The University of Newcastle
School of Electrical Engineering and Computer Science

COMP3260/COMP6360 Data Security

Week 12 Workshop – 24th and 26th May 2021

Job	Sex	Age	Disease
Engineering	Female	31	Fracture
Scientist	Female	33	Flu
Scientist	Female	35	HIV
Lawyer	Female	32	Flu
Doctor	Female	31	Flu
Cricketer	Male	23	Fracture
Cricketer	Male	25	Fracture
Golfer	Male	20	HIV

Table 1

Name	Job	Sex	Age
Anne	Engineering	Female	31
Betty	Scientist	Female	33
Claire	Scientist	Female	35
Donna	Lawyer	Female	32
Andrew	Doctor	Female	31
Bob	Cricketer	Male	23
Charlie	Cricketer	Male	25
Dennis	Golfer	Male	20
Peter	Doctor	Male	33
David	Lawyer	Male	32
Mark	Engineer	Male	24

Table 2

1. Suppose that a hospital has removed patients' names from the hospital records and intends to make these 'anonymised' patients' records in Table 1 available to a researcher. Suppose that the researcher has access to the external table Table 2 and knows that every person with a record in Table 1 has a record in Table 2. Would this lead to record or attribute linkage of hospital patients? Which patients would have their privacy compromised? With what probability can an adversary infer that Betty has HIV?

2. Generate k-anonymous tables from Table 1 and Table 2. What is the highest k you can achieve for each table?

3. Consider the anonymous data generated in Problem 2. Suppose the adversary knows that the target victim Betty is a scientist of age 30 and has a record in the Table 1. With what probability can an adversary infer that Betty has HIV? Compare this with the case when data was not k-anonymised.

4. Consider the Anonymous data generated in Problem 2. Calculate ℓ for ℓ-diversity for Table 1. (If there are different kinds of ℓ–diversity that we studied then calculate ℓ for all of them.)

5.

a) What is the basic idea behind ε-differential privacy? What problem is it addressing?

b) If we have $P(F(T1)=S)=0.5$ and $P(F(T2)=S)=0.4$, for $\varepsilon=1$ then is the ε-differential privacy model satisfied for that particular query?

c) If we have $P(F(T1)=S)=0.8$ and $P(F(T2)=S)=0.4$, for $\varepsilon=1$ then is the ε-differential privacy model satisfied for that particular query?