PRIVACY PROTECTION IN INFORMATION SYSTEMS

Manolis Terrovitis
Athena Research Center
mter@athenarc.gr

Security definition - CIA

- CIA triad
  - Confidentiality
    - Protection from un-authorized access
  - Integrity
    - Protection from deletion and modification
  - Availability
    - System works properly

Privacy vs Security

Security refers to protection against the unauthorized access of data. We put security controls in place to limit who can access the information.

Privacy is harder to define, but it is about protecting individuals from revelation of personal information.

it is possible to have security without privacy, but impossible to have privacy without security.

Example

Your privacy and security are maintained. The bank uses your information to open your account and provide you with products and services. They go on to protect that data.

Your privacy is compromised, and your security is maintained. The bank sells some of your information to a marketer. Your personal information is in more hands than you may have wanted.

Both your privacy and security are compromised. The bank gets hit by a data breach. Cybercriminals penetrate a bank database, a security breach. Your information is exposed and could be sold on the dark web. Your privacy is gone. You could become the victim of cyber fraud and identity theft.

Basic frameworks for privacy protection

- GDPR
- Privacy by design
  - 7 principles
- ISO/IEC 29100:2011/2017 Information technology — Security techniques — Privacy framework
  - 11 principles

ISO/IEC 29100

International Organization for Standardization/International Electrotechnical Commission

- specifies a common privacy terminology;
- defines the actors and their roles in processing personally identifiable information (PII);
- describes privacy safeguarding considerations; and
- provides references to known privacy principles for information technology.
  - ISO/IEC 29100:2011 is applicable to natural persons and organizations involved in specifying, procuring, architecting, designing, developing, testing, maintaining, administering, and operating information and communication technology systems or services where privacy controls are required for the processing of PII.

ISO/IEC 29100:2024

- 1. Consent and choice
- 2. Purpose legitimacy and specification
- 3. Collection limitation
- 4. Data minimization
- 5. Use, retention and disclosure limitation
- 6. Accuracy and quality
- 7. Openness, transparency and notice
- 8. Individual participation and access
- 9. Accountability
- 10. Information security
- 11. Privacy compliance

Other privacy standards

- ISO 27701: 2019 – Requirements for Personal Information Management System
  - Security focused
- ISO 31700:2023 – Privacy by Design
- ISO 29134:2023: Guidelines for Privacy Impact Assessments
- ISO 27559:2022 – Privacy Enhancing Data De-Identification Framework
- ISO 27555: Guidelines on PII Deletion

Techniques for privacy protection

- Access control
- Logging
- Encryption
- Pseudonymization
- Anonymization

Access control

- Access control
  - Identification – Document the identity of the user
  - Authentication – Verify that the account is used by the correct user, e.g. password, fingerprints
  - Authorization – Document what the user can do, e.g. read, write, alter etc
- Access control models
  - MAC models – Mandatory Access control, no transfer of rights
    - Access decisions are based on security labels assigned to subjects (users) and objects (resources)
  - DAC models – Discretional Access control, transferable rights, e.g. UNIX file system
    - .Access decisions are based on the discretion of the resource owner, who can grant or revoke access to specific users or groups
  - RBAC model – Role based Access control
    - Permissions are assigned to roles rather than individual users, simplifying access management and reducing administrative overhead

Logging

- Logging of user and system actions
  - Goal based
  - Requires well defined access policies
  - Detailed authorization framework
  - Log files must be protected
  - Administrator actions should be logged in detail
  - System clocks should be synchronized
  - Log files are personal data!!

Encryption

- Encryption is a very old method
- Plaintext is transformed to a ciphertext that only those that know the encryption secret can decrypt
- Encryption is used mainly to ensure confidentiality
  - Encryption protocols can provide additional privacy enhancing mechanisms

Homomorphic encryption

Partial

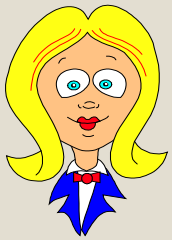
- $E(m_1)+E(m_2)=E(m_1+m_2)$ Additive
- $E(m_1)*E(m_2)=E(m_1*m_2)$ Multiplicative

Fully

- Any operation of the domain

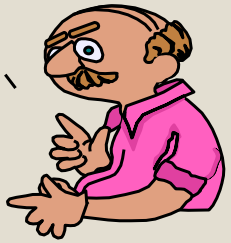
Secure multiparty Computation

Alice



a

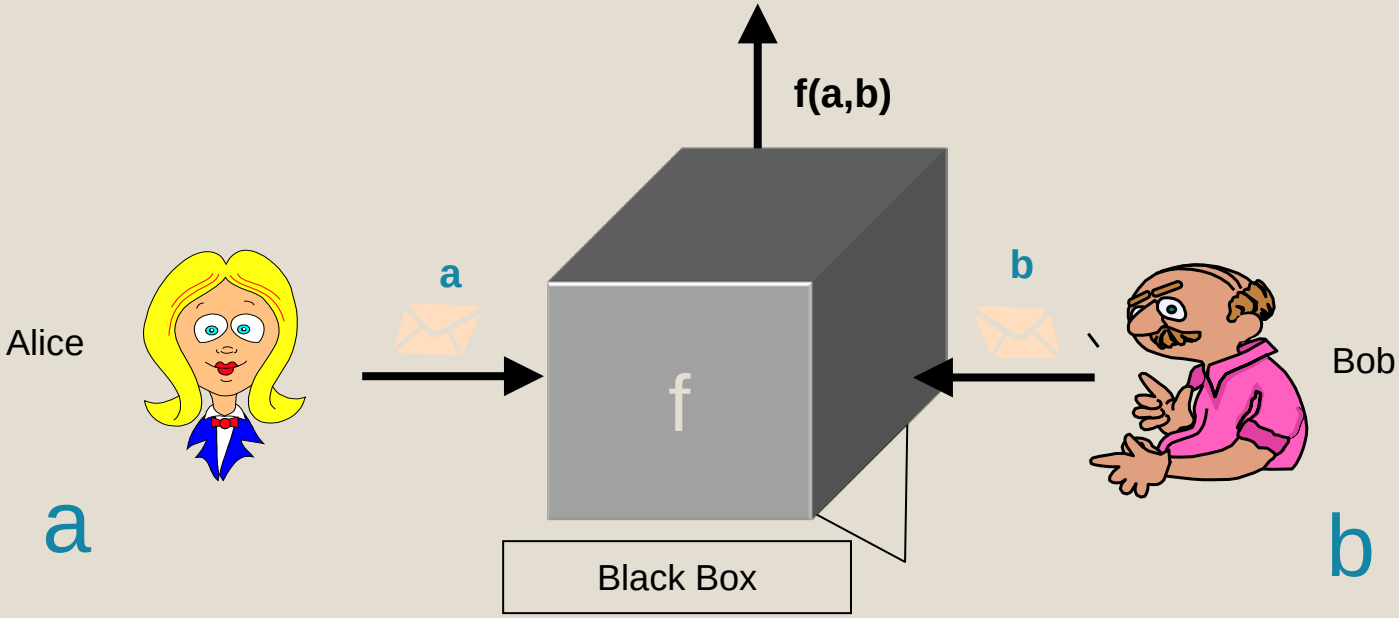
f(a,b)



Bob

b

Secure multiparty Computation



Anonymization and pseudonymization

- Closely related but very different in terms of GDPR
- Very often misused in practice
- Anonymization is claimed instead of pseudonymization

General Data Protection Regulation

- **Definitions:**
- ‘pseudonymisation’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person;

Anonymous personal data – Recital 26

- Personal data which have undergone **pseudonymisation**, which could be attributed to a natural person by the use of additional information should be considered to be **information on an identifiable natural person**. [3]
- To determine whether a natural person is identifiable, account should be taken of all the **means reasonably likely to be used**, such as **singling out**, either by the controller or by another person to identify the natural person directly or indirectly.
- [4]To ascertain whether means are reasonably likely to be used to identify the natural person, **account should be taken of all objective factors**, such as the **costs** of and the amount of **time** required for identification, taking into consideration the available **technology** at the time of the processing and technological developments. [5]
- **The principles of data protection should therefore not apply to anonymous information**, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. [6]This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

Pseudonymization vs Anonymization

PSEUDONYMIZATION

- Direct identifiers are replaced by arbitrary ones, which should be kept separately.
- “the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information.”
- (WG) Pseudonymisation is also addressed to clarify some pitfalls and misconceptions: pseudonymisation is not a method of anonymisation. It merely reduces the linkability of a dataset with the original identity of a data subject, and is accordingly a useful security measure.

ANONYMIZATION

- Irreversible
- No longer personal data

Not so simple

Article 29 Data protection working party (0829/14/EN WP216)

- Several anonymization techniques may be envisaged, there is no prescriptive standard in EU legislation.
- Importance should be attached to contextual elements: account must be taken of “all” the means “likely reasonably” to be used for identification by the controller and third parties, paying special attention to what has lately become, in the current state of technology, “likely reasonably” (given the increase in computational power and tools available).
- **A risk factor is inherent to anonymisation**: this risk factor is to be considered in assessing the validity of any anonymisation technique - including the possible uses of any data that is “anonymised” by way of such technique - and severity and likelihood of this risk should be assessed.

Problems

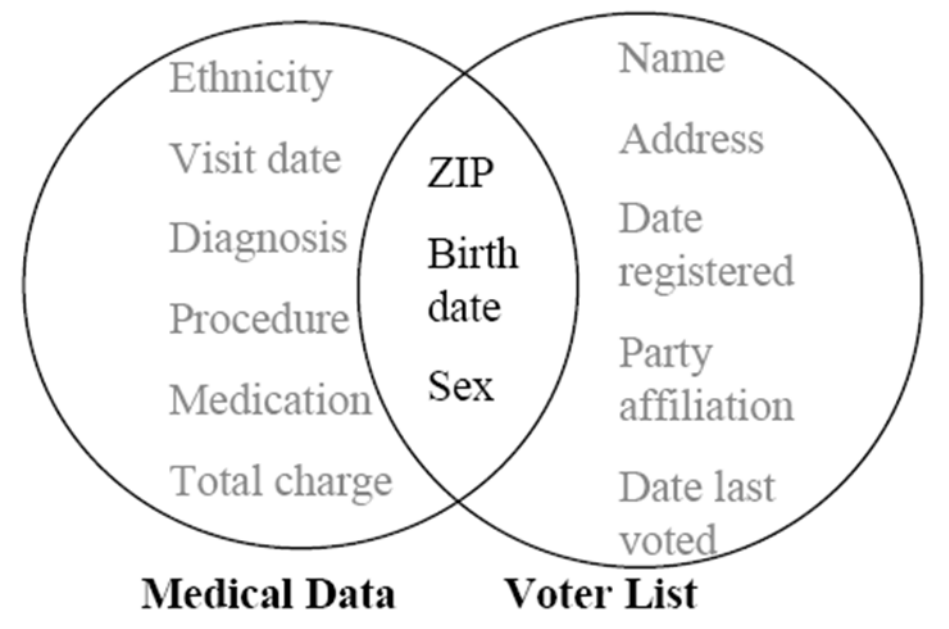
An effective anonymisation solution prevents all parties from singling out an individual in a dataset, from linking two records within a dataset (or between two separate datasets) and from INFERRING any information in such dataset

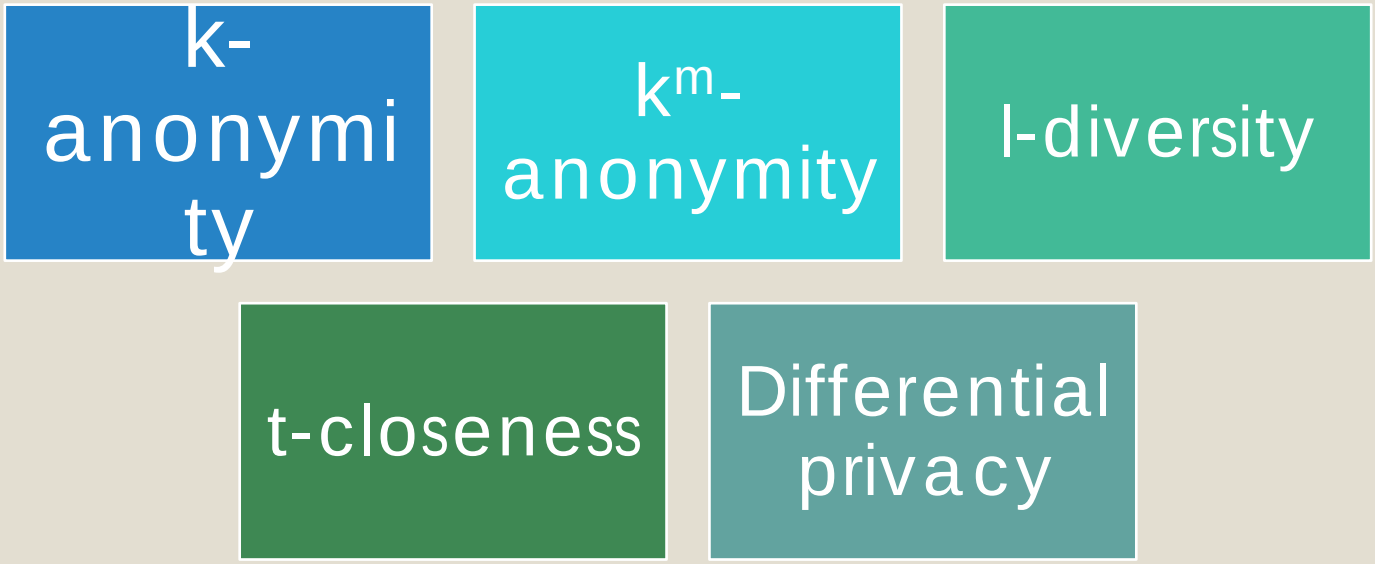
Inferring?

publication scenario

- A research group or an organization wants to share micro data
  - To support research, to allow the verification of results
- to protect privacy microdata are sanitized
  - explicit identifiers (SSN, name, phone #) are removed
- is this sufficient for preserving privacy?
- no! susceptible to link attacks
  - publicly available databases (voter lists, city directories) can reveal the “hidden” identity

Pseudonymization - Link attacks

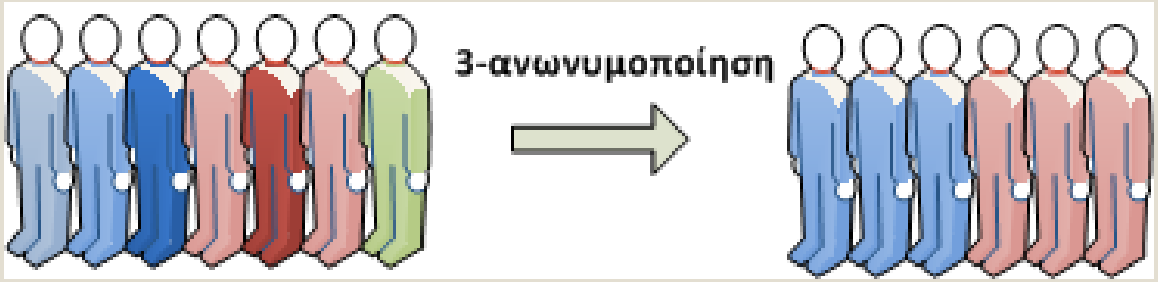




Different flavors

K-anonymity

- Suggested in 2001
- Intuitively each person is hidden in a group of similar person
- Straightforward interpretation of GDPR



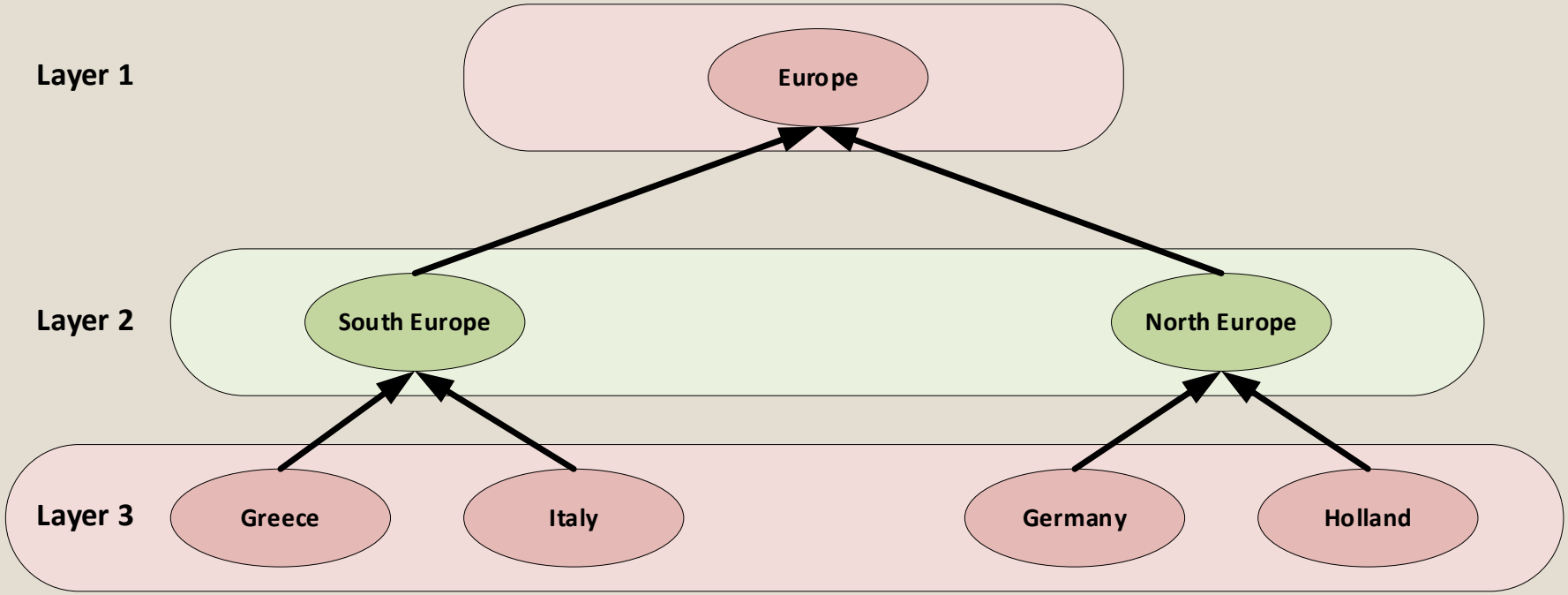
k-anonymity

- Each entry becomes indistinguishable from other *k*-1 entries
  - *k*-anonymity is achieved through suppression and generalization

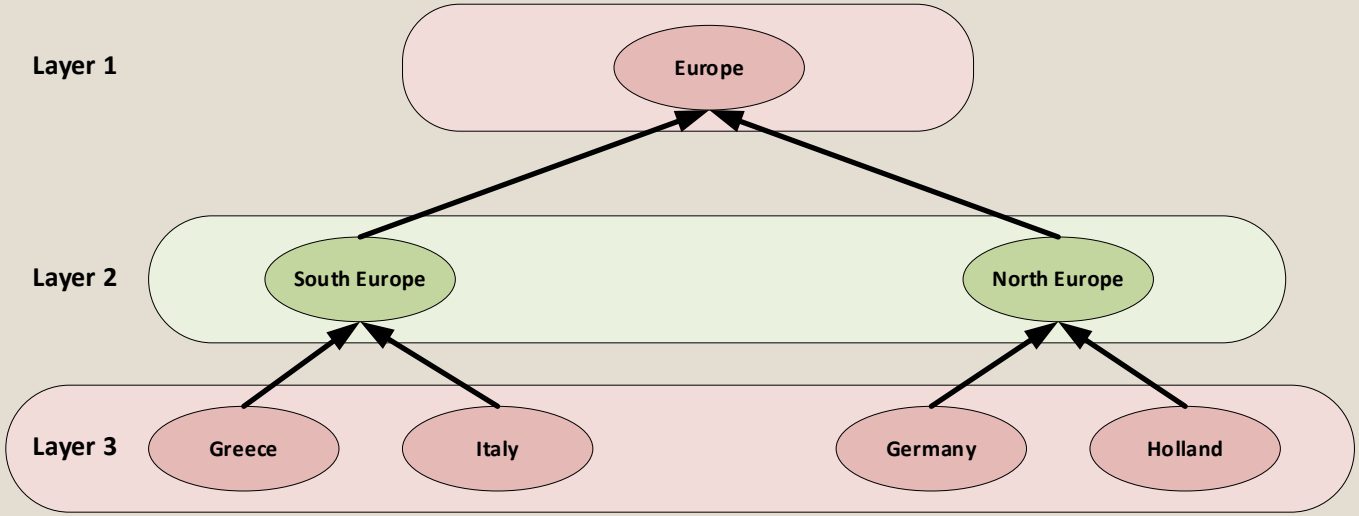
id	Zipcode	Age	National.	Disease
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

~~id~~	Zipcode	Age	National.	Disease
~~1~~	130**	<30	*	Heart Disease
~~2~~	130**	<30	*	Heart Disease
~~3~~	130**	<30	*	Viral Infection
~~4~~	130**	<30	*	Viral Infection
~~5~~	1485*	≥40	*	Cancer
~~6~~	1485*	≥40	*	Heart Disease
~~7~~	1485*	≥40	*	Viral Infection
~~8~~	1485*	≥40	*	Viral Infection
~~9~~	130**	3*	*	Cancer
~~10~~	130**	3*	*	Cancer
~~11~~	130**	3*	*	Cancer
~~12~~	130**	3*	*	Cancer

Data transformation – Full domain generalization



Full domain



Age	Place
28	Greece
33	Italy
21	Greece
24	Greece
37	Germany
36	Holland
35	Germany

Age	Place
[25-35]	South Europe
[25-35]	South Europe
[20-25]	South Europe
[20-25]	South Europe
[35-40]	North Europe
[35-40]	North Europe
[35-40]	North Europe

Local Recoding

- Original values are not globally generalized but only some of subsets of them

Age	Place
28	Greece
33	Italy
21	Greece
24	Greece
37	Germany
36	Holland
35	Germany

Age	Place
[25-35]	South Europe
[25-35]	South Europe
[20-25]	Greece
[20-25]	Greece
[35-40]	North Europe
[35-40]	North Europe
[35-40]	North Europe

Generalization graph in full domain generalization

Z2 ={537**}

↑

Z1 ={5371*, 5370*}

↑

Z0 ={53715, 53710, 53706, 53703}

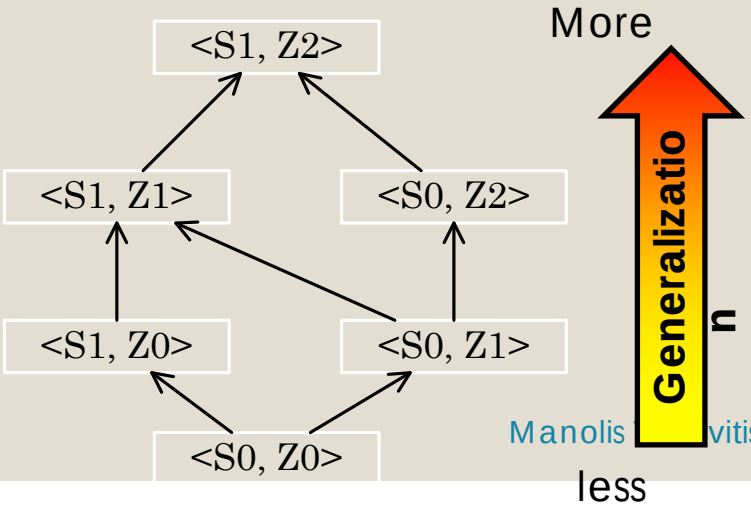
Zip Code

S1 ={person}

↑

S0 ={male,female,non-binary, other}

Gender

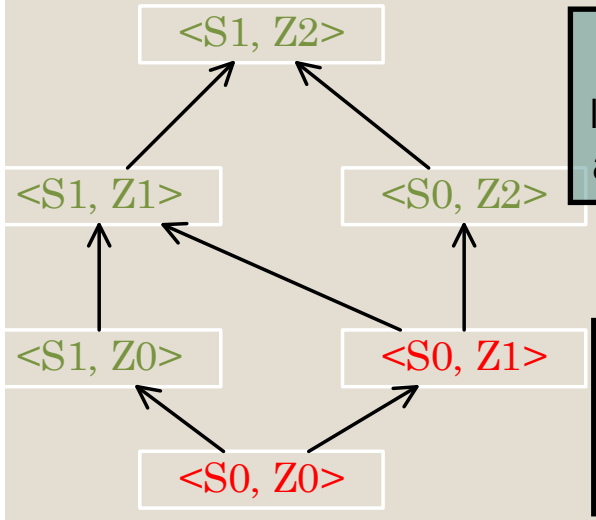


Target

Find the least generalization that satisfies k-anonymity

Generalization graph

Monotonicity



(I) Generalization property(~rollup)

If a node is k-anonymous, every descendant is also k-anonymous

(II) Subset property (~apriori)

If k-anonymity does not hold for a set of quasi identifiers, it will not hold for any superset of them

Incognito is one of the first algorithms that work on this graph

More complex situations

- Data are rarely just a table of independent tuples

Structural information

- We need to anonymize all relevant information about a person, not just a tuple
- Information tends to gather over time
- Information is linked through semantic properties, the schema is irrelevant
- Personal data tend to accumulate over time
- There are several solution for simple data and complicated attack models, but few solutions for complicated data and simple attack models

Limits of k-anonymity

	Fruits	Meat	Vegetables	Fish
Vassilis	X	X		
Manolis	X	X	X	
Eleni			X	
Maria		X	X	
Kostas	X			X

	Food
Vassilis	X
Manolis	X
Eleni	X
Maria	X
Kostas	X

- 2-anonymous

Limits of k-anonymity

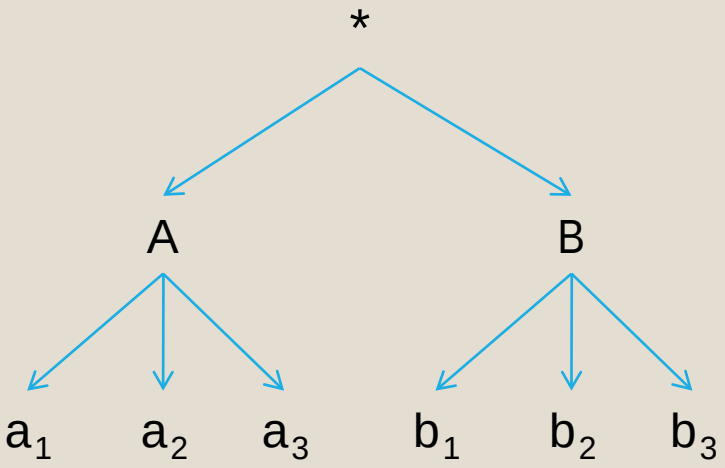
	Fruits	Meat	Vegetables	Fish
Vassilis	X	X		
Manolis	X	X	X	
Eleni			X	
Maria		X	X	
Kostas	X			X

	Fruits	Meat	Other food
Vassilis	X	X	
Manolis	X	X	X
Eleni			X
Maria		X	X
Kostas	X		X

- 2^2-anonymous
- Any combination of *m* items will not appear less than *k* times

Generalization of set values

Generalization Hierarchy



$$\{a_1, a_2\} \rightarrow \{A\}$$

id	contents
t_1	$\{a_1, b_1, b_2\}$
t_2	$\{a_2, b_1\}$
t_3	$\{a_2, b_1, b_2\}$
t_4	$\{a_1, a_2, b_2\}$

(a) original database (D)

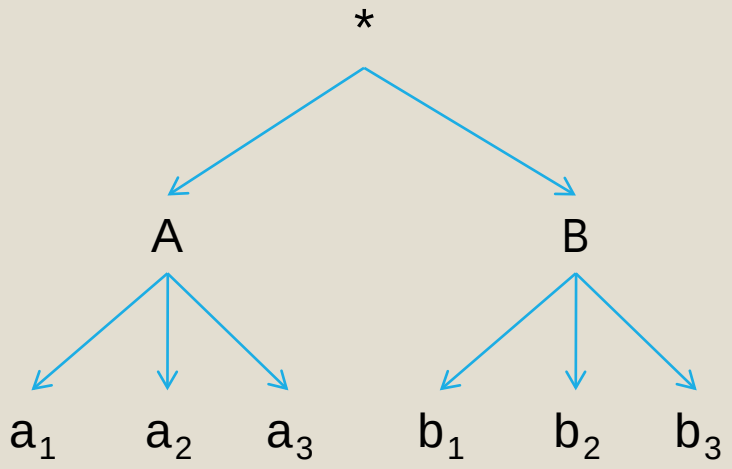
k=2
m=2

id	contents
t'_1	$\{A, b_1, b_2\}$
t'_2	$\{A, b_1\}$
t'_3	$\{A, b_1, b_2\}$
t'_4	$\{A, b_2\}$

(b) transformed database (D')

Method Description

- Data Generalization Hierarchy

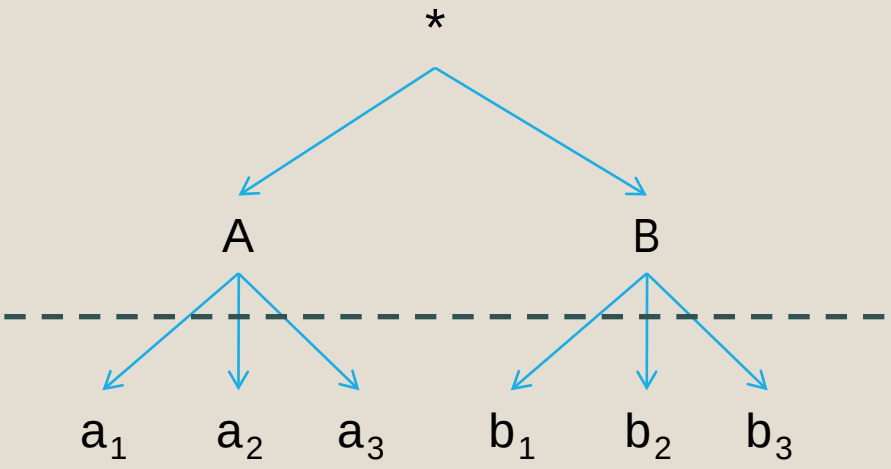


$c_0 = \{*\}$

Top-Down Specialization

Method Description

- Data Generalization Hierarchy

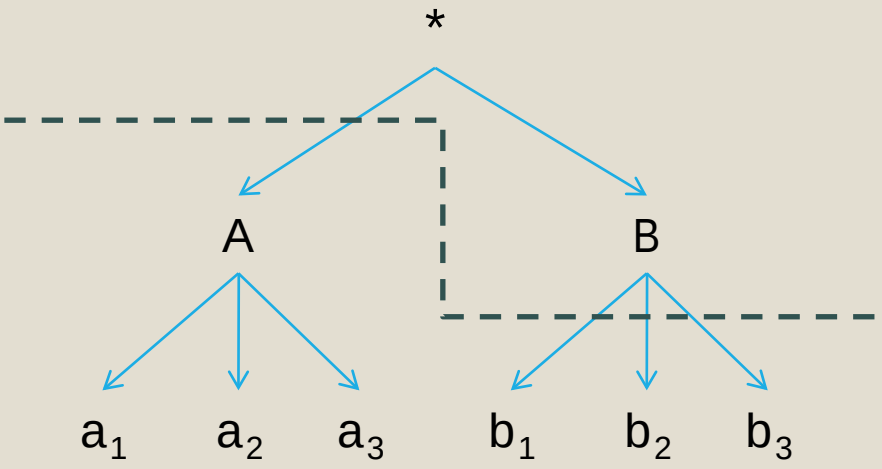


$c_4 = \{a_1, a_2, a_3, b_1, b_2, b_3\}$

no generalization

Method Description

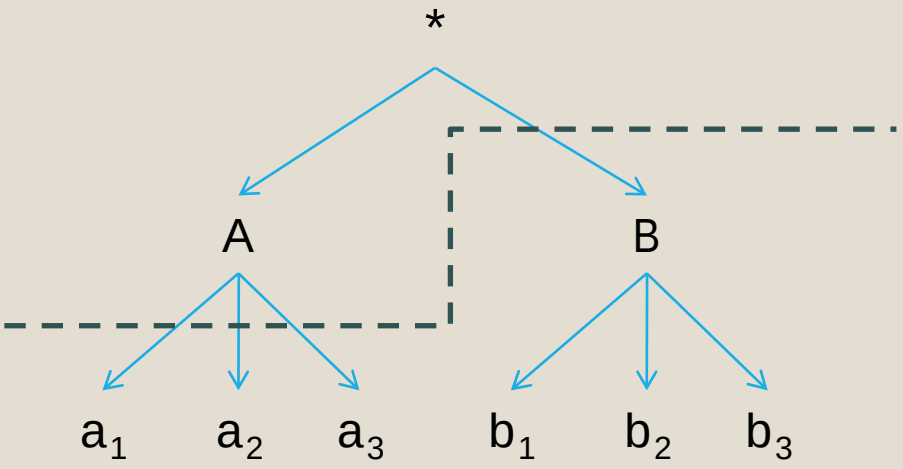
- Data Generalization Hierarchy



$c_3 = \{A, b_1, b_2, b_3\}$

Method Description

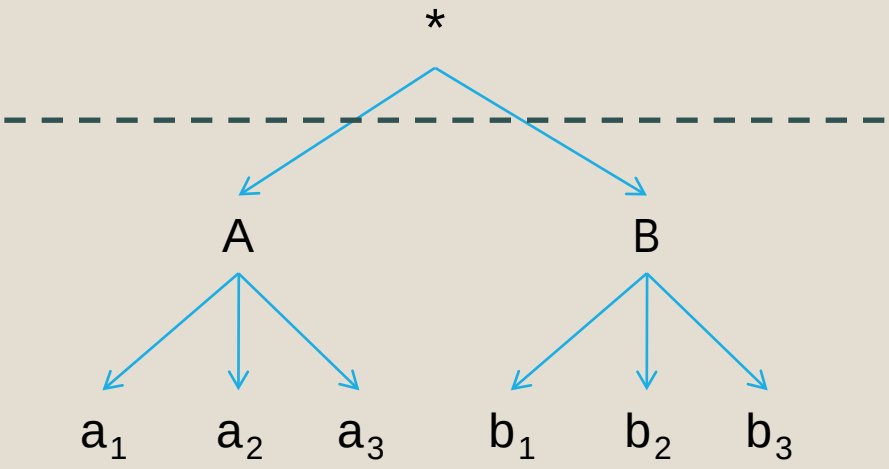
- Data Generalization Hierarchy



$c_2 = \{a_1, a_2, a_3, B\}$

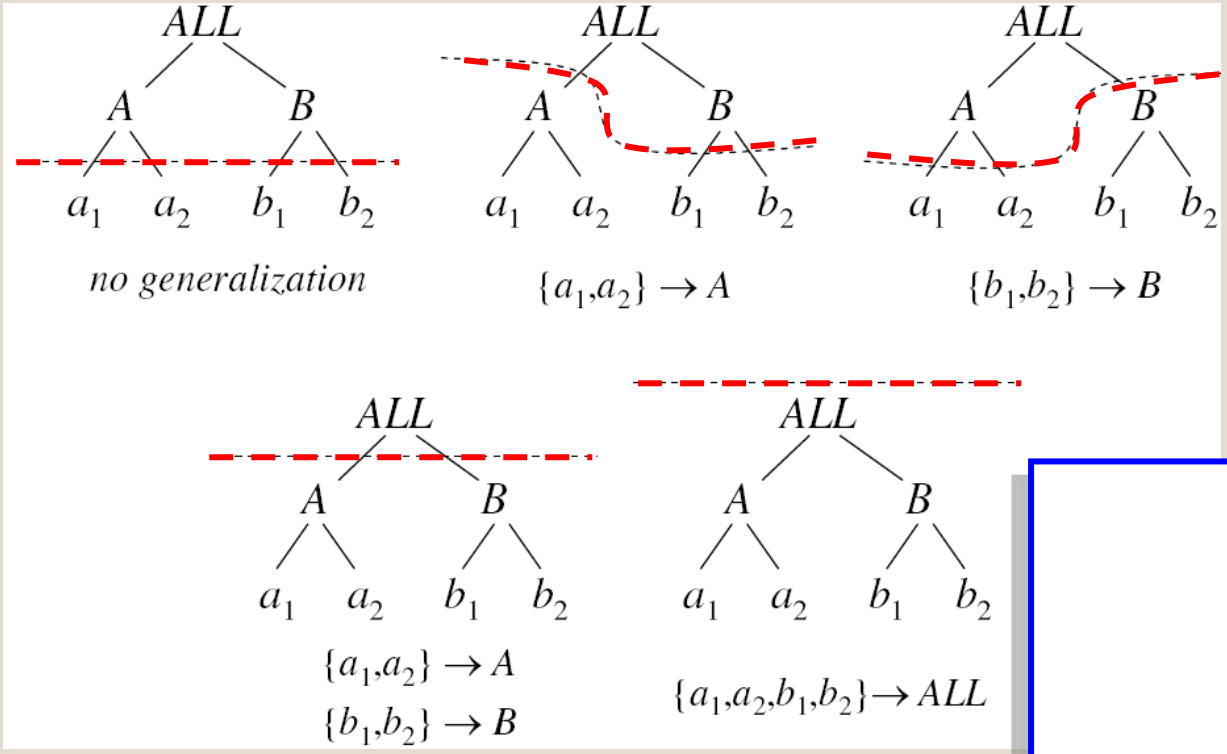
Method Description

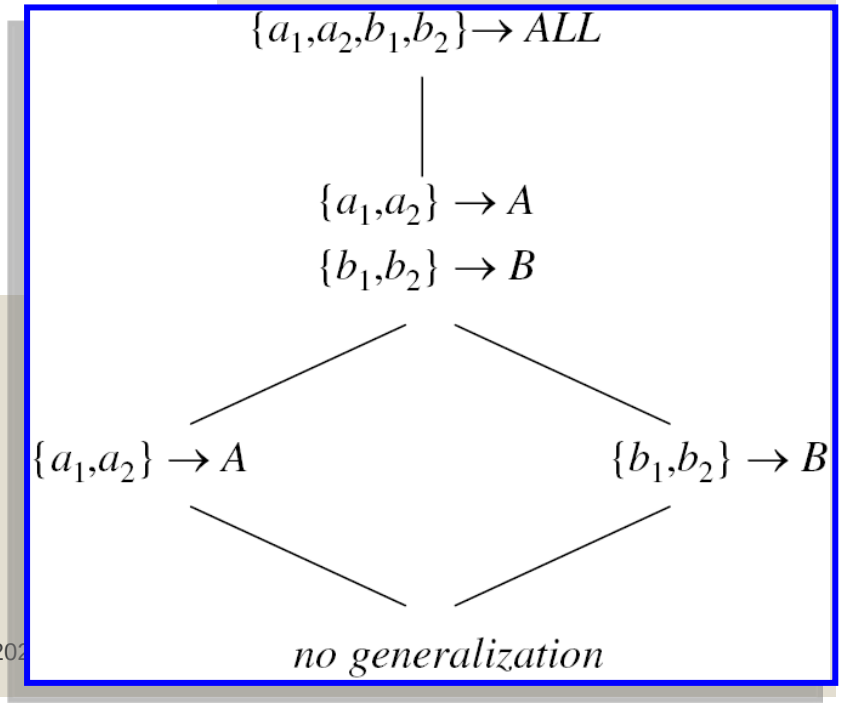
- Data Generalization Hierarchy



$c_1 = \{A,B\}$

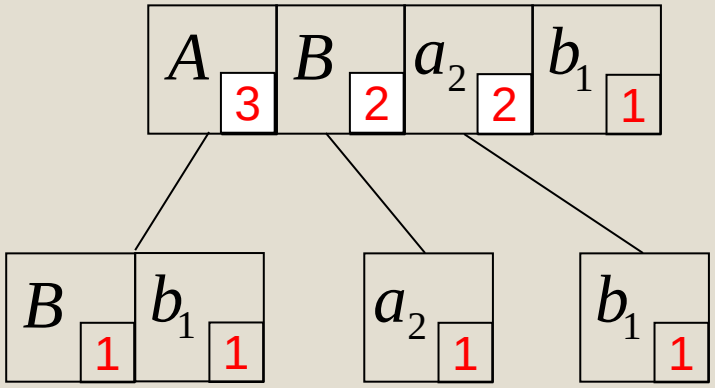
Lattice of Generalizations





Count Tree

$$t_2 = \{a_2, b_1\} \rightarrow \{a_2, b_1, A, B\} \rightarrow \left\{ \begin{array}{c} A, B, a_2, b_1, \\ AB, Ab_1, Ba_2, a_2b_1 \end{array} \right\}$$



"Apriori-based" Anonymization

Construct the count-tree incrementally

Prune unnecessary branches

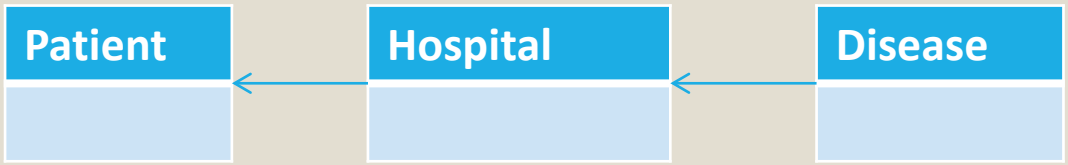
1: initialize c_{out}
2: **for** $i := 1$ to m **do** ▷ for each itemset length
3: initialize a new count-tree
4: **for all** $t \in D$ **do** ▷ scan D
5: extend t according to c_{out}
6: add all i-subsets of extended t to count-tree
7: run DA on count-tree for $m = i$ and update c_{out}

Even more complex stuff

- In relational databases data span in multiple tables
- Privacy guarantees should be provided at the level of a complete personal record, i.e., all data related to a specific person
- Additional structural information must be taken into account
- *Tree Structured data*
  - $k^{(m,n)}$-anonymity

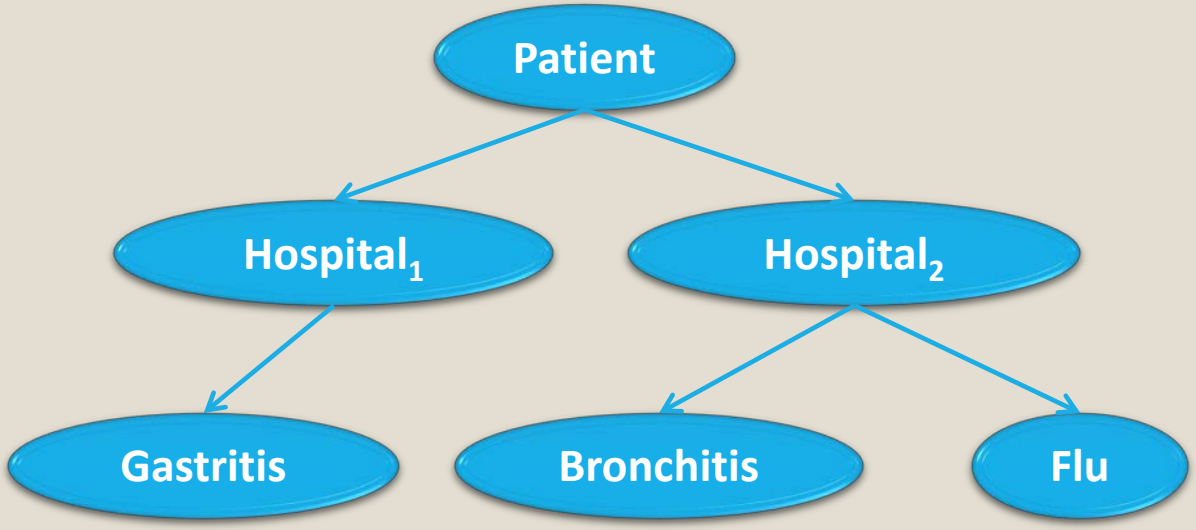
Example

- record = unordered, non-recursive, attribute tree
- XML / DB with various tables



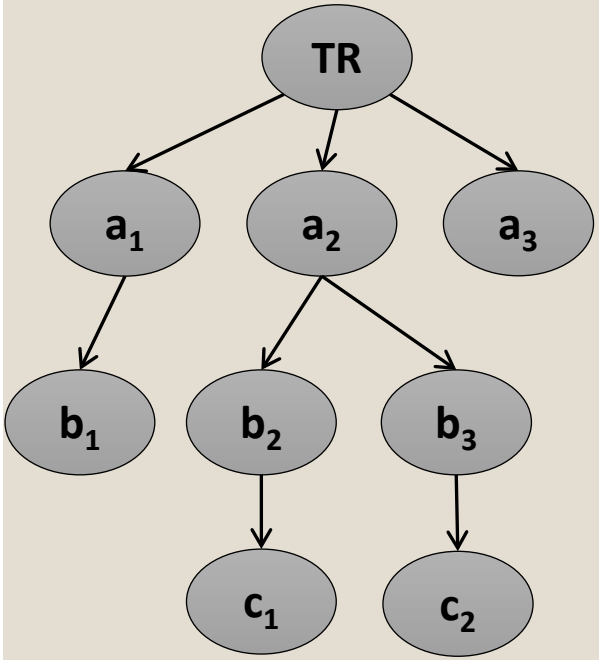
Tree-Structured Data

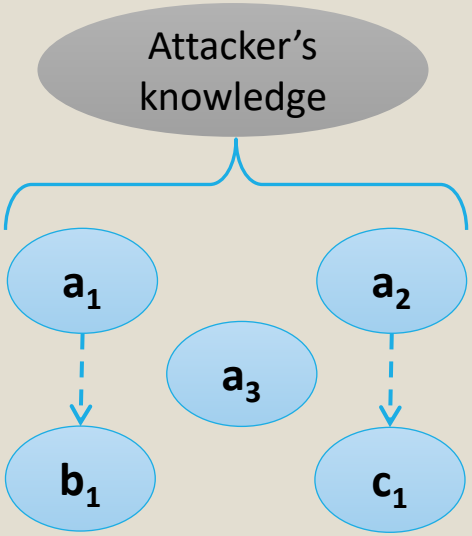
- record = unordered, non-recursive, attribute tree
- XML / DB with multiple tables

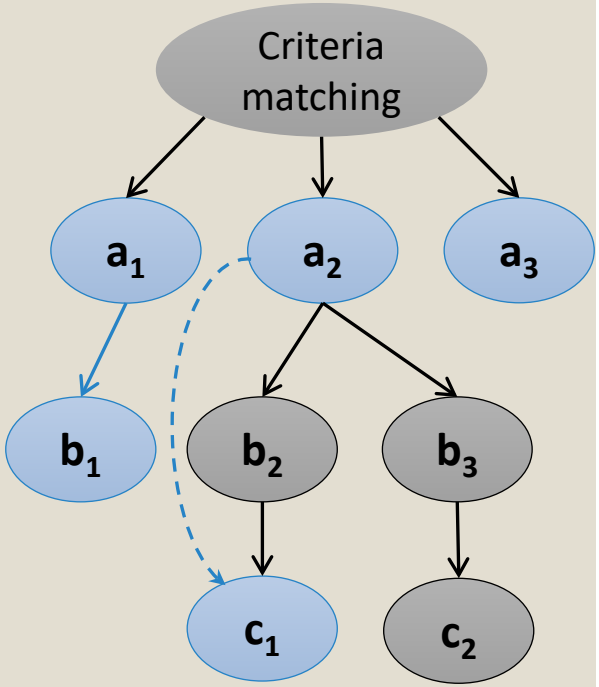


Attack Model

- ***m*** Node labels (values)
- ***n*** Structural relations
  - Ancestor-descendant

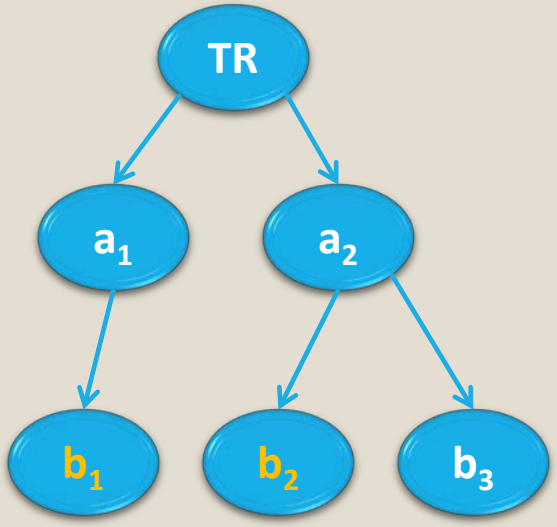


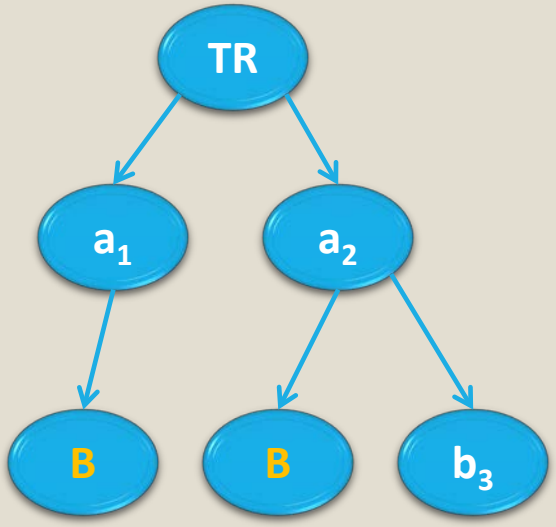




Basic Operations

- Value Generalization
  - b_1 --> B
  - b_2 --> B

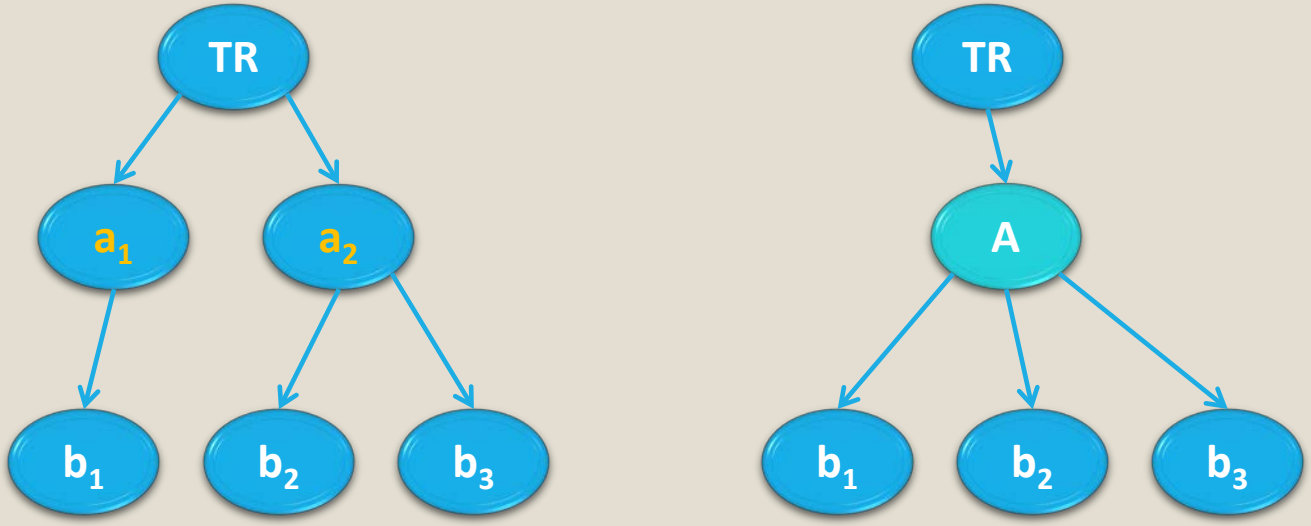




Generalization Rules:
$\{b_1 \rightarrow B\}$, $\{b_2 \rightarrow B\}$

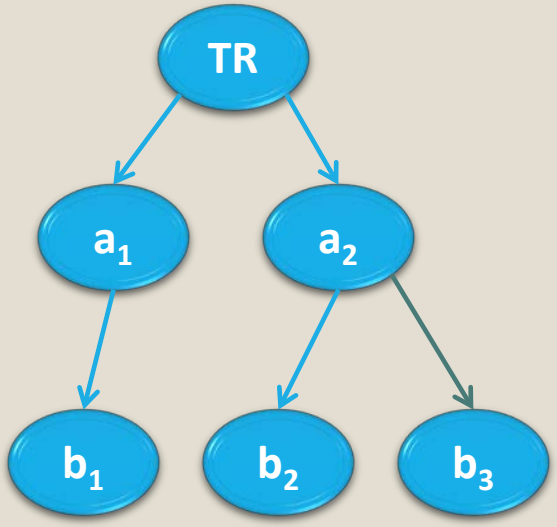
Basic Operations

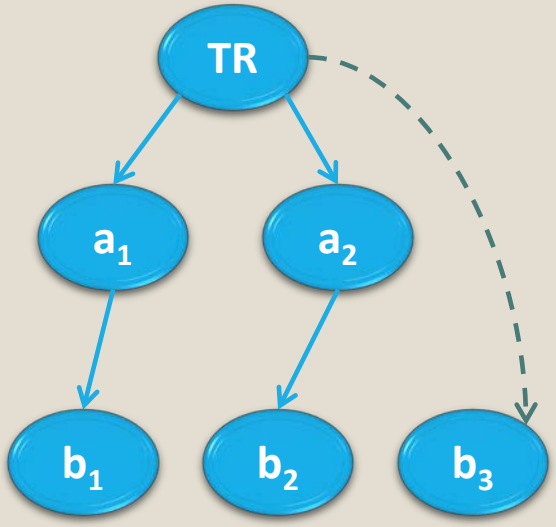
- Value Generalization
  - a_1 --> A
  - a_2 --> A
- Node Merging
  - When siblings are generalized to a common value



Basic Operations

- Value Generalization
- Node Merging
- Structural Disassociation
  - “Hide” relations



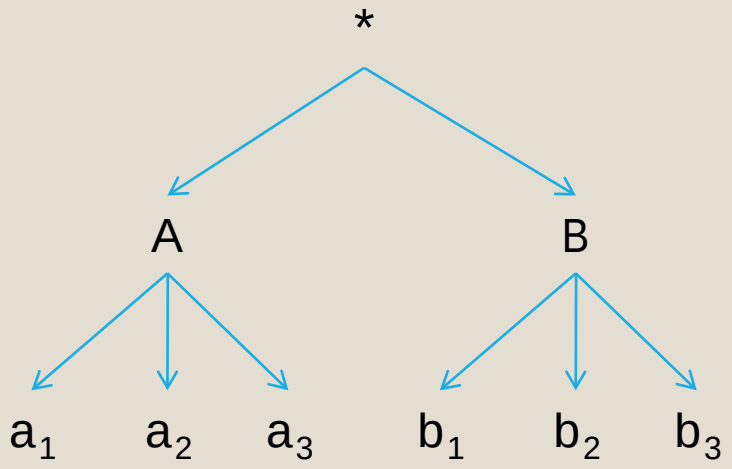


Problem Statement

- Attacker's prior knowledge
  - Up to *m* nodes' values.
  - Up to *n* structural relations of *known* nodes.
    - m: values existence
      - "Target has visited hospital a_1."
    - n: associations between *known* values
      - Ancestor-descendant relations
      - "Target was treated for the disease b_2 at hospital a_1."
- Guarantee
  - At least *k* or 0 records match the attacker's knowledge.

Method Description

- Data Generalization Hierarchy

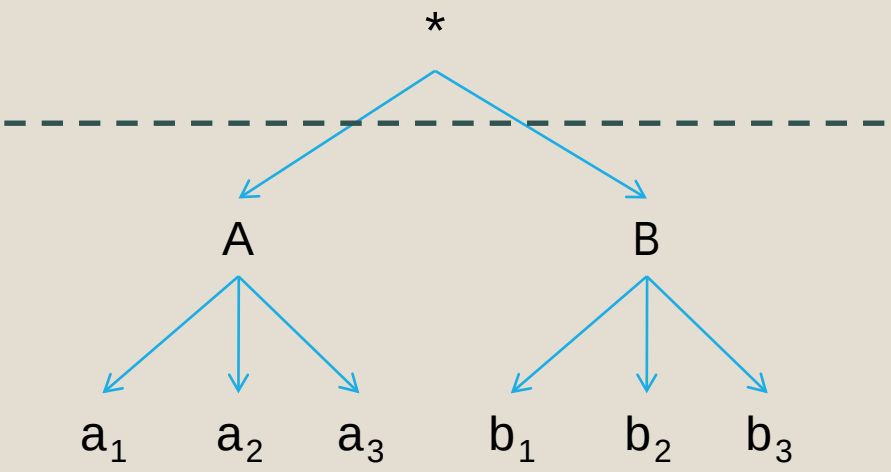


$c_0 = \{*\}$

Top-Down Specialization

Method Description

- Data Generalization Hierarchy

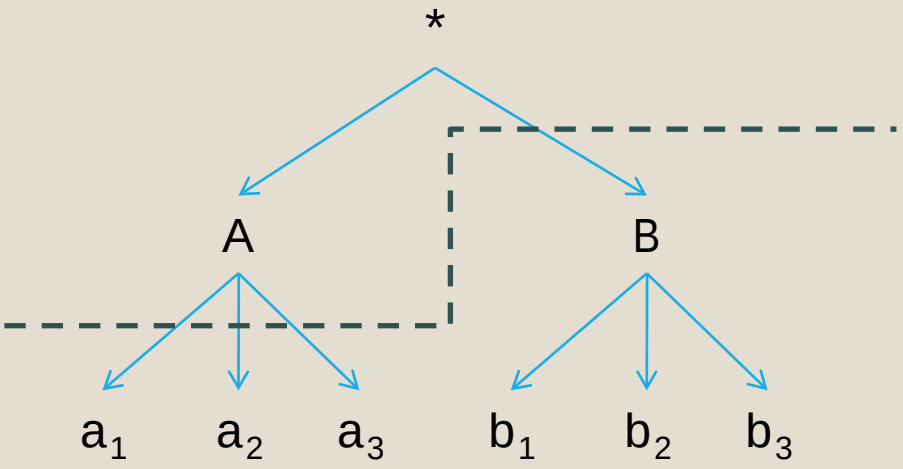


$c_1 = \{A,B\}$

Top-Down Specialization

Method Description

- Data Generalization Hierarchy

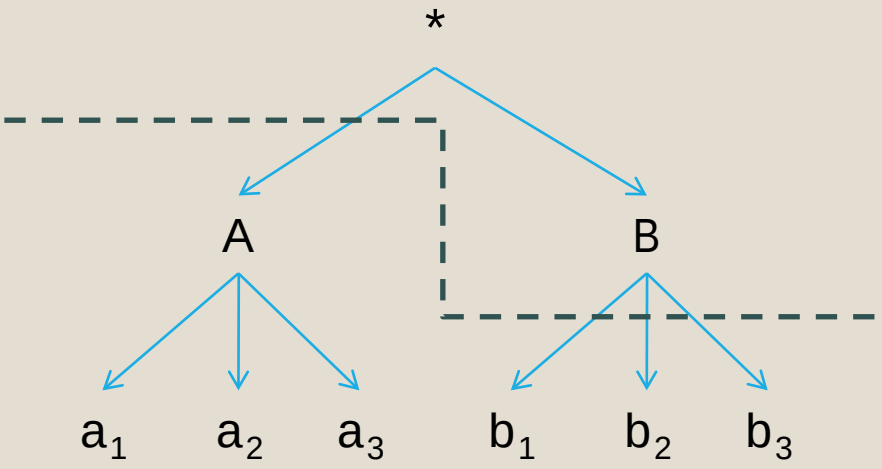


$c_2 = \{a_1, a_2, a_3, B\}$

Top-Down Specialization

Method Description

- Data Generalization Hierarchy

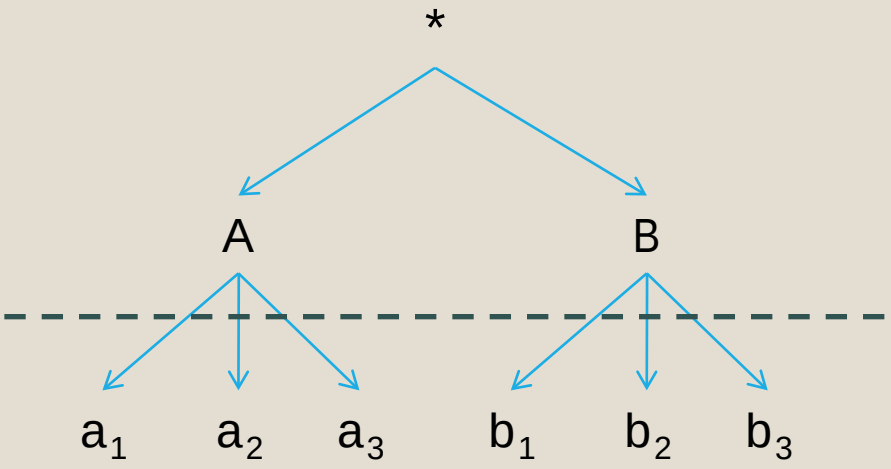


$c_3 = \{A, b_1, b_2, b_3\}$

Top-Down Specialization

Method Description

- Data Generalization Hierarchy



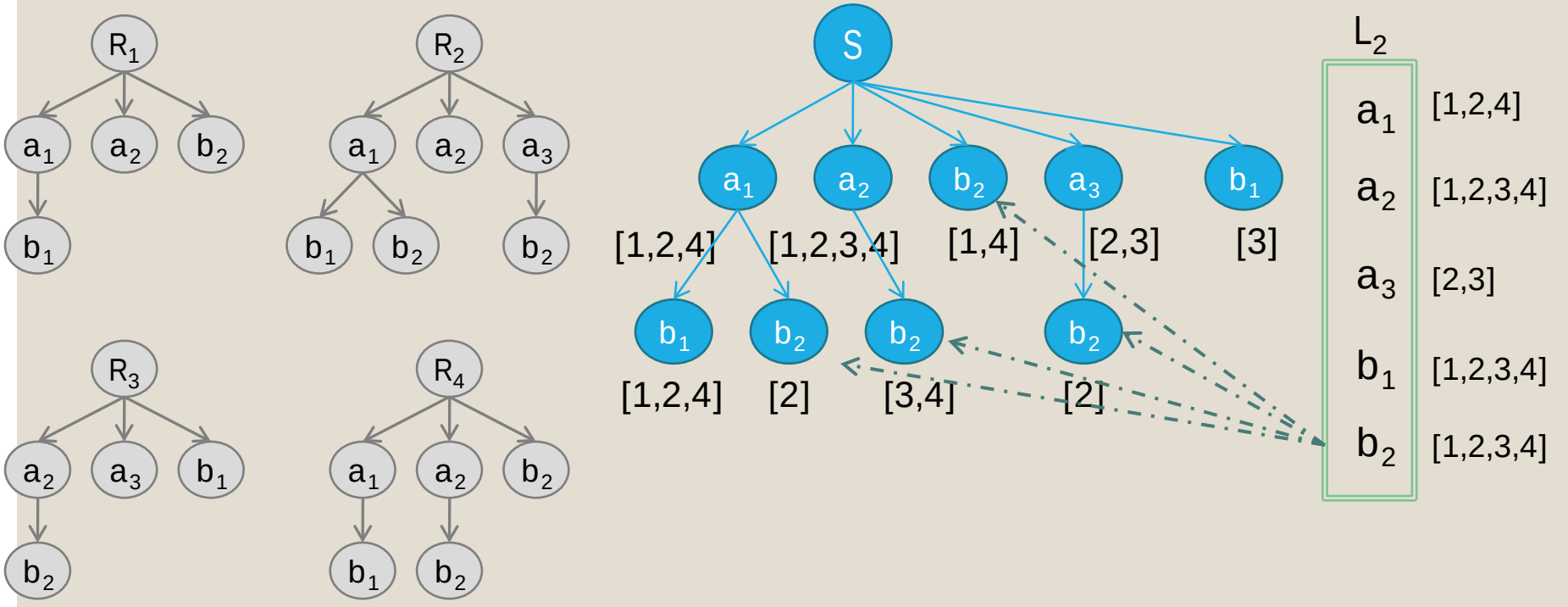
$c_4 = \{a_1, a_2, a_3, b_1, b_2, b_3\}$

no generalization

Top-Down Specialization

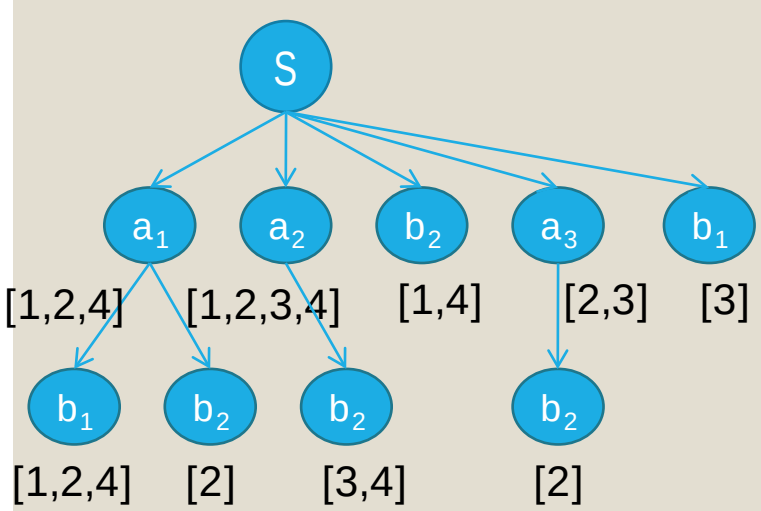
Method Description

- Synopsis tree

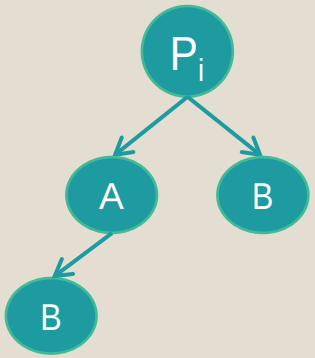


Method Description

- Algorithm



cut c_i



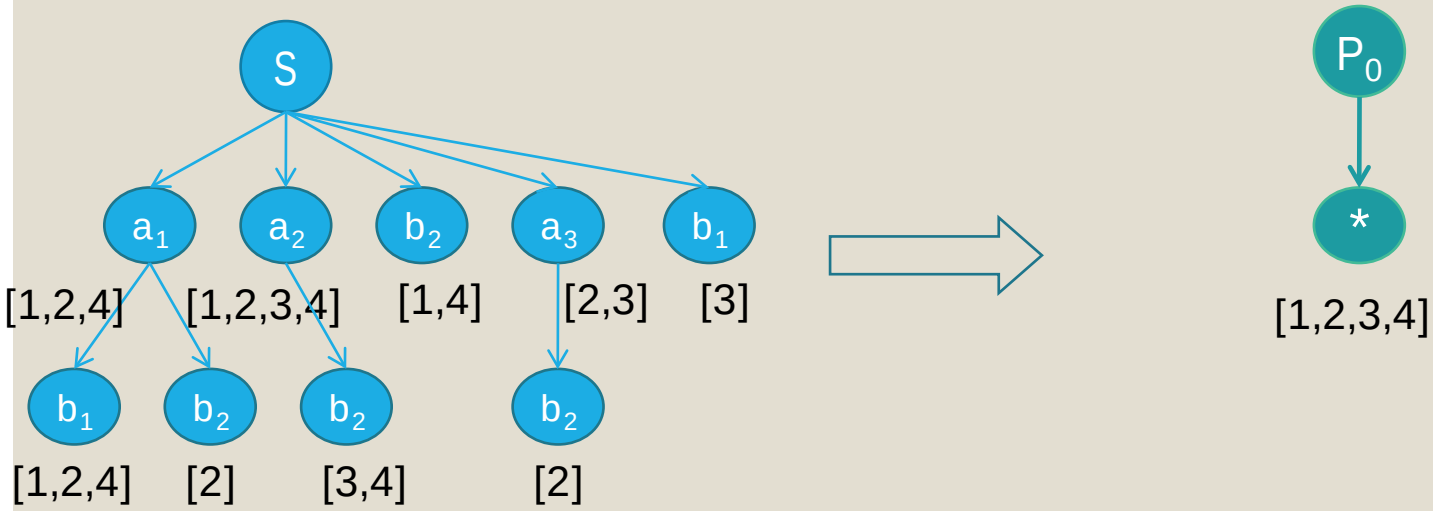
Privacy check,
Cost estimation

InfoLoss(c_i)

Next cut
$i=i+1$

Method Description

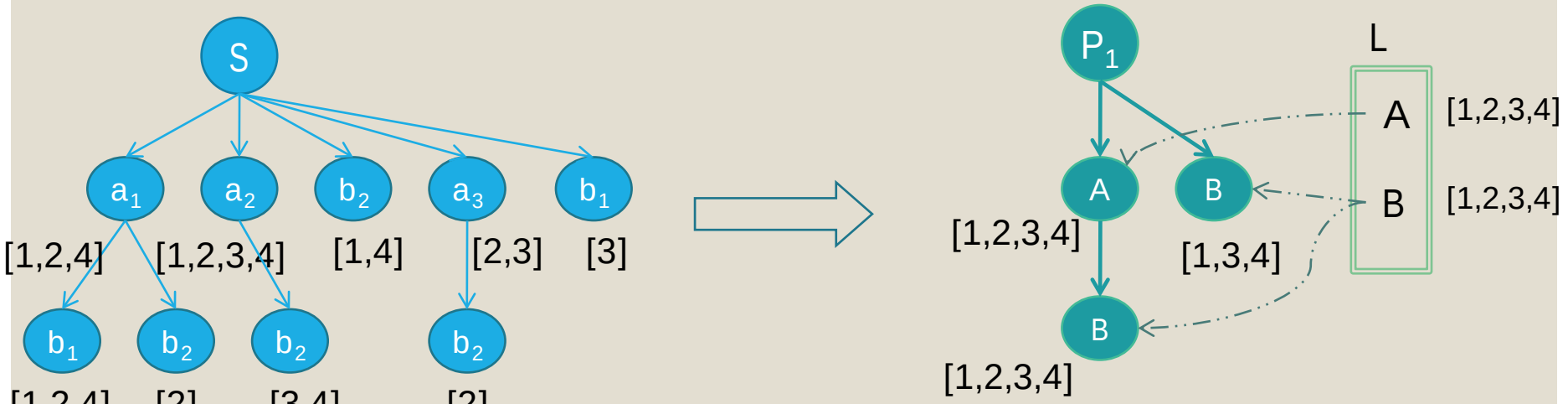
- Projection Tree



$c_0 = \{*\}$

Method Description

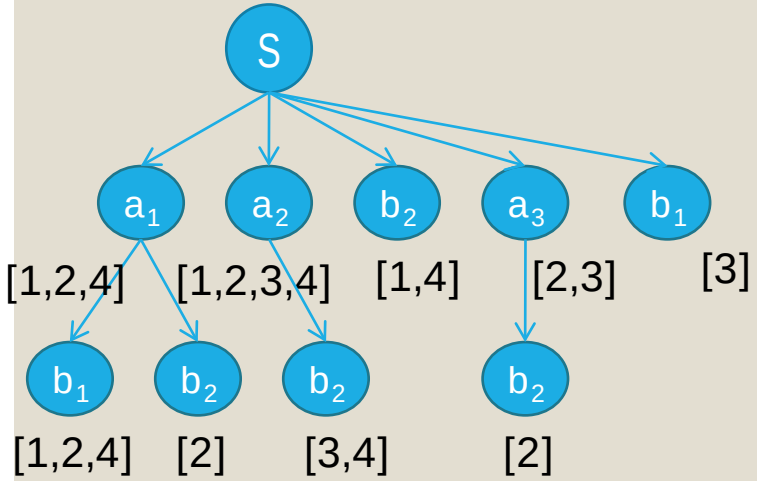
- Projection Tree

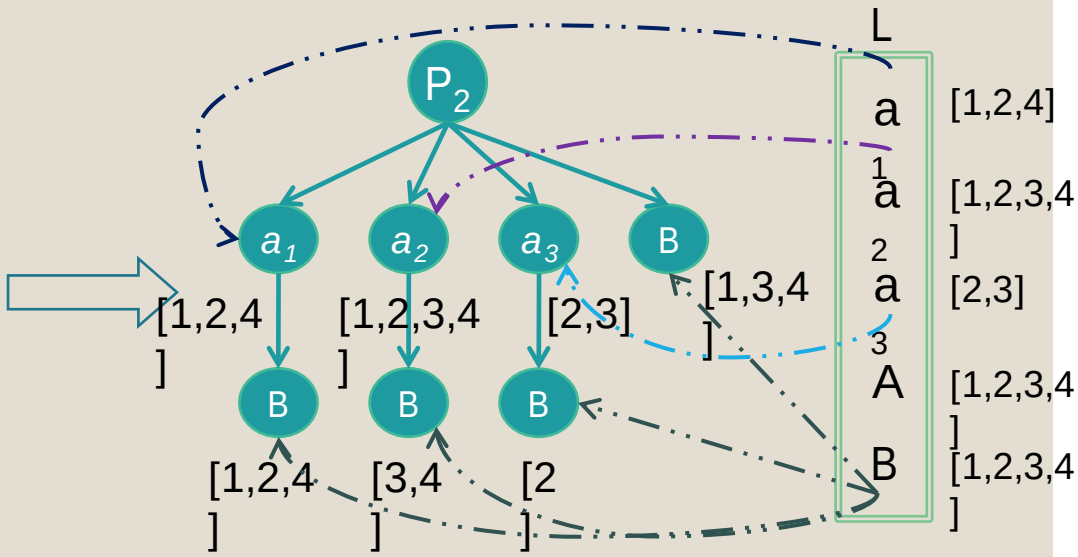


$c_1 = \{A,B\}$

Method Description

- Projection Tree





$$c_2 = \{a_1, a_2, a_3, B\}$$

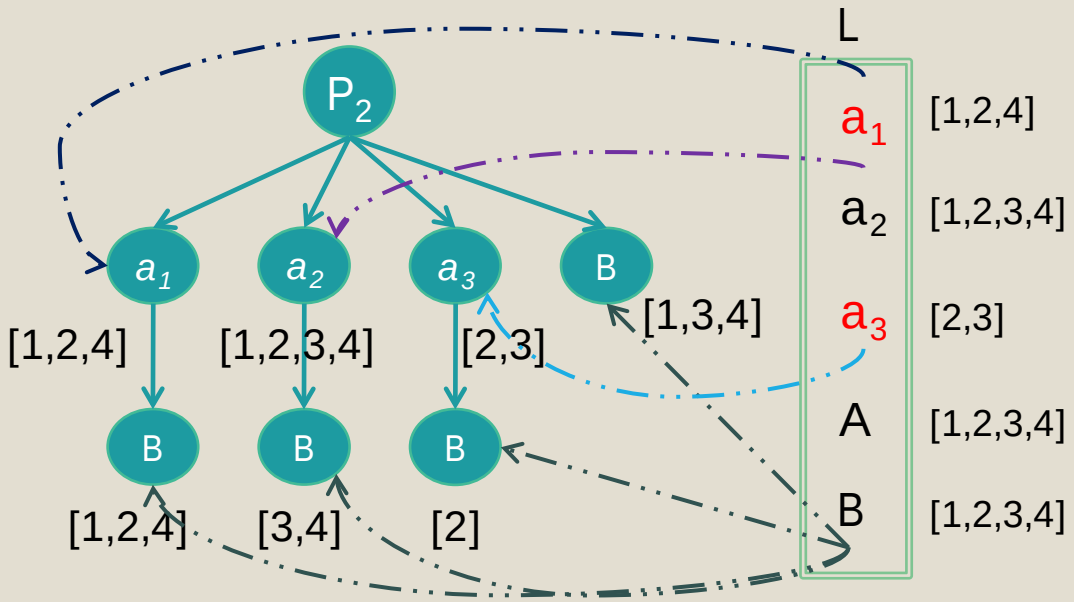
Method Description

- Two main steps:
  - Value check
    - *If it fails the solution is rejected and we check the next one*
  - Structure check
    - *Only if the value check succeeds*
    - Structural disassociation is used to tackle identification through structural information
- The solution is a greedy heuristic

Method Description

- Value check

knowledge={a_1, a_3}.



k=2

m=2

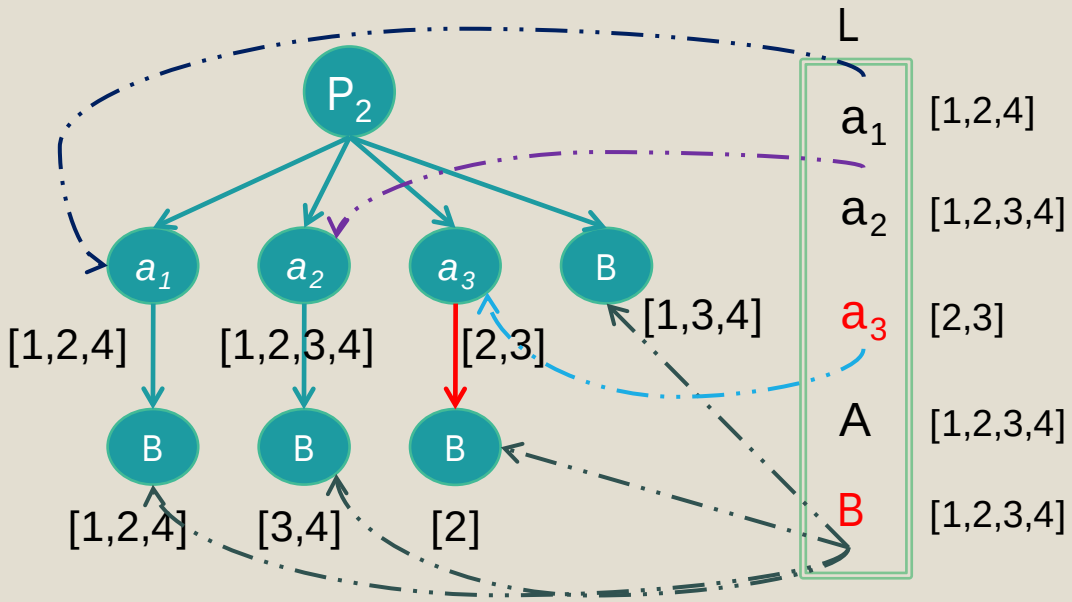
n=1

$L(a_1) \cap L(a_3)=\{2\}$ ➔ $Support(a_1,a_3) = 1 < k$ ➔ c_2 is not a solution

Method Description

- Structure check

knowledge={a_3, B, (a_3→B)}.



k=2

m=2

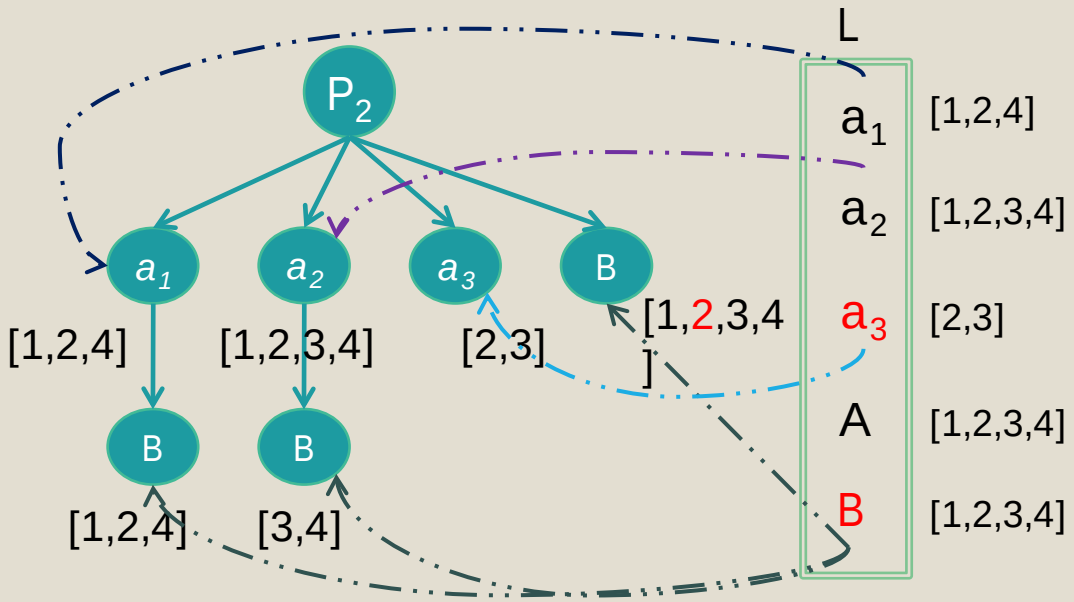
n=1

$L(a_3) \cap L(B)=\{2,3\}$, $L(a_3 \rightarrow B)=\{2\}$ $Support(a_3, B, a_3 \rightarrow B) = 1 < k$ ➔ *structural disassociation*

Method Description

- Structure check

knowledge={a_3, B, (a_3→B)}.



k=2

m=2

n=1

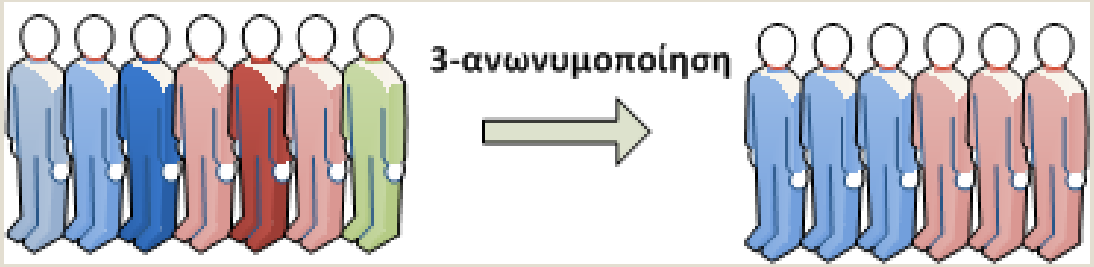
$L(a_3) \cap L(B)=\{2,3\}$, $L(a_3 \rightarrow B)=\{\}$ $Support(a_3, B, a_3 \rightarrow B) = 0$

Conclusions

- Data anonymization allows sharing the valuable information without leaking identifying information
- k^m-anonymity allows the anonymization of sparse, high dimensional data
- Two methods for set-valued data:
  - Generalization based one [PVLDB08]
  - Record splitting based one [PVLDB12]
- Anonymization algorithm for tree-structured data [TKDE15]
  - Structure acts as QI
  - We use generalization and Structural Disassociation

K-anonymity

- Suggested in 2001
- Intuitively each person is hidden in a group of similar person
- Straightforward interpretation of GDPR



k-anonymity

- Each entry becomes indistinguishable from other *k*-1 entries

id	Zipcode	Age	National.	Disease
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

~~id~~	Zipcode	Age	National.	Disease
~~1~~	130**	<30	*	Heart Disease
~~2~~	130**	<30	*	Heart Disease
~~3~~	130**	<30	*	Viral Infection
~~4~~	130**	<30	*	Viral Infection
~~5~~	1485*	≥40	*	Cancer
~~6~~	1485*	≥40	*	Heart Disease
~~7~~	1485*	≥40	*	Viral Infection
~~8~~	1485*	≥40	*	Viral Infection
~~9~~	130**	3*	*	Cancer
~~10~~	130**	3*	*	Cancer
~~11~~	130**	3*	*	Cancer
~~12~~	130**	3*	*	Cancer

l-diversity

- **A group is l-diverse if it contain at least l different "well represented" values**
  - Protection from background knowledge
  - Protection from homogeneity attacks
- "well represented" definition affects monotonicity
  - "well represented" = appear, then it is monotonous
  - "well represented" = probability 1/l, non monotonous

Anatomy

Age	Gender	Zip	Diagnosis
23	M	11000	Pneumonia
27	M	13000	Dyspepsia
35	M	59000	Dyspepsia
59	M	12000	Pneumonia
61	F	54000	Flu
65	F	25000	Gastritis
65	F	25000	Flu
70	F	30000	Bronchitis

Anatomy

1. Partition data

	Age	Gender	Zip	Diagnosis
QI group 1	23	M	11000	Pneumonia
	27	M	13000	Dyspepsia
	35	M	59000	Dyspepsia
	59	M	12000	Pneumonia
QI group 2	61	F	54000	Flu
	65	F	25000	Gastritis
	65	F	25000	Flu
	70	F	30000	Bronchitis

a 2-diverse partition

Anatomy

2. Create one table of quasi identifiers and one of sensitive values

Age	Gender	Zip	Group-ID
23	M	11000	1
27	M	13000	1
35	M	59000	1
59	M	12000	1
61	F	54000	2
65	F	25000	2
65	F	25000	2
70	F	30000	2

Πίνακας ψευδοαναγνωριστικών(QIT)

Group-ID	Diagnosis
1	pneumonia
1	dyspepsia
1	dyspepsia
1	pneumonia
2	flu
2	gastritis
2	flu
2	bronchitis

Πίνακας ευαίσθητων τιμών (ST)

Anatomy

Age	Gender	Zip	Group-ID
23	M	11000	1
27	M	13000	1
35	M	59000	1
59	M	12000	1
61	F	54000	2
65	F	25000	2
65	F	25000	2
70	F	30000	2

Group-ID	Diagnosis	Count
1	Dyspepsia	2
1	Pneumonia	2
2	Bronchitis	1
2	Flu	2
2	Gastritis	1

Negative knowledge?

Bob has zip code 1300 and does not have pneumonia...

Age	Gender	Zip	Group-ID
23	M	11000	1
27	M	13000	1
35	M	59000	1
59	M	12000	1
61	F	54000	2
65	F	25000	2
65	F	25000	2
70	F	30000	2

Πίνακας ψευδοαναγνωριστικών (QIT)

Group-ID	Diagnosis
1	pneumonia
1	dyspepsia
1	dyspepsia
1	pneumonia
2	flu
2	gastritis
2	flu
2	bronchitis

Πίνακας ευαίσθητων τιμών (ST)

Similarity attacks

Original

	ZIP Code	Age	Salary	Disease
1	47677	29	3K	gastric ulcer
2	47602	22	4K	gastritis
3	47678	27	5K	stomach cancer
4	47905	43	6K	gastritis
5	47909	52	11K	flu
6	47906	47	8K	bronchitis
7	47605	30	7K	bronchitis
8	47673	36	9K	pneumonia
9	47607	32	10K	stomach cancer

3-diverse

	ZIP Code	Age	Salary	Disease
1	476**	2*	3K	gastric ulcer
2	476**	2*	4K	gastritis
3	476**	2*	5K	stomach cancer
4	4790*	≥ 40	6K	gastritis
5	4790*	≥ 40	11K	flu
6	4790*	≥ 40	8K	bronchitis
7	476**	3*	7K	bronchitis
8	476**	3*	9K	pneumonia
9	476**	3*	10K	stomach cancer

Value distribution

Bob has zip code 1300. Only 1% of patients have pneumonia

Age	Gender	Zip	Group-ID
23	M	11000	1
27	M	13000	1
35	M	59000	1
59	M	12000	1
61	F	54000	2
65	F	25000	2
65	F	25000	2
70	F	30000	2

Πίνακας ψευδοαναγνωριστικών (QIT)

Group-ID	Diagnosis
1	pneumonia
1	dyspepsia
1	dyspepsia
1	pneumonia
2	flu
2	gastritis
2	flu
2	bronchitis

Πίνακας ευαίσθητων τιμών (ST)

δ-presence

- Existential certainty: The adversary is certain that an individual is described in a private dataset
  - Assumed in previous guaranties
  - Might reveal important information by itself
- The common case is that the adversary has existential uncertainty
- Given public background knowledge P and a private table T
  $\delta=(\delta_{min},\delta_{max})$ – presence holds if:

$$\delta_{min} \leq Pr(t \text{ in } T \mid T^*,P) \leq \delta_{max}$$

Example

P

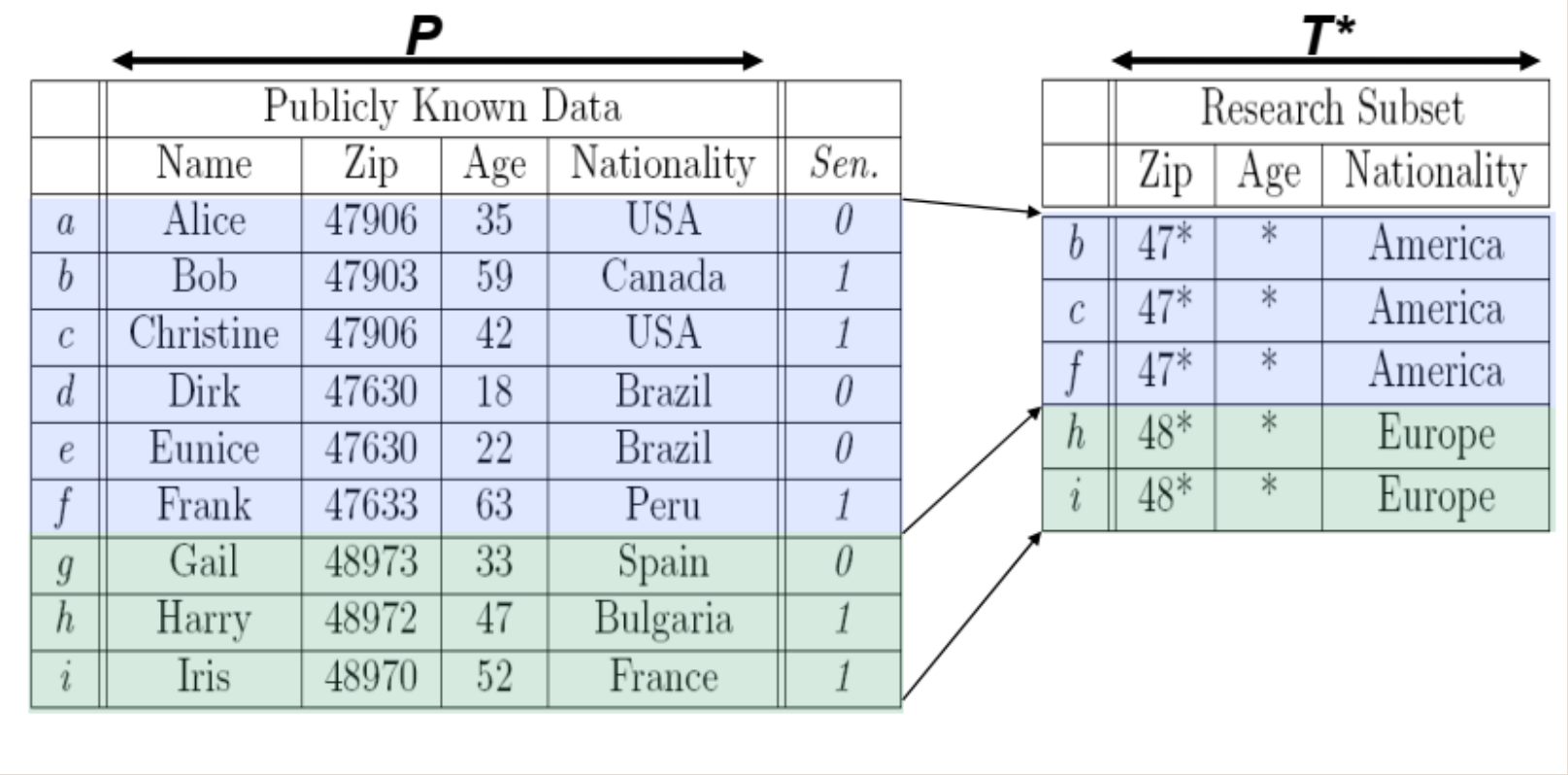
	Publicly Known Data				
	Name	Zip	Age	Nationality	*Sen.*
a	Alice	47906	35	USA	*0*
b	Bob	47903	59	Canada	*1*
c	Christine	47906	42	USA	*1*
d	Dirk	47630	18	Brazil	*0*
e	Eunice	47630	22	Brazil	*0*
f	Frank	47633	63	Peru	*1*
g	Gail	48973	33	Spain	*0*
h	Harry	48972	47	Bulgaria	*1*
i	Iris	48970	52	France	*1*

T

	Research Subset		
	Zip	Age	Nationality
b	47903	59	Canada
c	47906	42	USA
f	47633	63	Peru
h	48972	47	Bulgaria
i	48970	52	France

How to find δ-present generalization of *T*?

Example



	Publicly Known Data				
	Name	Zip	Age	Nationality	*Sen.*
a	Alice	47906	35	USA	*0*
b	Bob	47903	59	Canada	*1*
c	Christine	47906	42	USA	*1*
d	Dirk	47630	18	Brazil	*0*
e	Eunice	47630	22	Brazil	*0*
f	Frank	47633	63	Peru	*1*
g	Gail	48973	33	Spain	*0*
h	Harry	48972	47	Bulgaria	*1*
i	Iris	48970	52	France	*1*

	Research Subset		
	Zip	Age	Nationality
b	47*	*	America
c	47*	*	America
f	47*	*	America
h	48*	*	Europe
i	48*	*	Europe

Example

P*

	Public Dataset			Sen.
	Zip	Age	Nationality	
a	47*	*	America	0
b	47*	*	America	1
c	47*	*	America	1
d	47*	*	America	0
e	47*	*	America	0
f	47*	*	America	1
g	48*	*	Europe	0
h	48*	*	Europe	1
i	48*	*	Europe	1

T*

	Research Subset		
	Zip	Age	Nationality
b	47*	*	America
c	47*	*	America
f	47*	*	America
h	48*	*	Europe
i	48*	*	Europe

$Pr(t_a \in T \mid T^*) = 0.5$

$Pr(t_g \in T \mid T^*) = 0.66$

Privacy guaranties

- Against identity disclosure
  - K-anonymity
- Against attribute disclosure
  - l-diversity, t-closeness
- Against existence disclosure
  - δ-presence

Anonymization tools

ARX

- https://arx.deidentifier.org/

μ-argus

- http://neon.vb.cbs.nl/casc/mu.htm

Amnesia

- https://amnesia.openaire.eu

DIFFERENTIAL PRIVACY

Impossibility theorem

> *... nothing about an individual should be learnable from the database that cannot be learned without access to the database.*
>
> T. Dalenius, 1977

Problem with this approach:

- Analyst knows Bob has green hair.
- Analysts learns from published data that people with green hair have 99% probability of cancer.

- Therefore analyst knows Bob has high risk of cancer, **even if Bob is not in the published data.**

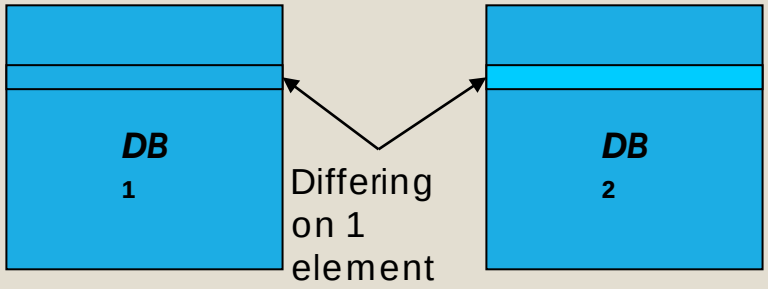
ε-differential privacy

a mathematical definition for the privacy loss associated with any data release drawn from a statistical database

The intuition for the definition of ε-differential privacy is that a person's privacy cannot be compromised by a statistical release if their data are not in the database. Therefore, with differential privacy, the goal is to give each individual roughly the same privacy that would result from having their data removed.

Key idea: **the results of a query or an algorithm should be more or less the same whether a single individual's record participates in the input or not**

Similar results in similar datasets



Differential Privacy

[Dwork'06]

- Randomized Mechanism K provides privacy to an individual, if individual's data effects the output of K only "little"

"adjacent" means "differ in one individual's entry"

K: $\mathcal{D} \rightarrow \mathcal{R}$ ensures ε-DP if for all adjacent datasets D_1, D_2 and for all subsets S of $\mathcal{R}$:

$$\frac{\Pr[K(D1) \in S]}{\Pr[K(D2) \in S]} \leq e^{\varepsilon}$$

Approximate Differential Privacy

K: $\mathcal{D} \rightarrow \mathcal{R}$ ensures ε-DP if for all adjacent datasets D_1, D_2 and for all subsets S of $\mathcal{R}$:

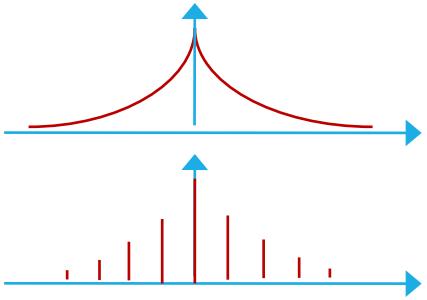
$$\Pr[K(D1) \in S \leq e^{\varepsilon} \Pr[K(D2) \in S] + \delta$$

How to create a randomized mechanism?

- How to randomize a count query?
  - Preserve accuracy?
  - Guarantee ε-differential privacy

For every value that is output:

- Add Laplacian noise Lap(ε/s):
- Or Geometric noise for discrete case:



How much noise?

(Global) Sensitivity of publishing:

- $s = \max_{D,D'} |F(D) - F(D')|$, D, D' differ by 1 individual
- E.g., count individuals satisfying property P: s = 1

The sensitivity is very important for the quality of the output

Group privacy

- If two datasets differ by c individuals then the privacy provided by an ε-differentially private mechanism is

$$\Pr[K(D1) \in S \leq e^{c\varepsilon} \Pr[K(D2) \in S]$$

$$\Pr[\mathcal{A}(D_1) \in S] \leq \exp(\epsilon c) \cdot \Pr[\mathcal{A}(D_2) \in S]$$

Composability

The sequential application of mechanisms{Mi}, each giving $\{\varepsilon_i\}$-differential privacy, gives $(\Sigma_i \varepsilon_i)$-differential privacy.

(We query the data using {Mi},

Parallel composition. If the previous mechanisms are computed on disjoint subsets of the private database then the result would be $(\max(\varepsilon_i))$-differentially private.

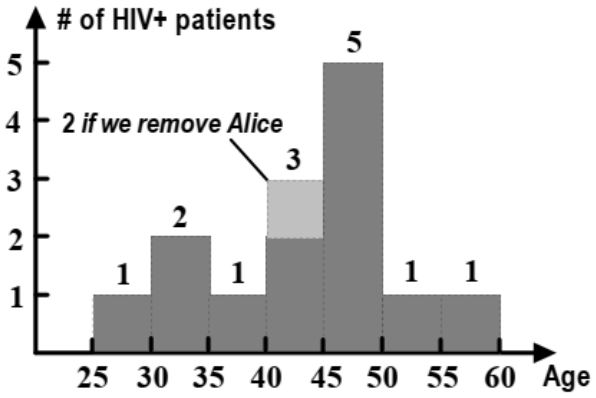
Post processing

- For any deterministic or randomized function F defined over the result of the mechanism A, if A satisfies ε-differential privacy, so does F(A)

Differentially private data release

Name	Age	HIV+
Alice	42	Yes
Bob	31	Yes
Carol	32	Yes
Dave	36	No
Ellen	43	Yes
Frank	41	Yes
Grace	26	Yes
...	...	...

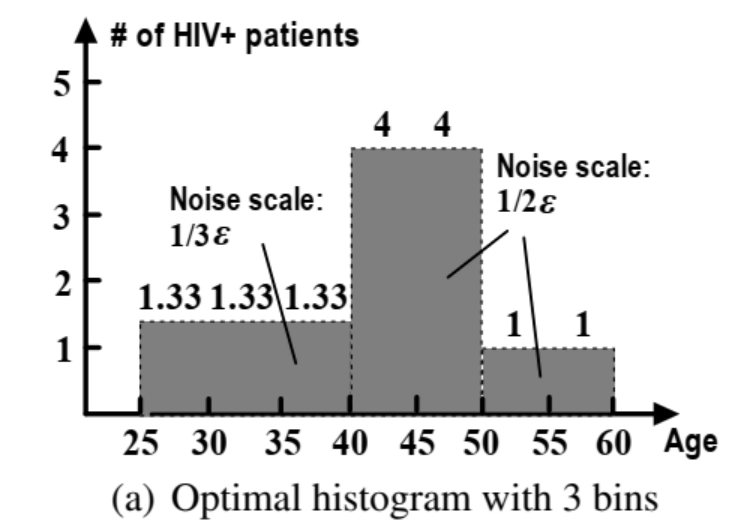
(a) Example sensitive data



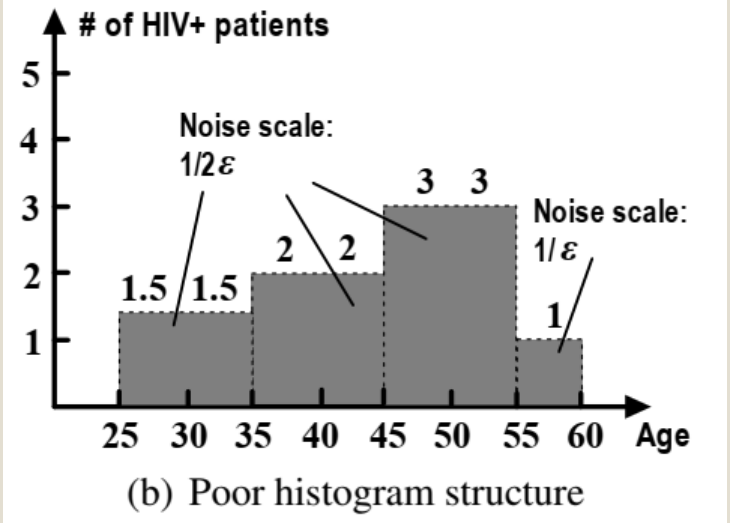
(b) Unperturbed histogram

Example

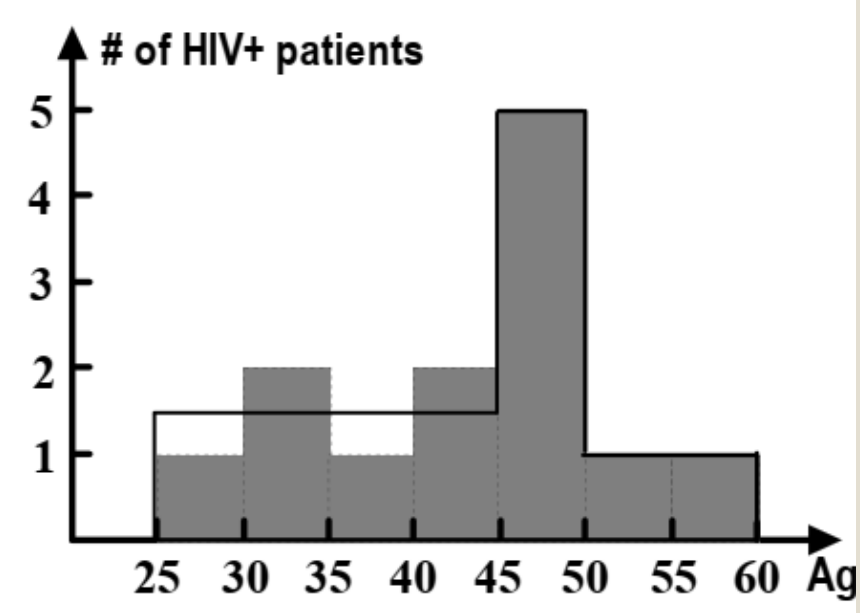
Laplace noise has a scale equal to 1/(number of units)



(a) Optimal histogram with 3 bins



(b) Poor histogram structure



Problems with optimality

Structure must be independent of original data

Solution

- Noise first
  - Add noise, before creating the optimal histogram
- Structure first
  - Use some noise for the structure of the histogram itself

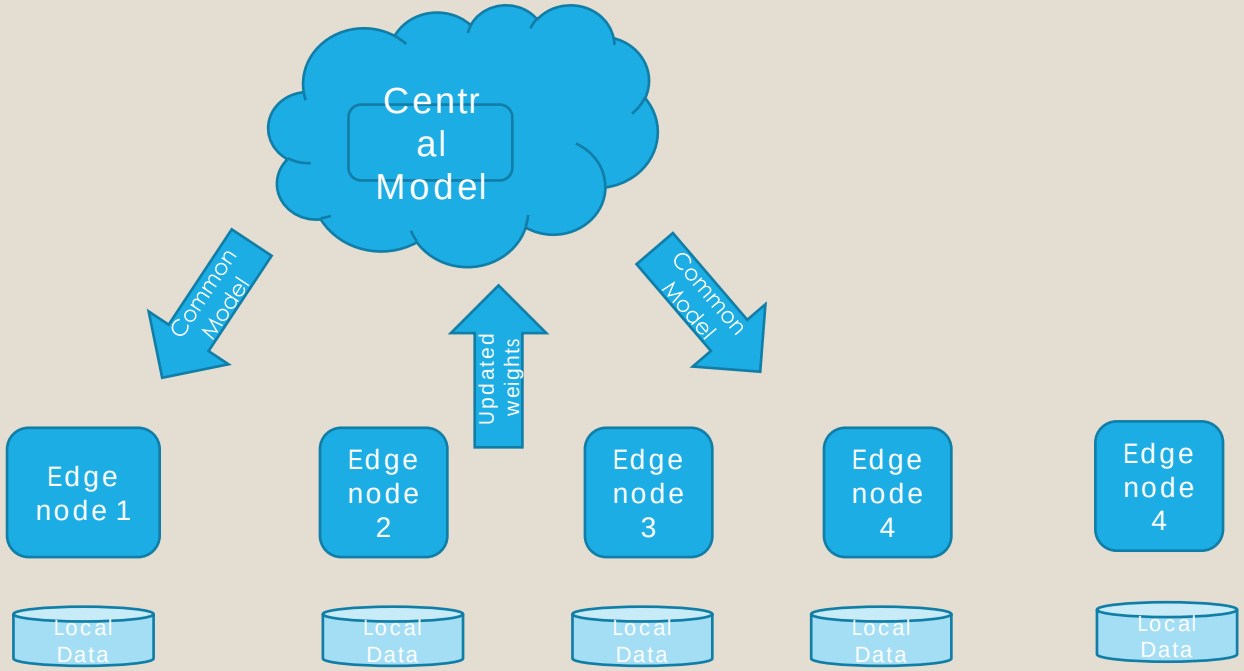
J. Xu, Z. Zhang, X. Xiao, Y. Yang and G. Yu, "Differentially Private Histogram Publication," *2012 IEEE 28th International Conference on Data Engineering*, Arlington, VA, USA, 2012, pp. 32-43, doi: 10.1109/ICDE.2012.48.

Comparison with combinatorial methods

- Property of the mechanism, not of the data
- Stricter, more robust
  - Less utility?
- Better mathematical properties
- Hard to understand parameters
  - Privacy budget ε
- Updates continue to be a problem

PRIVACY ISSUES IN MACHINE LEARNING

Federated Learning



Federated Learning

- Data processing is moving on device:
  - Improved latency
  - Works offline
  - Better battery life
  - Privacy advantages

E.g., on-device inference for mobile keyboards and cameras

- Centralized
  - There is a centralized server to coordinate learning
- Decentralized
  - No centralized server
- Heterogeneous
  - Nodes with different capabilities, i.e., mobile phones and IoT

Federated Learning limitations

- High communication costs
- Systems Heterogeneity
- Statistical Heterogeneity
- Privacy issues
  - *DP*
  - *MPC*

Privacy issues in ML

- ***Privacy attacks***
  - Training data
    - E.g., reveal the identity of patients whose data was used for training a model
  - ML model
    - E.g., reveal the architecture and parameters of a model that is used by an insurance company for predicting insurance rates
    - E.g., reveal the model used by a financial institution for credit card approval

Types of attacks on training data

- Membership inference attack
  - Adversarial goal: determine whether or not an individual data instance x^* is part of the training dataset $\mathcal{D}$ for a model
  - The attack typically assumes black-box query access to the model
  - Cause: overfitting
- Reconstruction attacks
  - Given output labels and partial knowledge of some features, try to recover sensitive features or the full data sample
- Property inference
  - The ability to extract dataset properties which were not explicitly encoded as features or were not correlated to the learning task, is called property inference. An example of property inference is the extraction of information about the ratio of women and men in a patient dataset when this information was not an encoded attribute or a label of the dataset

Defense

- DP
- Regularization.
  - Regularization techniques in machine learning aim to reduce overfitting and increase model generalization performance. Dropout is form of regularization that randomly drops a predefined percentage of neural network units during training.

- Maria Rigaki, Sebastian Garcia: A Survey of Privacy Attacks in Machine Learning. ACM Computing Surveys, Vol 56 (2023)

Training Data Extraction Attack

- ***Training data extraction attack***
  - Carlini (2020) Extracting training data from large language models
- Attack on ***GPT-2*** language model (LM)
  - GPT-2 has 1.5 billion parameters, it is trained on public data collected from the Internet
- The goal of the attack is to analyze output text sequences from GPT-2 and identify text that has been memorized by the model
  - The authors had black-box query access to the GPT-2 model
- Successfully extracted data examples include:
  - Personally identifiable information (PII): names, phone numbers, e-mail addresses
  - News headlines, log files, Internet forum conversations, code
- The extracted information was present in just one document in the training data

Continued..

- Example of training data extraction
  - The authors query GPT-2 by entering the ***prefix***: “East Stroudsburg Stroudsburg...”
  - The model outputted a block of text, which included the full name, phone number, email address, and physical address of the person
  - This information was included in the training data for GPT-2, it was memorized by the model, and extracted by using the training data extraction attack