COMP7370 Advanced Computer and Network Security

Generalizing Data to Provide Anonymity when Disclosing Information (4)

Topics:

1. Time management
2. Domain generalization hierarchy

Topic 1: Time management

- Emails
  - Google: “How to Read 100 Emails, Fast”
  - Check email once a day
  - Group emails
  - Reply to all the short emails - first with "yes" or "no" as an answer
  - Write brief emails
  - Long emails -> tasks -> must be prioritized

Review: **K-anonymity**

Definition 3.1 (k-minimal generalization – wrt a quasi-identifier) *Let T_i and T_j be two tables such that $T_i \leq T_j$. T_j is said to be a k-minimal generalization of a table T_i* wrt to a quasi-identifier QI *iff:*

1. T_j satisfies k-anonymity wrt QI

2. $\forall T_z : T_i \leq T_z, T_z \leq T_j$, T_z satisfies k-anonymity wrt $QI \Rightarrow T_z[QI] = T_j[QI]$.

(1) k-anonymity

(2) Minimal **Question:** (why minimal matters?)

Topic 2: Domain generalization hierarchy

Why we need a domain generalization hierarchy?

We have many generalization solutions, for example:

Eth:E_0	ZIP:Z_0
a	38
a	39
a	41
a	42
b	38
b	39
b	41
b	42
c	38
c	39
c	41
c	42

PT

Eth:E_1	ZIP:Z_0
p	38
p	39
p	41
p	42
p	38
p	39
p	41
p	42
p	38
p	39
p	41
p	42

$GT_{[1,0]}$

Eth:E_1	ZIP:Z_1
p	30
p	30
p	40
p	40
p	30
p	30
p	40
p	40
p	30
p	30
p	40
p	40

$GT_{[1,1]}$

Eth:E_0	ZIP:Z_2
a	00
a	00
a	00
a	00
b	00
b	00
b	00
b	00
c	00
c	00
c	00
c	00

$GT_{[0,2]}$

Eth:E_0	ZIP:Z_1
a	30
a	30
a	40
a	40
b	30
b	30
b	40
b	40
c	30
c	30
c	40
c	40

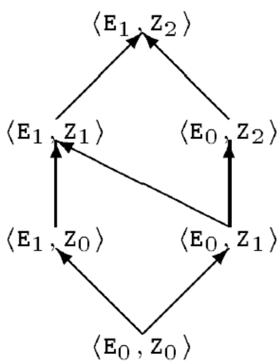
$GT_{[0,1]}$

Figure 3: Examples of generalized tables for PT

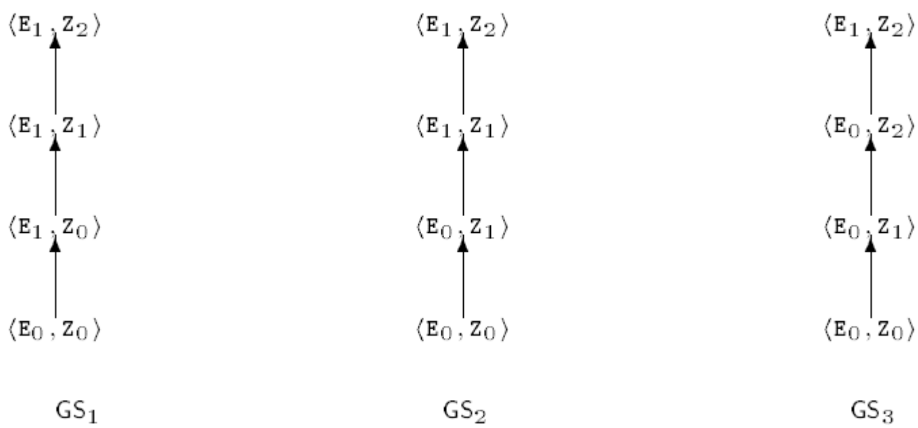
- Motivations of using domain generalization hierarchy:
  - Can we represent relations between these generation solutions?
  - The definition of k-minimal generalization depends on k value. Can we find an approach that is independent of k value?
  - How can we show generalization strategies or different ways of generalizing a DB?

Note: some generalizations are minimal some are not wrt quasi-identifier.

- Walk through this example:



low usability, high privacy



high usability, low privacy

- The number of different possible strategies for a domain hierarchy

Theorem 3.2 *Let $DT = \langle D_1, \ldots, D_n \rangle$ be a tuple such that $D_i \in \mathsf{Dom}, i = 1, \ldots, n$. The number of different strategies for DT is: $\frac{h_{DT}!}{h_1! \ldots h_n!}$, where each h_i is the length of the path from D_i to the top domain in DGH_{D_i} and $h_{DT} = \sum_{i=1}^{n} h_i$.* □

DT = (D1, …, Dn)
e.g., DT = (E0, Z0, E1, Z1, ..)