

# EN4720: Security in Cyber-Physical Systems

## Exercise — Privacy

Name: Thalagala B. P.

Index No: 180631J

June 15, 2023

**This is an individual exercise!**  
**Due Date: 23 June 2023 by 11.59 PM**

### Section 1

Provide answers to questions 1 through 7 in the given space referring to the Tables **1** and **2** given below.

Table 1: Example dataset on the weekly usage of taxis by certain individuals

Name	Age	Gender	Occupation	Average No. of taxi trips per week
943145	21	Female	Legal Counsel	15
416765	38	Male	Data Privacy Officer	2
356891	44	Female	Database Administrator	3
723145	25	Female	Administrative Assistant	1
239976	31	Male	Data Privacy Officer	5
562396	42	Female	Programmer	3
964825	22	Female	Administrative Assistant	4
873892	30	Female	Legal Counsel	2

1. What type of data anonymization technique is used for the dataset given in Table **1**?

*Your answer here*

2. Is this technique sufficient to protect the privacy of the associated individuals? If not, why?

*Your answer here*

3. Briefly describe three other data anonymization techniques that can be used to anonymize data.

*Your answer here*

Table 2: Modified example dataset on the weekly usage of taxis by certain individuals

Age	Gender	Occupation	Average No. of taxi trips per week
21 to 30	Female	Legal Counsel	15
31 to 40	Male	Data Privacy Officer	2
41 to 50	Female	IT	3
21 to 30	Female	Administrative Assistant	1
31 to 40	Male	Data Privacy Officer	5
41 to 50	Female	IT	3
21 to 30	Female	Administrative Assistant	4
21 to 30	Female	Legal Counsel	2

4. The example dataset given in Table 1 was modified to improve anonymity. The new dataset is provided in Table 2. Mention all the data anonymization techniques that were used to achieve this.

*Your answer here*

5. Can k-anonymity be observed in the data given in Table 2? If so, what is the value of k?

*Your answer here*

6. Is privacy guaranteed for the data given in Table 2? Justify your answer.

*Your answer here*

7. Calculate the risk of re-identification for data given in Table 2 (mention as a percentage). Justify your answer.

*Your answer here*

8. Suggest ways to enhance the privacy of this dataset considering l-diversity.

*Your answer here.*

## Section 2

1. What is differential privacy?

*Your answer here*

2. Briefly describe what is **protected** and **not protected** by applying differential privacy.

*Your answer here*

3. The following mechanism is used to satisfy differential privacy where  $Z$  denotes the noise added.

$$F(x) = f(x) + Z$$

- (a) Provide an equation to show the relationship between  $\epsilon$  (privacy budget) and  $Z$ ? Are they directly or inversely related?

*Your answer here*

- (b) What does it mean to have  $\epsilon = 0$ ?

*Your answer here*

- (c) What would be the result of adding a larger amount of noise?

*Your answer here*