Question 1: Privacy-preserving Data Publishing (10 points)

`BecomeAChef` is a website that publishes recipes. They distinguish themselves from other sites by having recipes tailored to people with dietary restrictions. To win new users, their marketing team launches a campaign in popular newspapers. They decide to publish in each newspaper the following pseudonymised table about the recipes viewed by users during one week:

```
| userID | recipe | dietary_restrictions | day| visit_duration | comments_on_recipe |

userID: random ID assigned by the system to a user viewing the page, persistent across
views
recipe: name of the recipe
dietary_restrictions: typical dietary restrictions that can eat this recipe safely (e.g.,
celiac, halal, vegan)
day: day of the week in which userID visited the recipe's webpage
visit_duration: amount of time userID spent on the recipe's webpage
comments_on_recipe: text of the comments the user left on the recipe's webpage (blank if no
comment)
```

The marketing team hopes that curious readers will visit their website to read more about their recipes.

Part 1 (1 point): Describe a privacy concern related to the publication of this table, and an adversary (background information, capabilities, and goal) under which that privacy concern would materialise.

Part 2 (2 points): Propose a defence based on k-anonymity that would mitigate the risk identified in Part 1. Justify your choice of quasi-identifiers, and the mechanisms you would use to achieve k-anonymity on each of them.

Part 3 (2 points): Describe an alternative threat model (or adversary) in which your defence would not protect the privacy concern described in Part 1. If such a threat model would not exist, justify why.

Part 4 (5 points): After seeing the success of their newspaper campaign, `BecomeAChef` decides to monetize their data by allowing supermarkets to query their database with queries of the type:

"how many recipes for dietary restriction Y received more than X visits in [day_start, day_end]"

3.1. Describe one legal basis under which BecomeAChef could legally monetize their data in this way, and explain why it applies. (1 point)

3.2 Describe a privacy concern that could arise from the answers of this query. Describe a threat model under which this risk would materialize. (1 point)

3.3 Describe a mechanism that allows BecomeAChef to reduce the privacy risk. Your answer should sketch an algorithm and specify the details that would be needed to implement this algorithm. Explain a privacy risk mitigated by your chosen mechanism. (3 points)

Question 2: Privacy-preserving Encryption (10 points)

The University of Privacy Studies (UPS) wants to know who is the best professor according to the 1500 students of the university. For this, they ask students to vote for their three favourite lecturers among the 10 professors of the university. As UPS staff members are very concerned with privacy, they want to avoid knowing the individual votes of each student.

Part 1 (4 points). First, the UPS IT team proposes to let the students organize and run the vote and have the student representative reveal the result to the University. They have a friend that took CS-523 that tells them how to do the vote, since this was a live exercise of the class. The representative additionally wants to know how many students vote for the same two professors (for all pairs of two professors). The IT team asks the representative to come up with a proposal to do it in a privacy-preserving way.

Specify how the representative can use SMC to learn how many students voted for the same two professors without anyone having to reveal their full list of three votes in the clear. You may use more than one circuit to solve this problem. For your solution describe:

- the input that each student gives to the circuit(s),
- how each student computes their shares,
- to whom each student sends shares,
- which circuit(s) need to be implemented,
- how the final result is computed,
- the threat model under which your solution would provide privacy guarantees. Specify, for each of the actors in the system, what assumptions are needed on their capabilities.

Part 2 (2 points). The UPS IT team has heard the representative has a very favourite professor and decides it is better not to trust them to reveal the result. They heard of a company called Triserve which offers SMC as a service which would enable the outsourcing of the computation: students would send the shares of their votes to Triserve servers. Triserve has the following statement on their web page: “We have servers in three jurisdictions: Brazil, Hungary, and Bangladesh. Thus, you can safely run your SMC computations”. Would you recommend the university to use Triserve to ensure that no actor in this system can know which professors each student voted for? Justify your answer.

Part 3 (2 points). UPS learns that Triserve also offers a service called Hompute, which permits the use of Fully Homomorphic Encryption in one of their nodes. The IT team thinks that this may be a better idea than SMC. They read in the brochure that after three multiplications, the homomorphic scheme implemented in Hompute requires an extremely expensive bootstrapping operation.

(1) Can UPS use Hompute to find the best professor within a reasonable time? Justify your answer. (1 point)

(2) Under which threat model would the use of Hompute provide privacy guarantees? Describe, for each of the actors in the system, what assumptions are needed on their capabilities. (1 point)

Part 4 (2 points). If a student is malicious, explain two security or privacy threats that this student could materialise in either of the protocols above (in Part 1 or Part 3)

Question 3: Anonymous authentication (10 points)

(All details in this question are invented. Any resemblance with reality is pure coincidence. No Balélec organiser was harmed while writing this question)

The Balélec festival wants to gather statistics about who goes to which scene. This year, attendees can buy tickets that will grant them access to a subset of the stages. The tickets will be checked at every stage entrance. The organisers plan to use students' CAMIPRO cards to store the permissions associated with the ticket the student bought and allow access to different stages. One student of CS-523 says: "but what about privacy? Better we use Attribute Based Credentials!" They propose a scheme as follows:

- The Balélec ticket booth takes the role of the Issuer, and gives a credential to every user when they buy their ticket
- The credential will have the following attributes:
    - school: which faculty at EPFL the student belongs to
    - time_purchase: hour and minute at which the student bought the ticket
    - stages_allowed: the set of scenarios the student can access to listen to music

- At every inside/outside stage there is a guard that takes the role of the Verifier

- To enter a stage, a student runs an ABC show protocol with a guard at the entrance of the stage in which they reveal all attributes in their credential.

For all parts, assume that: (1) the ABCs are implemented using a scheme that provides all desired properties described in the lecture, and the implementation is correct. (2) the personnel selling tickets at the ticket booth will not act as guard during the festival.

Part 1 (2 points). The student that designed this system claims that "because it is based on ABCs, it will enable Balélec to count the number of spectators per stage without being able to track visitors". Agree or disagree with this claim. If you agree, describe which properties of ABCs enable the protocol to hide all information. If you disagree, describe how visitors can be tracked.

Part 2 (3 points). Propose a way to use the specified attributes to minimise information leakage (even if the information would not be linkable) while allowing guards to filter students based on the scenarios they paid for. Justify your choices.

Hint: Describe which attributes you would hide (if any), you would show (if any), or you would proof something about (if any). If there is any of the latter, write the Zero-knowledge proof that would need to be implemented

Part 3 (3 points). The organizers of Balélec raise the concern that these schemes may enable groups of students to buy one ticket, duplicate it, and use the duplicates to attend concerts at one stage together. Specify if this concern is valid for both ABC-based schemes, the one in Part 1 (proposed by CS-523 student initially) and the one in Part 2 (proposed by you). If it is valid, explain the conditions under which this concern can materialise and describe a defence to prevent explaining how it would work (in a similar form as the scheme described above). If it is invalid, justify why. Once a credential is verified, students get a stamp on their arm, specific to the stage and concert, which is used for re-entry to that concert.

Part 4 (2 points). Explain what a malicious booth operator (i.e., malicious issuer) could do to ensure that they can track users given your proposal in Part 2. Justify your answer. If they can do nothing, justify why.