COMP3260/6360 Data Security

Lecture 11

Prof Ljiljana Brankovic

Lecture Overview

- What is Privacy?
- Technical Aspects of Privacy
  - Sample Database
  - Sample Attack
- Types of Attacks
- K-anonymity
- L-diversity
- Differential Privacy

Privacy

Resources:

Chapter 24, Legal and Ethical aspects, Section 24.3 Privacy

This lecture notes

Note that in-text references and quotes are omitted for clarity of the slides. When you write as essay or a report it is very important that you use both in-text references and quotes where appropriate.

What is Privacy?

DICTIONARY MEANING:

- "A state in which one is not observed or disturbed by other people." *(Oxford Dictionary)*
- "The state of being free from public attention." *(Oxford Dictionary)*
- "The state of being apart from other people or concealed from their view; solitude; seclusion." *(dictionary.com)*
- "The state of being free from unwanted or undue intrusion or disturbance in one's private life or affairs; freedom to be let alone." *(dictionary.com)*

What is Privacy?

LEGAL MEANING:

- Common Law: "The right of people to lead their lives in a manner that is reasonably secluded from public scrutiny, whether such scrutiny comes from a neighbor's prying eyes, an investigator's eavesdropping ears, or a news photographer's intrusive camera; and in statutory law, the right of people to be free from unwarranted drug testing and electronic surveillance." *(legal-dictionary/thefreedictionary.com)*

What is Privacy?

LEGAL MEANING: Australian privacy law and practice (ALRC Report 108):

- **Information privacy**, which involves the establishment of rules governing the collection and handling of personal data such as credit information, and medical and government records. It is also known as 'data protection';

- **Bodily privacy**, which concerns the protection of people's physical selves against invasive procedures such as genetic tests, drug testing and cavity searches;

- **Privacy of communications**, which covers the security and privacy of mail, telephones, e-mail and other forms of communication; and

- **Territorial privacy**, which concerns the setting of limits on intrusion into the domestic and other environments such as the workplace or public space. This includes searches, video surveillance and ID checks.

What is Privacy?

In this course we are mostly interested in information privacy and we will adopt the following definition:

- Privacy may be defined as the claim of individuals, groups or institutions to determine when, how and to what extent information about them is communicated to others (Alan F. Westin, *Privacy and Freedom, New York*: Atheneum, 1967, page 7).

- Privacy is right of individuals to control personal information about themselves.

History of Privacy

- Privacy as we know it is only 150 years old – often referred to as "luxury goods"

- However, there is evidence of privacy attitudes and behaviours across different cultures (Acquisti et al., 2015):
  - including Ancient Rome and Greece, preindustrial Javanese, Balinese and Tuareg society

History of Privacy

Privacy is mentioned in different religions (Acquisti et al., 2015):

- **Bible (Genesis 3.7)**
  "And the eyes of them both were opened, and they knew that they were naked; and they sewed fig leaves together, and made themselves aprons."

- **Quran (49.12)**
  "And do not spy or backbite each other."

- **Talmud (Bava Batra 60a)** instructs homebuilders to position the windows in such a way that they do not directly face windows of their neighbours.

Psychological aspect of privacy

Sidney M. Jourard (1926-1974), 1966, identifies the following:

- the pressures to conform (punishment or invalidation)
- conformity and health (repression)
- the therapeutic and socially necessary functions of privacy
- the psychological function of self-disclosure and concealment
- disclosure inhibited and privacy denied
  - the social risk of private places
  - institutional life (no privacy implies conformity and no individuality)
  - "hell is other people" (changelessness)
  - opting out: "the beatniks"
- possible solutions
  - Individual stratagems and check-out places
  - Education for private life

Psychological aspect of privacy

In short, privacy is experienced as "room to grow in," as freedom from interference, and as freedom to explore, to pursue experimental projects in science, art, work, play, and living. In the name of the *status quo* and other, even more attractive goals, privacy may be eroded. But without privacy and its concomitant, freedom, the cost to be paid for the ends achieved—in terms of lost health, weak commitment to the society, and social stagnation—may be too great.

Abstract model

Name	City	Age	Gender	Status	Post-traumatic stress disorder	Attempted suicide
White	Sydney	34	F	W	4.1	no
Scarlet	Dubbo	27	F	D	3.9	no
Brown	Sydney	45	M	M	4.3	no
Mustard	Perth	32	M	S	2.1	yes
Green	Ballina	76	NB	M	4.8	no
Green	Darwin	32	F	M	4.6	no
Plum	Hobart	25	M	D	2.9	no
Mustard	Darwin	24	M	W	4.2	no
White	Dubbo	51	F	D	3.8	no
Peacock	Sydney	40	NB	M	4.1	no
Black	Ballina	68	F	W	3.6	no
Violet	Dubbo	33	F	M	2.7	no
Aureate	Sydney	28	F	S	3.5	no

Compromise

Example 1

COUNT(City=Darwin and Sex=M and Age<30) = 1

COUNT(City=Darwin and Sex=M and Age<30 and AS=no) = 0

AVG(City=Darwin and Sex=M and Age<30; PTSD) = 4.2

Example 2

COUNT(City=Sydney and Age<37) = 2

COUNT(City=Sydney and Age<37 and AS=no) = 2

Basic Privacy Techniques

Restriction

- query set size control
- query set overlap control
- maximum order control
- partitioning
- cell suppression
- auditing

Modification

- data perturbation
- response perturbation
- data swapping (shuffling)
- random sample

Published Data Table

Name *	City	Age	Gender	Status	Post-traumatic stress disorder	Attempted suicide
White	Sydney	34	F	W	4.1	no
Scarlet	Dubbo	27	F	D	3.9	no
Brown	Sydney	45	M	M	4.3	no
Mustard	Perth	32	M	S	2.1	yes
Green	Ballina	76	NB	M	4.8	no
Green	Darwin	32	F	M	4.6	no
Plum	Hobart	25	M	D	2.9	no
Mustard	Darwin	54	M	W	4.2	no
White	Dubbo	51	F	D	3.8	no
Peacock	Sydney	40	NB	M	4.1	no
Black	Ballina	68	F	W	3.6	no
Violet	Dubbo	33	F	M	2.7	no
Aureate	Sydney	28	F	S	3.5	no

	ID - unique identifier ID={Name}
	QID - Quasi identifier QID={City,Age, Sex}
	Non-sensitive attributes
	Sensitive Attribute

* Strictly speaking, name itself can rarely be considered to be a unique identifier.

Attack Models

We can classify the main attack types into 2 broad categories:

1. **Linkage Attack Models:**
   1. **Record linkage,** where an intruder is able to link an individual to a record in the published data table.
   2. **Attribute linkage,** where an intruder is able to link an individual to a sensitive value in the published data table.
   3. **Table linkage,** where an intruder is able to link an individual to the published data table itself.

Attack Models

2. **Probabilistic attack**. Ideally, the published data should reveal to an intruder as little additional knowledge about individuals as possible, beyond what he/she already knew before seeing the data (background knowledge, or supplementary knowledge). *Probabilistic attack* occurs when the difference between the prior and the posterior knowledge regarding an individual is "significant".

Table 2.1: Privacy models

Privacy Model	Attack Model			
	Record linkage	Attribute linkage	Table linkage	Probabilistic attack
k-Anonymity [201, 217]	✓			
MultiR k-Anonymity [178]	✓			
ℓ-Diversity [162]	✓	✓		
Confidence Bounding [237]		✓		
(α, k)-Anonymity [246]	✓	✓		
(X, Y)-Privacy [236]	✓	✓		
(k, e)-Anonymity [269]		✓		
(ϵ, m)-Anonymity [152]		✓		
Personalized Privacy [250]		✓		
t-Closeness [153]		✓		✓
δ-Presence [176]			✓	
(c, t)-Isolation [46]	✓			✓
ϵ-Differential Privacy [74]			✓	✓
(d, γ)-Privacy [193]			✓	✓
Distributional Privacy [33]			✓	✓

Record Linkage

The intruder is able to link an individual to a record in the published data table.

Recall that in published data tables Unique Identifiers (UIs) are typically removed, so record linkage typically relies on QIDs.

Suppose that an individual A, which the intruder is after, has a value *qid* of the QID, and that the value qid is known to the intruder.

In general, *qid* identifies a group of records in the table. If the size of the group is 1, we have record linkage.

If the size of the group is more than 1, an intruder may still be able to uniquely identify A with the help of additional knowledge.

Record Linkage Example 1

Name *	City	Age	Gender	Status	Post-traumatic stress disorder	Attempted suicide
White	Sydney	34	F	W	4.1	no
Scarlet	Dubbo	27	F	D	3.9	no
Brown	Sydney	45	M	M	4.3	no
Mustard	Perth	32	M	S	2.1	yes
Green	Ballina	76	NB	M	4.8	no
Green	Darwin	32	F	M	4.6	no
Plum	Hobart	25	M	D	2.9	no
Mustard	Darwin	24	M	W	4.2	no
White	Dubbo	51	F	D	3.8	no
Peacock	Sydney	40	NB	M	4.1	no
Black	Ballina	68	F	W	3.6	no
Violet	Dubbo	33	F	M	2.7	no
Aureate	Sydney	28	F	S	3.5	no

QID={City,Age,Sex}

ID(A) = Scarlet

qid=QID(A) = {Dubbo,27,F}

Record Linkage Example 2

Table 2.2: Original patient data

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV
Dancer	Female	30	HIV
Dancer	Female	30	HIV

Table 2.3: External data

Name	Job	Sex	Age
Alice	Writer	Female	30
Bob	Engineer	Male	35
Cathy	Writer	Female	30
Doug	Lawyer	Male	38
Emily	Dancer	Female	30
Fred	Engineer	Male	38
Gladys	Dancer	Female	30
Henry	Lawyer	Male	39
Irene	Dancer	Female	32

A hospital intends to release the Patient Data table (Table 2.2.) to a researcher.

The researches already has access to an external data (Table 2.3.). Additionally, they know that every patient in Table 2.2. also has a record in Table 2.3.

What are ID, QID, non-sensitive and sensitive attributes in each table?

What can the researcher learn by linking these two tables?

k-anonymity

In her famous 2002 paper [5], Sweeny showed that 87% of respondents in 1990 US census (216,000,000) can be uniquely identified using only 3 attributes:

- ZIP code
- Date of Birth
- Gender

In the same paper she famously demonstrated how linking different data sets can be used to compromise sensitive information about individuals.

k-anonymity

The National Association of Health Data Organizations (NAHDO) reported that 37 states in the USA have legislative mandates to collect hospital level data and that 17 states have started collecting ambulatory care data from hospitals, physicians offices, clinics, and so forth [2]. The leftmost circle in Figure 1 contains a subset of the fields of information, or *attributes*, that NAHDO recommends these states collect; these attributes include the patient's ZIP code, birth date, gender, and ethnicity.

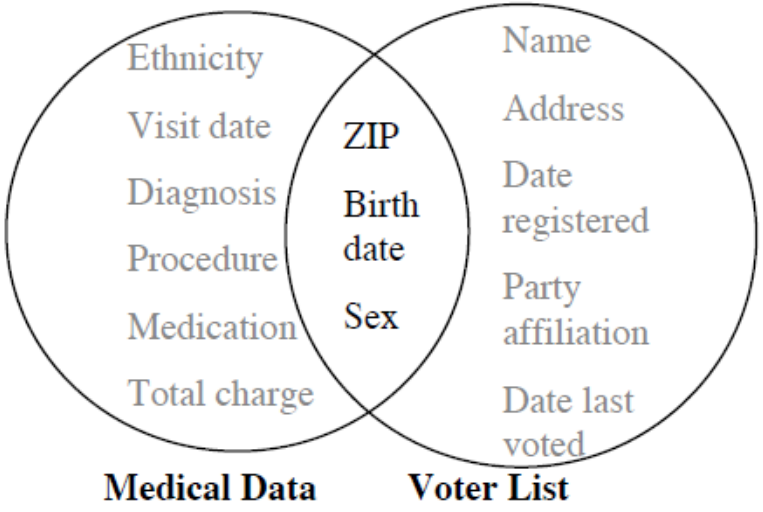

Figure 1 Linking to re-identify data

k-anonymity

In Massachusetts, the Group Insurance Commission (GIC) is responsible for purchasing health insurance for state employees. GIC collected patient-specific data with nearly one hundred attributes per encounter along the lines of the those shown in the leftmost circle of Figure 1 for approximately 135,000 state employees and their families. Because the data were believed to be anonymous, GIC gave a copy of the data to researchers and sold a copy to industry [3].

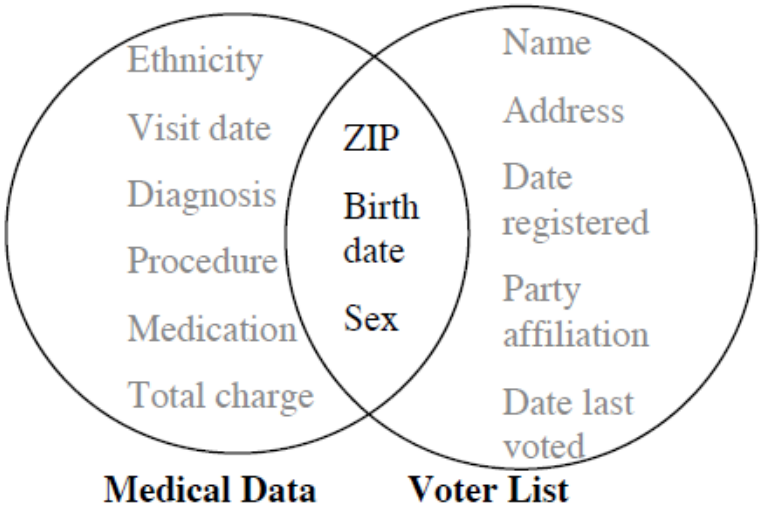

Figure 1 Linking to re-identify data

k-anonymity

For twenty dollars I purchased the voter registration list for Cambridge Massachusetts and received the information on two diskettes [4]. The rightmost circle in Figure 1 shows that these data included the name, address, ZIP code, birth date, and gender of each voter. This information can be linked using ZIP code, birth date and gender to the medical information, thereby linking diagnosis, procedures, and medications to particularly named individuals.

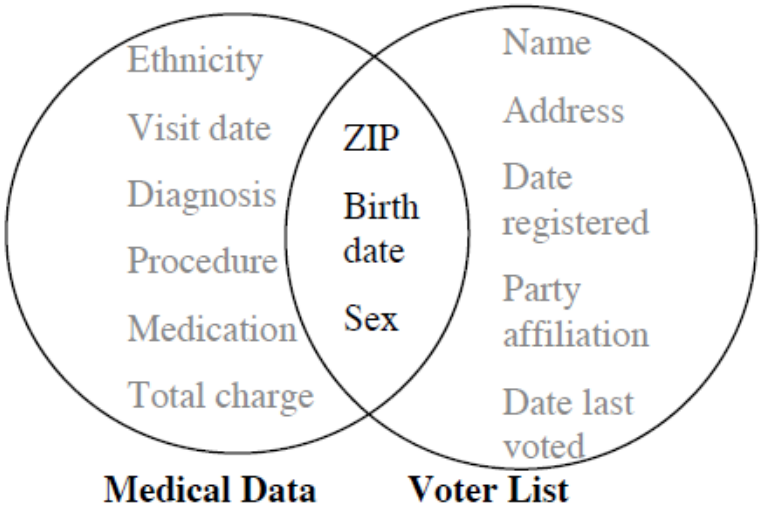

Figure 1 Linking to re-identify data

k-anonymity

For example, William Weld was governor of Massachusetts at that time and his medical records were in the GIC data. Governor Weld lived in Cambridge Massachusetts. According to the Cambridge Voter list, six people had his particular birth date; only three of them were men; and, he was the only one in his 5-digit ZIP code.

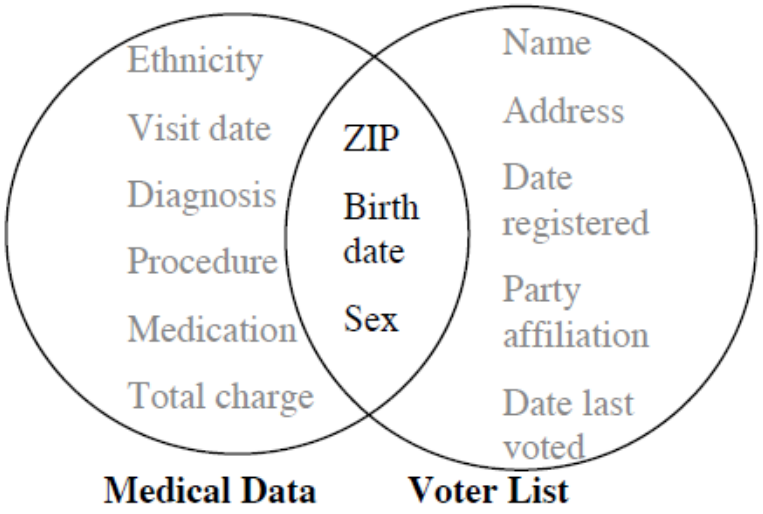

Figure 1 Linking to re-identify data

k-anonymity

In a series of papers [4,5,6], together with Samarati and Sweeny proposed *k-anonymity* in order to prevent record linkage.

For each value *qid* of QID that exist in the data table, there are at least k record having value *qid* in QID.

If a table satisfies this requirement, we say it is *k-anonymous*.

In a *k-anonymous* table, a probability of successfully linking a record to another table on QID is at most $\frac{1}{k}$.

Question: Is the probability of linking a record to another table the same as probability of linking a record to an individual?

k-anonymity Example 1 [1]

Table 2.2: Original patient data

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV
Dancer	Female	30	HIV
Dancer	Female	30	HIV

Table 2.3: External data

Name	Job	Sex	Age
Alice	Writer	Female	30
Bob	Engineer	Male	35
Cathy	Writer	Female	30
Doug	Lawyer	Male	38
Emily	Dancer	Female	30
Fred	Engineer	Male	38
Gladys	Dancer	Female	30
Henry	Lawyer	Male	39
Irene	Dancer	Female	32

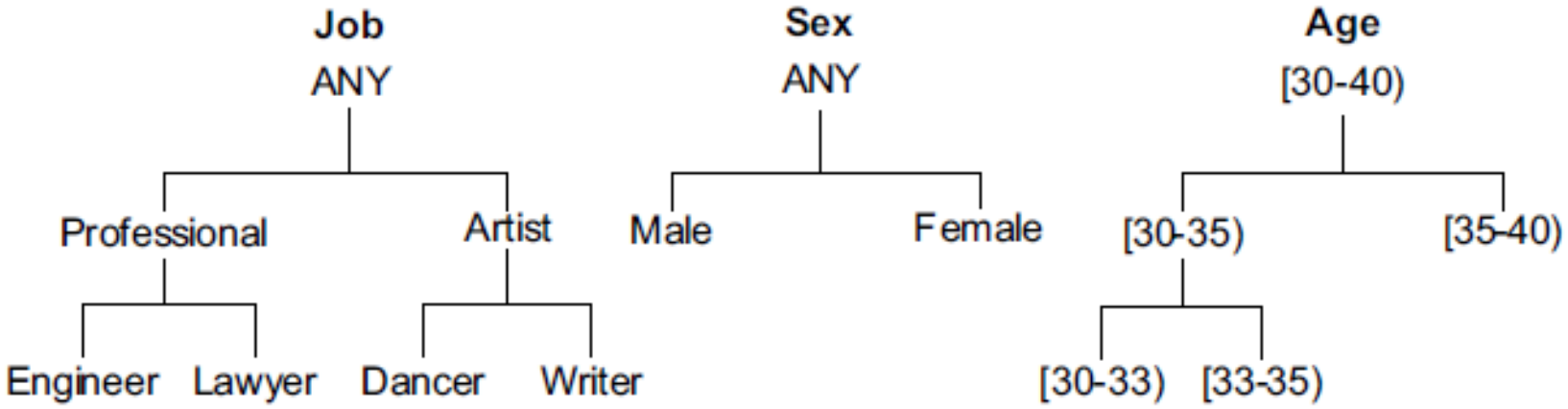

FIGURE 2.1: Taxonomy trees for *Job*, *Sex*, *Age*

k-anonymity Example 1 [1]

Table 2.2: Original patient data

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV
Dancer	Female	30	HIV
Dancer	Female	30	HIV

Table 2.3: External data

Name	Job	Sex	Age
Alice	Writer	Female	30
Bob	Engineer	Male	35
Cathy	Writer	Female	30
Doug	Lawyer	Male	38
Emily	Dancer	Female	30
Fred	Engineer	Male	38
Gladys	Dancer	Female	30
Henry	Lawyer	Male	39
Irene	Dancer	Female	32

Table 2.4: 3-anonymous patient data

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

Table 2.5: 4-anonymous external data

Name	Job	Sex	Age
Alice	Artist	Female	[30-35)
Bob	Professional	Male	[35-40)
Cathy	Artist	Female	[30-35)
Doug	Professional	Male	[35-40)
Emily	Artist	Female	[30-35)
Fred	Professional	Male	[35-40)
Gladys	Artist	Female	[30-35)
Henry	Professional	Male	[35-40)
Irene	Artist	Female	[30-35)

k-anonymity and Quasi-identifiers

We assume that QID is known to the adversary.

How to select a QID?

- Include all attributes that could be known to the adversary into the QID. This would increase privacy but decrease utility of the data.
- Use multiple QIDs. This is useful if the data manager (owner, holder) knows the tables the adversary may use for linking.

k-anonymity and Multiple QIDs

Example 1 [1].

Data manager wants to publish a table $T=(A,B,C,D,S)$, where attributes A, B, C and D are not considered sensitive, but attribute S is. Data user already has access to two other tables, $T_1=(A,B,X)$ and $T_2=(C,D,Y)$. Then data manager can use two QUIs to provide k-anonymity: $QID_1=(A,B)$ and $QID_2=(C,D)$.

Question 1: Does k-anonymity on $QID=(A,B,C,D)$ provide k-anonymity on $QID_1=(A,B)$ and $QID_2=(C,D)$?

Question 2: Does k-anonymity on $QID_1=(A,B)$ and $QID_2=(C,D)$ provide k-anonymity on $QID=(A,B,C,D)$?

Question 3: Let $QID' \subseteq QID$. Does k-anonymity on QID provide k-anonymity on QID'? What about the other way around?

k-anonymity and Multiple Records per Individual

So far we have assumed that each individual corresponds to at most one record in the data table.

However, it is possible that the table contains multiple records per an individual. Such tables are typically obtained by joining multiple tables. For example, a patient can have a record for each visit to his doctor, or a record for each doctor he visits.

Example [1].

Consider a table Patient=(PID, Age, Gender, Disease) and let QID={Age, Gender}. Note that a patient can have more that one disease, therefore more than one record in the table. Thus a group of k records with a same qid may contain less than k patients.

Linkage Attack Models

1. **Linkage Attack Models:**
   1. **Record linkage,** where an intruder is able to link an individual to a record in the published data table.
   2. **Attribute linkage,** where an intruder is able to link an individual to a sensitive value in the published data table.
   3. **Table linkage,** where an intruder is able to link an individual to the published data table itself.

Table 2.1: Privacy models

Privacy Model	Attack Model			
	Record linkage	Attribute linkage	Table linkage	Probabilistic attack
k-Anonymity [201, 217]	✓			
MultiR k-Anonymity [178]	✓			
ℓ-Diversity [162]	✓	✓		
Confidence Bounding [237]		✓		
(α, k)-Anonymity [246]	✓	✓		
(X, Y)-Privacy [236]	✓	✓		
(k, e)-Anonymity [269]		✓		
(ϵ, m)-Anonymity [152]		✓		
Personalized Privacy [250]		✓		
t-Closeness [153]		✓		✓
δ-Presence [176]			✓	
(c, t)-Isolation [46]	✓			✓
ϵ-Differential Privacy [74]			✓	✓
(d, γ)-Privacy [193]			✓	✓
Distributional Privacy [33]			✓	✓

Attribute Linkage

Even if an intruder is not able to perform record linkage, they still may be able to perform attribute linkage and disclose the sensitive attribute value of an individual, or significant information about the sensitive value.

Attribute Linkage Example [1]

Table 2.4: 3-anonymous patient data

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

Table 2.5: 4-anonymous external data

Name	Job	Sex	Age
Alice	Artist	Female	[30-35)
Bob	Professional	Male	[35-40)
Cathy	Artist	Female	[30-35)
Doug	Professional	Male	[35-40)
Emily	Artist	Female	[30-35)
Fred	Professional	Male	[35-40)
Gladys	Artist	Female	[30-35)
Henry	Professional	Male	[35-40)
Irene	Artist	Female	[30-35)

l-diversity

In order to prevent attribute linkage, Machanavajjhala et al. in 2007 proposed the l-diversity privacy model.

Informally, ***l-diversity*** requires that every qid equivalence group contains at least l "well-represented" values in each sensitive attribute.

Distinct l-diversity

The simplest version of l-diversity is ***distinct l-diversity***, where every qid equivalence group contains at least l distinct values in each sensitive attribute.

Note that distinct l-diversity satisfies k-anonymity for k = l.

Distinct l-diversity cannot prevent probabilistic attack as the frequency of different sensitive values can vary greatly.

Entropy l-diversity

Entropy l-diversity requires that for every qid equivalence group and each sensitive attribute we have

$$\sum_{s \epsilon S} p(qid, s) \lg \frac{1}{p(qid, s)} \geq \log(l)$$

where S is an actual domain of the sensitive attribute (the set of values that actually exist in the data table), and p(qid,s) is a probability that a record in qid equivalence group has a sensitive value s.

Entropy l-diversity Example [1]

$$\sum_{s \epsilon S} p(qid, s) \lg \frac{1}{p(qid, s)} \geq \log(l)$$

Example: Calculate l for Distinct and Entropy l-diversity for the following table.

Table 2.4: 3-anonymous patient data

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

Entropy l-diversity Example [1]

Solution:

$$\sum_{s \epsilon S} p(qid, s) \lg \frac{1}{p(qid, s)} \geq \log(l)$$

Table 2.4: 3-anonymous patient data

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

For the first equivalence class with qid={Professional, Male, [35-40]} we have: $\frac{2}{3}\lg\frac{3}{2} + \frac{1}{3}\lg\frac{3}{1} \geq \lg(l)$. Therefore, $l \leq 1.9$

For the second equivalence class with
qid={Artist, Female, [30-35]}
we have: $\frac{3}{4}\lg\frac{4}{3} + \frac{1}{4}\lg\frac{4}{1} \geq \lg(l)$. Therefore, $l \leq 1.8$.

Putting the two inequalities together we get $l \leq 1.8$.

Table 2.1: Privacy models

Privacy Model	Attack Model			
	Record linkage	Attribute linkage	Table linkage	Probabilistic attack
k-Anonymity [201, 217]	✓			
MultiR k-Anonymity [178]	✓			
ℓ-Diversity [162]	✓	✓		
Confidence Bounding [237]		✓		
(α, k)-Anonymity [246]	✓	✓		
(X, Y)-Privacy [236]	✓	✓		
(k, e)-Anonymity [269]		✓		
(ϵ, m)-Anonymity [152]		✓		
Personalized Privacy [250]		✓		
t-Closeness [153]		✓		✓
δ-Presence [176]			✓	
(c, t)-Isolation [46]	✓			✓
ϵ-Differential Privacy [74]			✓	✓
(d, γ)-Privacy [193]			✓	✓
Distributional Privacy [33]			✓	✓

ε-differential privacy

In 2006 Dwork proposed a new model that she termed "different privacy", based on the following principle: "the risk to the record owner's privacy should not substantially increase as a result of participating in a statistical database".

Importantly, unlike other models, differential privacy does not prior and posterior knowledge.

Instead, Dwork argues if the responses from the data table are the same or very similar with and without one particular record, there is no (are at least not significant) privacy risk for that individual.

ε-differential privacy

In other words, if a data set is protected using differential privacy model, then removing or adding a single record will not affect the results of any analysis very much. Therefore, data linkage will not pose a risk to privacy.

Therefore, if an individual chooses not to provide their record, the responses from the data set will remain largely unchanged.

ε-differential privacy

Differential privacy uses a parameter ε which is meant to be a "small" positive value.

We are now ready for the formal definition of ε-differential privacy.

A randomized function F ensures ε-differential privacy if for all data sets T_1 and T_2 differing on at most one record,

$$\left| \ln \frac{p[F(T_1) = s]}{p[F(T_2) = s]} \right| \leq \varepsilon$$

For all s ∈ Range(F), where Range(F) is the set of possible outputs of the randomized function F.

ε-differential privacy

It is important to note that ε-differential privacy does not prevent record and attribute linkages which are prevented by k-anonymity and l-diversity.

On the other hand, ε-differential privacy signals to an individual that if they provide their record to the dataset, then nothing, or almost nothing, can be discovered from the dataset that could not have been discovered without their record.

From this we see that ε-differential privacy prevents table linkage.

ε-differential privacy

For a dataset with n records, Dwork proves that if the number of queries is sub-linear in n, the noise to achieve differential privacy is bounded by $o(\sqrt{n})$.

Differential privacy can be used for both interactive (online) and non-interactive (off-line) query models.

References

[1] B. C. M. Fung, K. Wang, A. W.-C. Fu and P. S. Yu, *Introduction to Privacy-Preserving Data Publishing - Concepts and Techniques*, CRC Press, Tylor & Francis Group, 2011.

[2] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM* Transactions on Knowledge Discovery from Data (TKDD), 1(1), March 2007.

[3] C. Dwork. Differential privacy. In Proc. of the 33rd International Colloquium on Automata, Languages and Programming (ICALP), pages 1–12, Venice, Italy, 2006.

References

[4] P. Samarati and L. Sweeney. Generalizing data to provide anonymity when disclosing information. In *Proc. of the 17th ACM SIGACTSIGMOD-SIGART Symposium on Principles of Database Systems (PODS)*, page 188, Seattle, WA, June 1998.

[5] L. Sweeney, "k-anonymity: a model for protecting privacy", *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, **10**(5), 557-570, 2002.

[6] P. Samarati. Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 13(6):1010–1027, 2001.