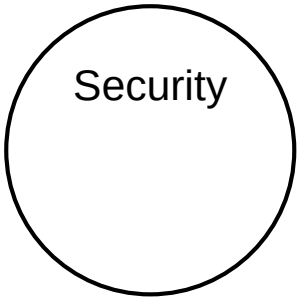
Privacy

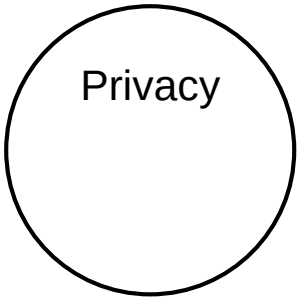
Prof. Ravi Sandhu
Executive Director and Endowed Chair

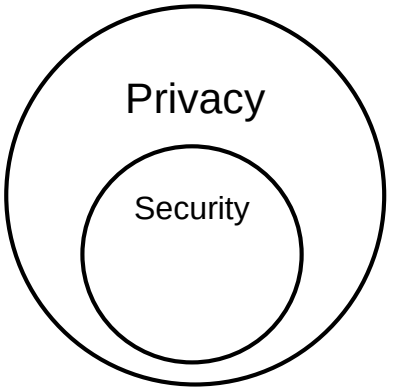
Lecture 17

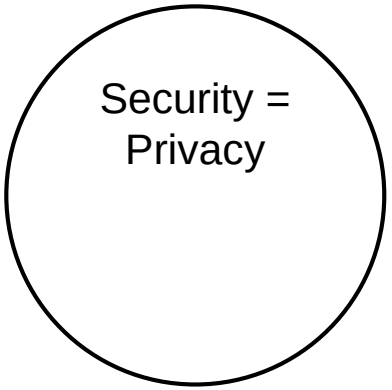
ravi.utsa@gmail.com
www.profsandhu.com

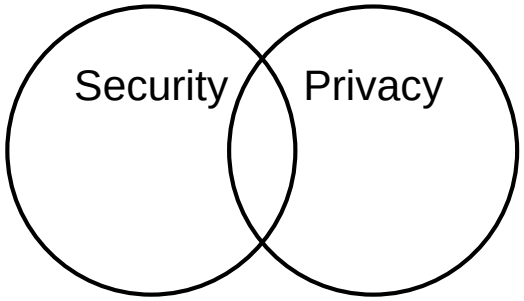
Security versus Privacy

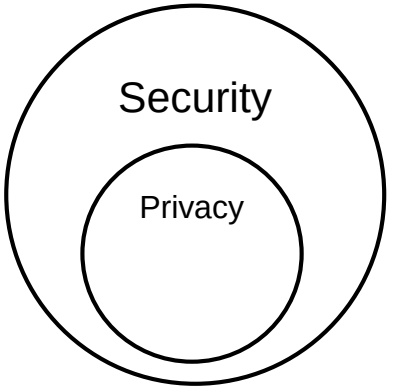




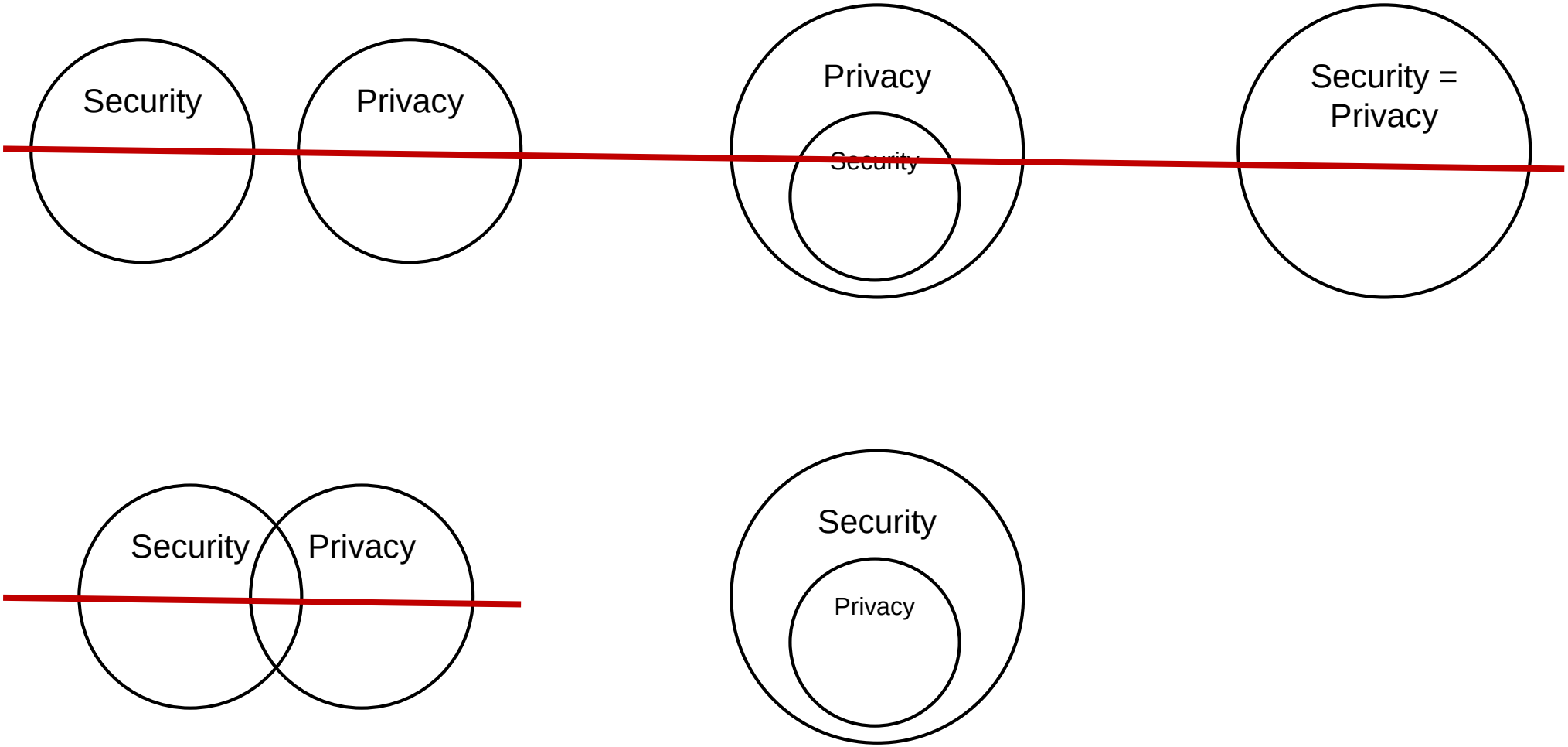




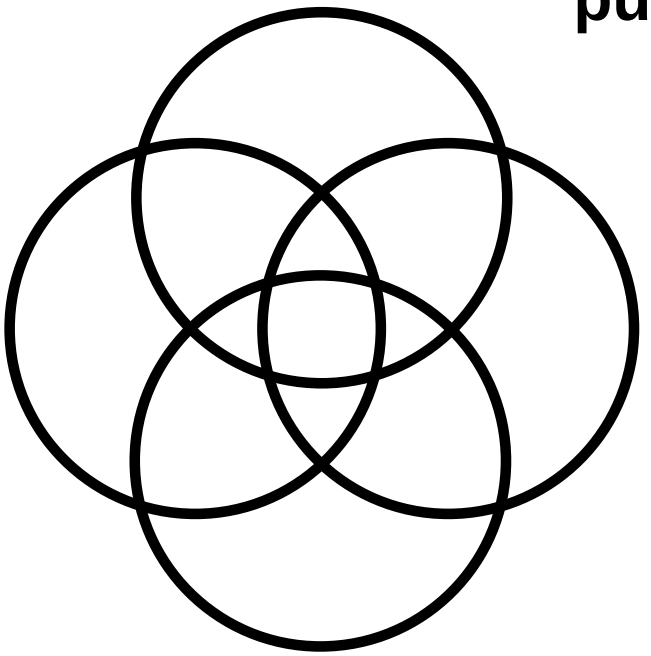




Security versus Privacy



Security and Privacy Objectives



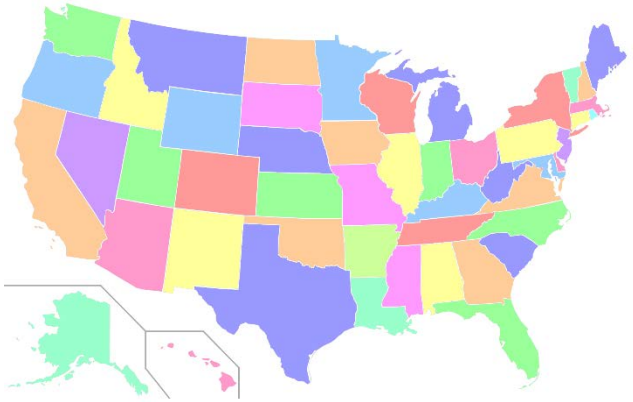
Privacy Goals

- Browsing privacy
- Interaction privacy
- Drive-by privacy
- Data disclosure privacy
- ……

Privacy Legislation in USA

- Case by case approach
  - Fragmented
  - Reactive
- Fair Credit Reporting Act (FCRA), 1970
- Privacy Act 1974
- Family Educational Rights & Privacy Act (FERPA), 1974
- Video Privacy Protection Act (VPAA) 1988
- Health Insurance Portability and Accountability Act (HIPAA) 1996
- Gramm-Leach-Bliley Act (GLBA) 1999
- ……

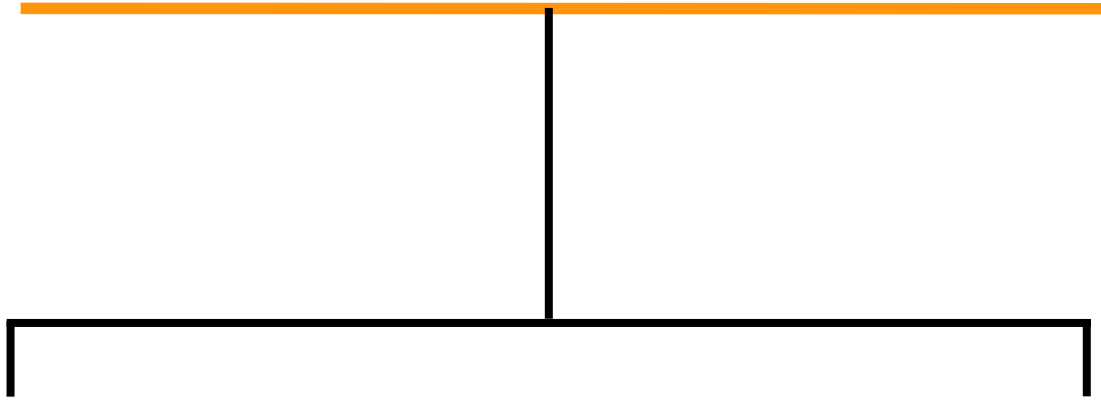
3 Ongoing Experiments



Data Disclosure
Syntactic Anonymity
Differential Privacy

Clifton, Chris, and Tamir Tassa. "On Syntactic Anonymity and Differential Privacy." Transactions on Data Privacy 6, no. 2 (2013): 161-183.

Data Disclosure



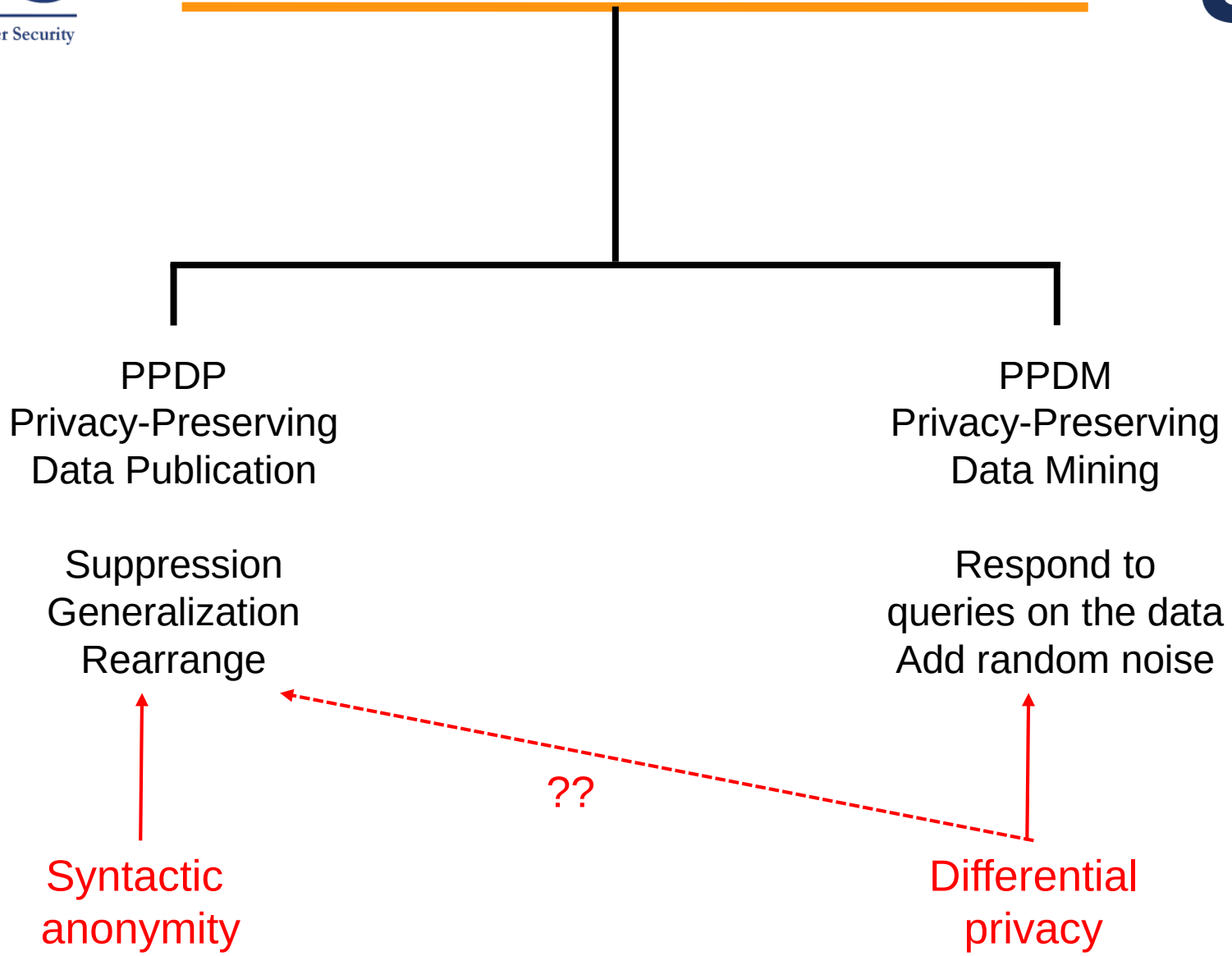
PPDP
Privacy-Preserving
Data Publication

Suppression
Generalization
Rearrange

PPDM
Privacy-Preserving
Data Mining

Respond to
queries on the data
Add random noise

Data Disclosure



Data Disclosure
Syntactic Anonymity

k-Anonymization

quasi-Identifiers QIs

sensitive attribute

generalization

age	zipcode	disease
28	10145	measles
21	10141	hepatitis
21	12238	hepatitis
55	12256	flu
53	12142	angina
48	12204	angina

age	zipcode	disease
[21 – 28]	1****	measles
[21 – 28]	1****	hepatitis
[21 – 28]	1****	hepatitis
[48 – 55]	12***	flu
[48 – 55]	12***	angina
[48 – 55]	12***	angina

QI block

QI block

Table 1: (a) A table (left); (b) a corresponding 3-anonymization (right).

l-Diversity

age	zipcode	disease
[21 – 53]	1****	measles
[21 – 53]	1****	hepatitis
[21 – 55]	122**	hepatitis
[21 – 55]	122**	flu
[21 – 53]	1****	angina
[21 – 55]	122**	angina

Table 2: A 3-anonymization of Table 1(a) that respects 3-diversity.

QI block

Frequency of occurrence of sensitive attribute ≤ 1/3

t-Closeness

age	zipcode	disease
[21 – 53]	1****	measles
[21 – 53]	1****	hepatitis
[21 – 55]	122**	hepatitis
[21 – 55]	122**	flu
[21 – 53]	1****	angina
[21 – 55]	122**	angina

Table 2: A 3-anonymization of Table 1(a) that respects 3-diversity.

QI block

Frequency of occurrence of sensitive attribute ≤ 1/3

Distribution in each QI block should be t-close to distribution in entire population

stricter than l-diversity

p-Sensitivity

age	zipcode	disease
[21 – 53]	1****	measles
[21 – 53]	1****	hepatitis
[21 – 55]	122**	hepatitis
[21 – 55]	122**	flu
[21 – 53]	1****	angina
[21 – 55]	122**	angina

Table 2: A 3-anonymization of Table 1(a) that respects 3-diversity.

QI block

at least p distinct values in each QI block

weaker than l-diversity

Anatomy Anonymization

randomly shuffle
Sensitive attribute
in each QI block

age	zipcode	disease
28	10145	hepatitis
21	10141	angina
53	12142	measles
21	12238	flu
55	12256	hepatitis
48	12204	angina

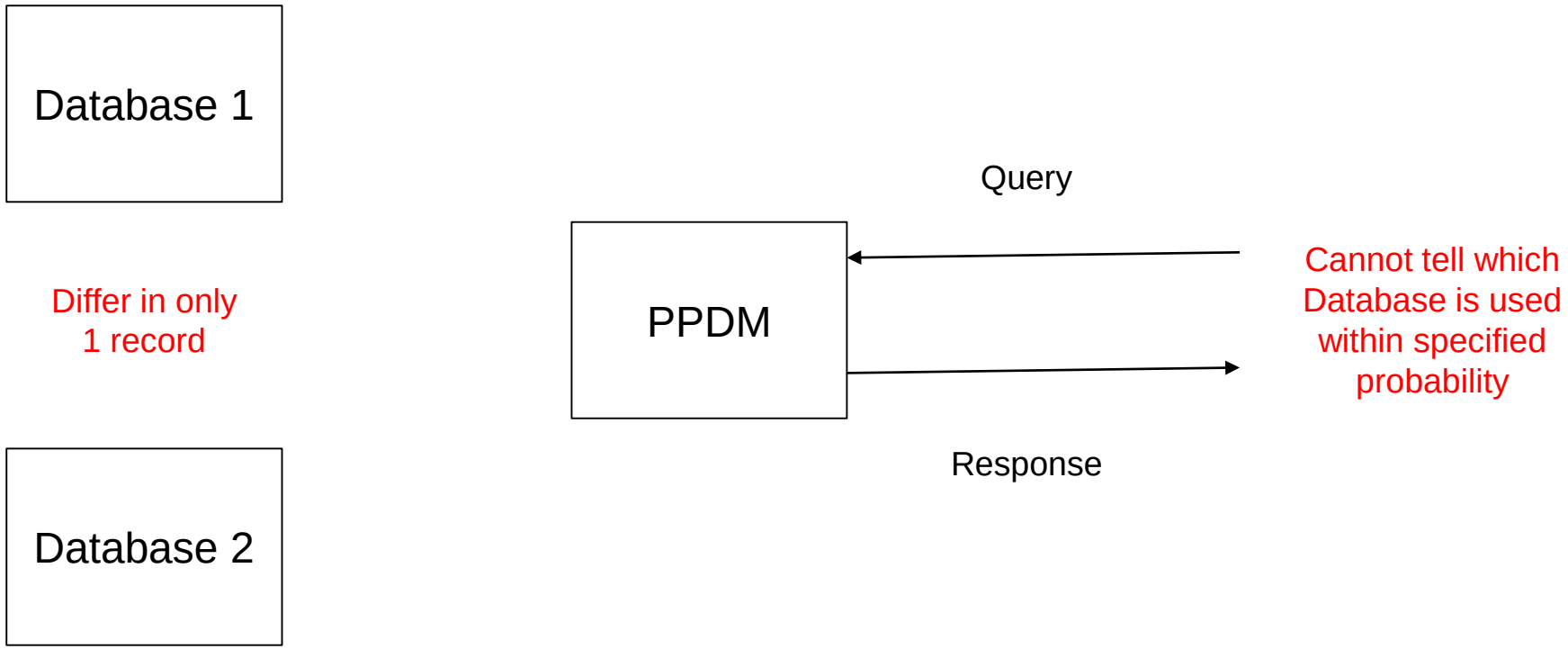
Table 3: An Anatomy anonymization of Table 1(a) that uses the same partitioning into QI-blocks as the anonymization in Table 2.

Curse of Dimensionality

- Given a large number of Quasi-Identifiers, need to suppress most of the table to achieve k-anonymity
- LKC-privacy for high dimensional data:
  - Attacker knows at most L quasi-identifiers
  - Every combination of L quasi-identifiers is shared by at least K records
  - Diversity in each group of K is not more than 1/C
- Release different versions of the data with different Quasi-Identifiers for different research purposes

Data Disclosure
Differential Privacy

Intuition


Database 1
Differ in only 1 record
Database 2
Query
PPDM
Response
Cannot tell which Database is used within specified probability