The University of Newcastle
School of Electrical Engineering and Computer Science

COMP3260/COMP6360 Data Security
Week 12 Workshop – 30 & 31 May 2019

Privacy

1. For each of the following attack models, describe the Attack Model, and name a Privacy Model that addresses that kind of attack.
   a. Record Linkage
   b. Attribute Linkage
   c. Table Linkage

2. One set of techniques for privacy involves restricting access to the dataset: query set size control, query set overlap control, maximum order control, partitioning, cell suppression and auditing. In this context, what is partitioning?

3. In information theory, what is entropy, and how is it calculated? What is equivocation (conditional entropy), and how is it calculated?

4. Consider the following dataset:
   a. Categorise the attributes into Identifier, Quasi-identifier, Non-sensitive attribute and Sensitive attribute
   b. What level of k-anonymity is achieved by the original table? (What is the smallest equivalence class?)
   c. Create your own taxonomy and generalize the data values so that 4-anonymity is achieved.
   d. Find ℓ so that the anonymized data set achieves ℓ-diversity.

Degree	Sex	Name	Age	Average grade
Civil Engineering	Female	Anne	20	HD
Electrical Engineering	Female	Betty	23	D
Mechanical Engineering	Female	Claire	25	D
Software Engineering	Female	Donna	22	HD
Mathematics	Male	Andrew	21	C
Chemistry	Male	Bob	23	HD
Biology	Male	Charlie	25	HD
Physics	Male	Dennis	20	D

5.
   a. What is the basic idea behind ε-differential privacy? What problem is it addressing?
   b. If we have $P(F(T_1) = S) = 0.5$ and $P(F(T_2) = S) = 0.4$, for $\varepsilon = 1$ then is the ε-differential privacy model satisfied for that particular query?
   c. If we have $P(F(T_1) = S) = 0.8$ and $P(F(T_2) = S) = 0.4$, for $\varepsilon = 1$ then is the ε-differential privacy model satisfied for that particular query?