is assumed to be honest-but-curious and thus cannot lie about his input values**.** This is incorrect (see also MRE above). An honest-but-curious adversary is bound to honestly follow the protocol and correctly execute it but might still supply arbitrary inputs. Thus, Sauron, even if he acts within the honest-but-curious model, could achieve his adversarial goal by lying about his passport number.

ABCs: Referendum

Q4

Argue that attribute age should be voter-defined. Some answers argued that the attribute age in the credential should be voter-defined because a voter's age, like their name or ZIP code, is an attribute naturally linked to a voter. However, this argument ignores that in order to achieve Genovia's voting requirement, that only people above 18 should be allowed to vote, age must be an issuer-defined attribute. Otherwise, the credential-based voting scheme would be subject to a trivial attack by which voters who do not fulfil the age requirement could lie and would be able to vote.

Q5

Invent additional voting requirements. Some answers suggested to disclose more attributes than the minimal set needed to achieve Genovia's voting requirements. In many cases, the disclosure of these attributes was justified by inventing some voting requirements, such as ensuring a unique vote, which were not in the original question. Because the question explicitly asked for the minimal set of attributes that needs to be disclosed *to fulfil the given voting requirements*, those answers did not receive full points.

Incomplete or missing justification for selected attributes. Many answers stated the correct set of attributes to be disclosed but did not fully justify their choice. These answers did not receive full points.

Assume additional steps during registration or voting. Some answers assumed some additional steps (e.g., an age check) during registration or voting that were not defined in the original question. Answers who disclosed an incorrect set of attributes based on these assumptions or incorrectly justified their choice of attributes with these assumptions got points deducted.

Zero-knowledge proofs as a wildcard solution. Several answers argued that certain attributes did not need to be disclosed because zero-knowledge proofs could be used to verify them. Modifying the ABC scheme is out of the scope of the question: we only ask about what attributes to disclose.

Q6

Assume that collusion between issuer and verifier leaks undisclosed attributes. Some answers assumed a collusion between the credentials' issuer, the central office, and the verifier, the counting office. These answers then argued that such a collusion would enable the counting office to reveal a voter's ZIP code. However, this would not work due to the unlinkability properties of anonymous credentials. Neither the issuer nor the verifier can link a credential at the showing step back to a voter. A collusion would thus not allow the counting office to learn a voter's ZIP code if the attribute is not disclosed.

ZIP code statistics based on assumed correlations with a voter's age. Some answers argued that the counting office could compute the average vote per ZIP code purely based on correlations with a voter's age; the attribute disclosed to the counting office during the showing step. These answers did not receive full points. First, in most cases, the assumption about correlations between a voter's age and ZIP code was not well justified or explained. While there might indeed be a (weak) correlation between these two attributes, it would need to be a close to perfect correlation for this to be a valid answer. Second, these answers did not answer the question which explicitly asked for the average vote per ZIP code; which would be different from its approximation based on correlations with voters' ages.

Q7

Assume that voters can *create* new credentials. Some answers assumed that voters could create new credentials without involvement of the issuer, the central office. However, due to the unforgeability property of anonymous credentials, a valid credential cannot be created by a malicious voter without the central office.

Assume that voters can ask the central office to issue multiple credentials. Some answers assumed that a malicious voter could repeatedly ask the central office to issue new credentials. However, the question stated that the central office would only issue a credential to voters once, i.e., if they had not come before and that this would be validated by a name check. The assumption that the central office would issue multiple credentials to a single malicious voter thus directly contradicted the question statement.

Crafted value in the vote field. Some answers suggested that a malicious voter could manipulate the outcome of the vote by putting a crafted value, e.g., -1000 or +1000, in the vote field but did not specify under which assumptions this attack would work. This is a great answer *under the assumptions that* the counting office counts votes in a blind/automated way and that the standard format of a referendum vote (a binary value) is not enforced in any way.

Modified value in vote field. Attacks where the vote value is modified in between the issuance and showing step (i.e., the credential issued has a “correct” value 0/1 which is then modified to +/-1000 before showing) are not valid, as the modified credential would not verify anymore. Otherwise, this would break the unforgeability property of anonymous credentials, since voters could create valid credentials with any attribute value.

Data Publishing: ICBC

Q8

Miss a possible quasi-identifier. Some answers missed that an internal adversary could use the number of family members as an additional quasi-identifier to single out a target family in the anonymised dataset. Because the generalised data achieves k-anonymity only with respect to attributes “City” and “Distribution month”, it is possible to conduct linkage attacks that uniquely identify a target family using family size as an additional quasi-identifier.

Missing specification of adversary’s background knowledge. Many answers failed to clearly define the adversary’s background knowledge and how it can be used to compromise privacy. As mentioned in the course, at the time of data publishing, we cannot know what auxiliary data may be available to the adversary. Hence, it is important to clearly state what additional information, other than the shared data, the adversary possesses, and how it can be used in combination with the shared data to perform de-anonymisation attacks.

Q9

Missing specification of adversary’s background knowledge. Many answers missed to clearly define what background knowledge an adversary must have to single out a target family. Often, these answers only vaguely mentioned that an adversary can use “background knowledge that is only linked to a single family”, without explaining what exactly this knowledge is about. These answers did not receive full points.

State a privacy concern without giving an attack. Some answers correctly stated a privacy concern but missed to describe an attack under which this concern might materialise.

Assume that an adversary needs to be malicious to conduct privacy attacks. Some answers stated that an adversary must to be malicious to materialize the privacy concern. This is incorrect because the adversary does not need to deviate from the protocol to launch privacy attacks. Instead, even an “honest-but-curious" adversary that has the necessary background knowledge is sufficient to materialise certain privacy concerns.

Q10

Ill-defined privacy entity. Many students chose the wrong privacy entity, like an individual, instead of the privacy entity defined in the question (family).

Ill-defined sensitivity. Many students did not choose the correct sensitivity for the Laplace mechanism. Sensitivity is defined as the maximum possible impact a *single privacy-entity* can have on the output of the computation. In this case, sensitivity of the computation could either be defined as “the maximum amount of meal kits a family could receive in a month” (impact on one statistic) OR “the maximum amount of meal kits a family could receive in a year” (impact on all statistics). Depending on the chosen definition of sensitivity, parallel and sequential compositions then needed to be applied differently in the second part of the question (see MRE below).

Wrong choice of sequential and parallel composition. Many answers incorrectly applied sequential and parallel composition. If the sensitivity was defined as “the maximum amount of meal kits a family could receive in a year”, then there was no need to apply sequential composition and divide the privacy budget among statistics. However, if the chosen sensitivity was “the maximum amount of meal kits a family could receive in a month”, then parallel composition should be applied across cities/family size **but** sequential composition **needs** to be applied across the 12 months of the year i.e. the privacy budget needed to be divided by the number of statistics a family could appear in, which is 12.