Privacy-Preserving Information Access

Gabriel Rovesti

Academic Year 2024/2025

Contents

Chapter 1

Introduction to Privacy

1.1 Definition of Privacy

Privacy is a complex and multifaceted concept that has evolved significantly over time. At its core, privacy refers to an individual's right to keep their personal matters and relationships secret. The Cambridge Dictionary defines privacy as "someone's right to keep their personal matters and relationships secret." This definition highlights three key aspects:

- **Legal aspect**: Privacy is recognized as a right rather than merely a preference.
- **Personal sphere**: Privacy concerns information related to the personal domain of an individual.
- **Secrecy or control**: Privacy implies the ability to maintain certain information confidential or to control its dissemination.

The complexity in defining privacy stems from its subjective nature—what constitutes private information varies across cultures, individuals, and contexts. Privacy can be understood as:

- A state of being free from unwanted observation or intrusion
- The right to control information about oneself
- The ability to maintain boundaries between personal and public spaces

Privacy is deeply intertwined with human psychology. There appears to be a genetic predisposition toward privacy concerns—our ancestors developed wariness about being observed as a survival mechanism against predators. In modern contexts, privacy concerns extend to social acceptance, as many legal but socially stigmatized behaviors require privacy for individuals to avoid judgment.

1.1.1 The Value of Privacy

Privacy holds significant value across multiple dimensions:

Human Rights Perspective

Article 12 of the Universal Declaration of Human Rights states: "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."

This establishes privacy as a fundamental human right, inherent to human dignity and autonomy.

Societal Value

Society recognizes the importance of privacy through extensive legal frameworks:

- National privacy laws
- International frameworks such as the General Data Protection Regulation (GDPR)
- Oversight bodies like the European Data Protection Board (EDPB)

Violating privacy regulations can result in severe penalties—under GDPR, organizations may face fines of up to 4% of annual global revenue or €20 million, whichever is higher.

Consumer Value

Research by CISCO reveals that users place high value on privacy:

- 84% of users express concern about data privacy
- 80% are willing to invest time and money to protect their data
- 48% have switched service providers due to data policies or practices
- 32% are classified as "Privacy Actives"—users who care deeply about privacy and will change their behavior to protect it

Businesses that fail to adequately address privacy concerns risk substantial revenue loss, as nearly half of users will abandon services they perceive as privacy-invasive.

Business Value

From a business perspective, privacy protection is not merely a legal compliance issue but a competitive advantage:

- Building user trust increases retention
- Privacy-respecting practices can become a market differentiator
- Avoiding privacy scandals protects brand reputation

1.2 Solove's Taxonomy of Privacy

Daniel J. Solove developed a comprehensive taxonomy of privacy to categorize the various ways privacy can be compromised or violated. This taxonomy provides a structured approach to understanding privacy issues in different contexts.

Solove's taxonomy identifies four principal categories of privacy-harmful activities:

1.2.1 Information Collection

Activities that gather information about individuals:

- **Surveillance**: Watching, listening to, or recording an individual's activities
- **Interrogation**: Questioning or probing for information

1.2.2 Information Processing

Activities that manipulate data already collected:

- **Aggregation**: Combining various pieces of information about a person
- **Identification**: Linking information to particular individuals
- **Insecurity**: Careless protection of stored information
- **Secondary use**: Using collected information for purposes different from those initially specified
- **Exclusion**: Failure to allow individuals to know about the data collected about them and participate in its handling and use

1.2.3 Information Dissemination

Activities that spread or transfer information:

- **Breach of confidentiality**: Breaking promises to keep personal information confidential
- **Disclosure**: Revealing truthful information about a person that impacts the way others judge their character
- **Exposure**: Revealing another's nudity, grief, or bodily functions
- **Increased accessibility**: Amplifying the accessibility of information
- **Blackmail**: Threatening to disclose personal information
- **Appropriation**: Using the identity of another person to serve someone else's aims
- **Distortion**: Disseminating false or misleading information about individuals

1.2.4 Invasion

Activities that directly affect the individual:

- **Intrusion**: Invasive acts that disturb one's tranquility or solitude
- **Decisional interference**: Government interference with individuals' decisions regarding private matters

Solove's taxonomy provides a valuable framework for analyzing privacy issues in information systems, helping to identify potential vulnerabilities and threats that might otherwise be overlooked.

1.3 European Privacy Regulation

1.3.1 Overview of the GDPR

The General Data Protection Regulation (GDPR) is the cornerstone of privacy legislation in the European Union. Implemented in May 2018, it represents the most comprehensive privacy and data protection framework globally. The GDPR applies to all organizations processing personal data of EU residents, regardless of the organization's location.

1.3.2 Key EU Privacy Authorities

Several authorities oversee privacy protection in the European context:

- **European Union Agency for Cybersecurity (ENISA)**: Provides regulations, best practices, and guidelines on cybersecurity.
- **European Data Protection Board (EDPB)**: Ensures consistent application of the GDPR across the EU through guidelines, recommendations, and identification of best practices.
- **National Data Protection Authorities**: Each EU member state has a national authority (e.g., Italy's Garante per la Protezione dei Dati Personali) responsible for enforcing privacy laws at the national level.

1.3.3 Core GDPR Principles

The GDPR establishes several fundamental principles for data processing:

- **Lawfulness, fairness, and transparency**: Processing must be legal, fair, and transparent to the data subject.
- **Purpose limitation**: Data should be collected for specified, explicit, and legitimate purposes.
- **Data minimization**: Only data necessary for the specified purposes should be processed.
- **Accuracy**: Personal data must be accurate and kept up to date.
- **Storage limitation**: Data should be kept in a form that permits identification of data subjects for no longer than necessary.
- **Integrity and confidentiality**: Data must be processed securely, with protection against unauthorized or unlawful processing and accidental loss.
- **Accountability**: Data controllers must demonstrate compliance with these principles.

1.3.4 Legal Bases for Data Processing

Under GDPR, processing personal data is lawful only if at least one of these conditions applies:

- **Consent**: The data subject has given clear consent for processing their data for a specific purpose.
- **Contract**: Processing is necessary for the performance of a contract with the data subject.
- **Legal obligation**: Processing is necessary for compliance with a legal obligation.
- **Vital interests**: Processing is necessary to protect someone's life.
- **Public interest**: Processing is necessary for performing a task in the public interest or in the exercise of official authority.
- **Legitimate interests**: Processing is necessary for the legitimate interests of the controller or a third party, unless overridden by the interests or rights of the data subject.

1.3.5 Key Rights Under GDPR

The GDPR grants individuals several rights regarding their personal data:

- **Right to be informed**: Individuals have the right to know how their data is being collected, used, and processed.
- **Right of access**: Individuals can request access to their personal data.
- **Right to rectification**: Individuals can have inaccurate personal data rectified or completed if incomplete.
- **Right to erasure (right to be forgotten)**: Individuals can request the deletion of their personal data under certain circumstances.
- **Right to restrict processing**: Individuals can request restriction of processing of their personal data.
- **Right to data portability**: Individuals can obtain and reuse their personal data across different services.
- **Right to object**: Individuals can object to processing of their personal data.
- **Rights related to automated decision-making and profiling**: Protections against purely automated decisions having legal or significant effects.

1.3.6 Privacy by Design and by Default

Two key concepts embedded in the GDPR:

- **Privacy by Design**: Privacy considerations must be integrated into systems and processes from the initial design stage—not added as an afterthought.
- **Privacy by Default**: Systems must be configured with the highest privacy settings by default, requiring no action from the user to protect their privacy.

These principles shift the responsibility for privacy protection from users to organizations, recognizing the significant information and power asymmetry between data subjects and data controllers.

1.4 Data Protection Impact Assessment (DPIA)

1.4.1 Definition and Purpose

A Data Protection Impact Assessment (DPIA) is a structured process designed to describe data processing activities, assess their necessity and proportionality, and help manage risks to the rights and freedoms of natural persons resulting from the processing of personal data.

The primary purposes of conducting a DPIA are:

- Ensuring compliance with data protection laws (GDPR and national guidelines)
- Enhancing organizational accountability
- Offering transparency to data subjects

1.4.2 When to Conduct a DPIA

According to the European Data Protection Board (EDPB), a DPIA is required whenever processing is likely to result in a high risk to the rights and freedoms of individuals, particularly in the following cases (as per Article 35 of GDPR):

- A systematic and extensive evaluation of personal aspects of an individual, including profiling
- Processing of sensitive data on a large scale
- Systematic monitoring of public areas on a large scale
- Implementation of AI models that deal with sensitive data (as specified in the AI Act)

1.4.3 The DPIA Process

A typical DPIA follows this structured process:

1. **Identify the needs**:
   - Identify the data processor and controller
   - Define legal bases for ethical data treatments
   - Define data to be stored
2. **Description of Data Processing**:
   - Detail how data will be collected, stored, and accessed
   - Specify data querying methods
   - Describe how data will be processed and for what purposes
3. **Risk Assessment**:
   - Evaluate external/internal threats from malicious users
   - Assess physical threats to data security
   - Identify potential leakage points for sensitive information
4. **Risk Mitigation**:
   - Implement encryption mechanisms
   - Apply privacy-preserving mechanisms
   - Secure network protocols
   - Establish access controls
5. **Final Steps**:
   - Consultation with stakeholders
   - Regular review and updates (approximately every six months)

1.4.4 Tools for Conducting DPIAs

Several software tools can facilitate the DPIA process:

- **PIA**: Privacy Impact Assessment Tool by the Commission nationale de l'informatique et des liberté (CNIL), recommended by the Italian Personal Data Protection Authority
- **ENISA Software**: Tools developed by the European Union Agency for Cybersecurity
- **Commercial software**: Proprietary solutions like IusPrivacy

A properly conducted DPIA not only ensures regulatory compliance but also helps organizations identify and address privacy risks before they materialize, potentially saving significant resources and reputational damage.

Chapter 2

Foundations of Information Access

2.1 Databases

Databases form the foundation of most information storage and retrieval systems. They represent structured collections of related data kept for long periods and designed for specific purposes.

2.1.1 Definition and Characteristics

A database (DB) is defined as a permanent collection of related data, where data indicates known facts with intrinsic meaning. Key characteristics include:

- Representing aspects of the real world, often called the mini-world or universe of discourse (UoD)
- Forming a coherent and integrated collection of data maintained for extended periods (potentially decades)
- Being designed for specific purposes to satisfy well-identified users' needs

A database management system (DBMS) is the general-purpose software that allows users to create, manage, and update databases. DBMSs offer several important features:

- Managing large volumes of data that exceed available memory
- Facilitating data sharing among applications and users while reducing redundancy and inconsistency
- Providing concurrency control to avoid undesired interactions between users/applications
- Ensuring durability by preserving data integrity even during hardware/software failures
- Guaranteeing data security through authentication and authorization
- Optimizing resource usage for efficiency in both space and time

2.1.2 Database Architecture

The standard ANSI/SPARC architecture divides database systems into three schemas:

- **External schema**: One for each application, describing only the data relevant to that specific application
- **Logical schema**: An integrated representation of data, independent of physical representation

- **Internal schema**: The physical representation of data in storage structures and units

This layered approach provides data independence, allowing changes at one level without necessitating changes at another.

2.1.3 Entity-Relationship Model

The Entity-Relationship (ER) model is a conceptual representation of data structures that allows designers to express the logical schema in graphical form. It identifies:

- Entities: Objects or concepts about which data is stored
- Attributes: Properties of entities
- Relationships: Connections between entities

2.1.4 Relational Model

The relational model, proposed by E. F. Codd at IBM in 1970, represents data in tables or relations. Each relation consists of:

- Tuples (rows): Individual data entries
- Attributes (columns): Properties or characteristics
- Domains: Sets of allowable values for attributes

Structured Query Language (SQL) is the standard language for interacting with relational databases, enabling data definition, manipulation, and querying.

2.1.5 Security Vulnerabilities: SQL Injection

SQL Injection (SQLI) is a critical security vulnerability in database-driven applications. It occurs when malicious SQL statements are inserted into entry fields for execution, allowing attackers to:

- Spoof identity
- Tamper with existing data
- Cause repudiation issues (e.g., voiding transactions)
- Disclose all system data
- Destroy data or make it unavailable
- Gain administrative privileges

The Open Web Application Security Project (OWASP) consistently ranks injection attacks among the most critical web application security risks.

Example of SQL Injection:

```
Original URL:
https://bankingwebsite/show_balances?user_id=984

Resulting SQL:
SELECT accountNumber, balance
```

```
FROM accounts
WHERE account_owner_id=984

Malicious URL:
https://bankingwebsite/show_balances?user_id=0%20OR%201%3D1

Resulting SQL:
SELECT accountNumber, balance
FROM accounts
WHERE account_owner_id=0 OR 1=1
```

The condition ‘1=1‘ is always true, causing the query to return all account information in the database.

2.2 Information Retrieval and Search Engines

2.2.1 Overview of Information Retrieval

Information Retrieval (IR) is concerned with finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections. Modern search engines are the most visible applications of IR technology.

IR vs. Database Retrieval

IR differs fundamentally from database retrieval in several ways:

- **Data structure**: Databases operate on structured data with well-defined attributes, while IR typically works with unstructured or semi-structured text.
- **Query matching**: Databases employ exact matching (Boolean conditions), while IR uses best-match principles (ranking documents by relevance).
- **Result certainty**: Database retrieval guarantees that returned records match the query criteria, whereas IR offers no such guarantee—results are ranked by estimated relevance.
- **World assumption**: Databases operate under a closed-world assumption (what isn't in the database doesn't exist), while IR uses an open-world assumption (relevant documents might not be retrieved).
- **Interaction model**: Database queries are typically one-shot, while IR involves iterative refinement of queries based on results.

2.2.2 Document Representation and Indexing

To make text retrieval efficient, documents must be processed and represented in a way that facilitates comparison with queries.

Document Preprocessing

Before indexing, documents typically undergo several preprocessing steps:

1. **Lexical Analysis/Tokenization**: Breaking text into individual terms or tokens.
2. **Stopword Removal**: Eliminating common words (e.g., "the," "and") that provide little discriminatory value.

3. **Stemming**: Reducing words to their root forms (e.g., "running," "runner," "runs" all reduce to "run").
4. **Lemmatization**: Similar to stemming but producing actual dictionary words as output.

Inverted Index

The primary data structure in IR systems is the inverted index, which maps from terms to their locations in documents:

- **Dictionary**: Contains all unique terms in the collection
- **Postings lists**: For each term, maintains a list of documents containing that term

More sophisticated inverted indexes may include:

- Document frequencies for each term
- Term frequencies within each document
- Positional information (where terms occur in documents)

The inverted index significantly reduces computational complexity by allowing the system to examine only documents containing query terms and reduces spatial complexity by converting words into pointers.

2.2.3 Retrieval Models

Retrieval models provide a framework for ranking documents according to their estimated relevance to a query.

Boolean Model

The Boolean model uses Boolean logic (AND, OR, NOT) to combine terms:

- Documents either match or don't match the query (no ranking)
- Example: "quick" AND "fox" retrieves only documents containing both terms

Vector Space Model

The Vector Space Model represents both documents and queries as vectors in a high-dimensional space where each dimension corresponds to a term:

- Documents and queries become vectors in term space
- Similarity is calculated using measures like cosine similarity
- Enables partial matching and ranking

TF-IDF Weighting

Term Frequency-Inverse Document Frequency (TF-IDF) is a numerical statistic that reflects the importance of a term to a document in a collection:

- **Term Frequency (TF)**: Measures how frequently a term appears in a document (normalized by document length)
- **Inverse Document Frequency (IDF)**: Measures how rare or common a term is across all documents
- **TF-IDF**: The product of TF and IDF, giving higher weight to terms that appear frequently in a document but rarely across the collection

Other Retrieval Models

More advanced retrieval models include:

- **BM25**: A probabilistic ranking function that improves upon TF-IDF with better term saturation and document length normalization
- **Language Models**: Model the probability of generating the query from each document
- **Learning to Rank**: Treat ranking as a machine learning problem
- **Neural IR Models**: Use neural networks to model semantic relationships between queries and documents

2.2.4 Evaluation of IR Systems

Evaluating IR systems involves assessing how well they satisfy users' information needs through standardized methodologies.

Cranfield Paradigm

The Cranfield Paradigm, established in the 1960s, provides a framework for comparative evaluation using:

- Document collections (corpora)
- Topics (surrogate information needs)
- Relevance judgments (assessments of which documents are relevant to which topics)

This approach ensures experiments are comparable and repeatable across different systems and settings.

Evaluation Measures

Set-based Measures

- **Precision**: Proportion of retrieved documents that are relevant
- **Recall**: Proportion of relevant documents that are retrieved
- **F-measure**: Harmonic mean of precision and recall

Rank-based Measures

- **Average Precision (AP)**: Average of precision values calculated at each relevant document in the ranked list
- **Discounted Cumulated Gain (DCG)**: Measures the usefulness of retrieved documents based on their position in the result list, with relevant documents appearing lower in the list receiving a reduced (discounted) contribution to the overall score
- **Normalized Discounted Cumulated Gain (nDCG)**: DCG normalized by the ideal DCG (achieved by perfect ranking) to produce a score between 0 and 1

2.3 Recommender Systems

2.3.1 Overview of Recommender Systems

Recommender Systems (RS) are specialized information filtering tools that suggest items (products, services, information) likely to be of interest to users based on their preferences, past behavior, or other factors.

The Paradox of Choice

Research has shown that while users prefer having options, too many choices can be overwhelming and counterproductive:

- When presented with 24 choices, 60% of consumers approach a display (vs. 40% for 6 choices)
- However, purchase rates drop dramatically with too many options (2% for 24 choices vs. 30% for 6 choices)

This paradox highlights the value of recommender systems in reducing cognitive overload by filtering and prioritizing options.

2.3.2 Types of Recommender Systems

Content-based Recommender Systems

Content-based systems recommend items similar to those a user has liked in the past:

- Each item is characterized by a set of attributes (e.g., genre, actors, director for movies)
- User profiles are built by analyzing the attributes of items they've previously engaged with
- Recommendations are made by matching user profiles with item attributes

Collaborative Filtering Systems

Collaborative filtering (CF) leverages the collective behavior and preferences of multiple users to make recommendations.

Non-personalized (Popularity-based)

- Recommends the most popular items across all users
- Simple to implement and privacy-friendly (can be computed using differential privacy)
- Often performs surprisingly well despite its simplicity
- Lacks personalization, which limits its effectiveness for diverse user bases

Personalized Memory-based CF

- Computes similarity between users or items to infer missing ratings
- User-User CF: "Users who are similar to you liked these items"
- Item-Item CF: "Users who liked this item also liked these other items"
- Requires maintaining the entire rating matrix in memory
- Formula for predicting rating r_{ui} for user u and item i:

$$r_{ui} = \frac{\sum_{v \in U_i} \text{sim}(u, v) \cdot r_{vi}}{\sum_{v \in U_i} \text{sim}(u, v)}$$

  where U_i is the set of users who have rated item i, and $\text{sim}(u, v)$ is the similarity between users u and v

Personalized Model-based CF

- Reduces memory consumption by building a model of user preferences
- Matrix factorization is a common approach: decomposing the rating matrix into two (or more) smaller matrices representing latent features
- More scalable for large-scale applications
- Examples: Singular Value Decomposition (SVD), Neural Network approaches

2.3.3 User Feedback in Recommender Systems

Feedback from users is critical for personalizing recommendations:

- **Explicit feedback**: Direct expressions of preference (e.g., ratings, likes/dislikes)
- **Implicit feedback**: Behavioral signals indicating interest (e.g., viewing time, click-through, purchase history)

This feedback is typically organized into a rating matrix with users as rows, items as columns, and cells containing feedback values. This matrix is usually very sparse (less than 1% filled), presenting challenges for recommendation algorithms.

2.3.4 Evaluation of Recommender Systems

Recommender systems can be evaluated from multiple perspectives:

Accuracy Measures

- **Root Mean Square Error (RMSE)**: Measures the difference between predicted and actual ratings
- **Precision, Recall, F1-Score**: When treating recommendation as a classification task
- **Area Under the ROC Curve (AUC)**: Assesses the system's ability to distinguish between relevant and non-relevant items

Ranking Measures

- **Normalized Discounted Cumulative Gain (nDCG)**: Evaluates the ranking quality of recommendations
- **Average Precision (AP)**: Measures precision at different recall levels

Beyond Accuracy

Modern evaluation considers multiple factors beyond accuracy:

- **Diversity**: Variety in recommendations
- **Novelty**: Recommending previously unknown items
- **Serendipity**: Unexpectedly interesting recommendations
- **Coverage**: Proportion of items that can be recommended
- **Privacy**: Protection of user data during the recommendation process

Chapter 3

Privacy Techniques and Protection Methods

3.1 Microdata Protection

3.1.1 Understanding Data and Privacy

In the context of privacy protection, data refers to information, especially facts or numbers, collected for examination, consideration, and decision-making. Microdata specifically refers to information at the level of individual respondents, where each record relates to a specific person, patient, user, or entity requiring privacy protection.

3.1.2 Types of Attributes in Microdata

When dealing with microdata, attributes can be classified into four categories according to their privacy implications:

- **Identifiers**: Attributes that allow (almost) univocal identification of an individual, such as Social Security Numbers, names, and addresses.
- **Quasi-identifiers**: Attributes that individually provide limited identification capability but, when combined, can identify individuals. Examples include sex, date of birth, ZIP code, and marital status.
- **Confidential attributes**: Attributes concerning the personal sphere of the respondent, such as medical conditions, political opinions, and religious beliefs.
- **Non-confidential attributes**: Attributes that (at current time) cannot be used to disclose users' identity or sensitive information.

3.1.3 Security vs. Privacy

It's important to distinguish between security and privacy in data protection:

- **Security**: Protecting data from unauthorized access, theft, or corruption. Security concerns data access and is fundamental to achieving privacy.
- **Privacy Protection**: Protecting the information contained in or inferable from the data. Privacy concerns arise when data is released or utilized.

3.1.4 Data Release Methods

Data can be released in several forms, each with different privacy implications:

- **Macrodata**: Aggregated data presented in tables (counts, frequencies, magnitudes)
- **Statistical databases**: Databases that accept only distributional queries
- **Microdata**: Information about individual respondents

3.1.5 Protection Methods for Macrodata

Macrodata is typically released in tabular form. Sensitive cells (those providing too much information about few subjects) can be protected through:

- **Cell suppression**: Removing values below a threshold (primary suppression) and additional values to prevent inference (secondary suppression)
- **Rounding**: Rounding values to the nearest multiple of a base number
- **Category roll-up**: Reducing table dimensions by combining categories
- **Sampling**: Using survey data instead of census data
- **Controlled tabular adjustment (CTA)**: Replacing low values with threshold values and adjusting others to maintain sums

3.1.6 Microdata Protection: Risks and Natural Defenses

Microdata Risks

- Highly visible outliers are most vulnerable to identification
- Re-identification attacks through joining microdata with external sources
- Disclosure of identity, attributes, or inferences from statistical properties

Natural Defenses

- Microdata often represents only a subset of the population
- Data may not be up-to-date
- Natural noise complicates linking to other sources
- Format differences between microdata and external sources

3.1.7 Microdata Protection: Masking Techniques

Microdata protection aims to balance privacy protection with maintaining statistical properties. Masking techniques fall into two main categories:

Non-perturbative Methods

These methods do not alter original data values:

- **Sampling**: Removing some observations to decrease identification risk
- **Local suppression**: Replacing attributes or cells with "missing value"
- **Global recoding**: Partitioning attribute domains into intervals and replacing values with interval labels
- **Top/bottom coding**: Converting values above/below thresholds with codes
- **Generalization**: Replacing attributes with more general versions using hierarchies

Perturbative Methods

These methods modify the original data:

- **Rounding**: Replacing values with points in intervals
- **Resampling**: Resampling attribute columns multiple times and sorting within samples
- **Lossy compression**: Treating the data as an image and applying compression algorithms
- **PRAM (Post-Randomized Method)**: Randomly changing categorical values based on probability distributions
- **Random noise**: Adding noise sampled from distributions
- **Swapping**: Exchanging sensitive attribute values between records
- **Rank swapping**: Swapping values among records with limited distance
- **Micro-aggregation/blurring**: Clustering similar records and replacing sensitive values with cluster means

3.1.8 Synthetic Data Generation

Instead of masking original data, synthetic data generation creates entirely new data:

- **Cholesky decomposition**: Using matrix decomposition of the covariance matrix
- **Random response**: Statistical technique that falls under differential privacy
- **Blank and impute**: Removing sensitive values and imputing replacements

3.1.9 Measuring Confidentiality and Information Loss

Disclosure Risk Measures

- **Uniqueness**: Probability that attribute combinations are unique in the population or sample
- **Record linkage**: Assessing the risk of connecting records across datasets using deterministic, probabilistic, or distance-based approaches

Information Loss Measures

- **For continuous data**: MSE, MAE, Mean Variation
- **For categorical data**: Direct comparison, contingency tables, entropy
- **Machine learning performance**: Decrease in prediction accuracy between original and protected data

Trade-off Assessment

The R-U confidentiality map plots utility against disclosure risk, helping to visualize the trade-off between data utility and privacy protection.

3.2 k-Anonymity, ℓ-Diversity, t-Closeness

3.2.1 Identity Linking Attacks

Despite de-identification and sanitization (removing identifiers), privacy breaches remain possible through identity linking attacks, where multiple data sources are combined to de-anonymize individuals. These attacks are facilitated by:

- Increasing availability of public information
- High computational power for data linkage

The most affected domains include medical, financial, electoral, census, customer, and legal data.

3.2.2 k-Anonymity

Definition

k-anonymity addresses identity linking attacks by ensuring that information for any person cannot be distinguished from at least k-1 other individuals in the dataset.

Formally: A table T satisfies k-anonymity with respect to a set of quasi-identifiers QI if for every sequence of values in T[QI] appears at least k times in T[QI].

Implementation Approaches

k-anonymity is typically achieved through two main operations:

- **Generalization**: Replacing specific values with more general ones using domain and value generalization hierarchies.
- **Suppression**: Removing outlier information to reduce the required level of generalization. Suppression can be applied at the tuple, attribute, or cell level.

Generalization Process

Generalization involves establishing domain generalization hierarchies (DGH) and value generalization hierarchies (VGH) that define how values can be generalized. Generalization can be applied at the attribute or cell level.

k-Minimal Generalization

A table T_j is a k-minimal generalization of table T_i if:

- T_j satisfies k-anonymity with minimal required suppression
- The number of suppressed records does not exceed a specified threshold
- No other generalization of T_i satisfying the above conditions requires less generalization

Algorithms

Several algorithms can achieve k-anonymity, including:

- Mondrian: Creates a spatial partition of the data
- Optimal Lattice Anonymization: Finds the optimal generalization
- Incognito: Uses a bottom-up breadth-first search

3.2.3 Limitations of k-Anonymity

k-anonymity protects against identity disclosure but remains vulnerable to:

- **Homogeneity Attack**: When all records in an equivalence class have the same sensitive value, the attacker learns this value despite anonymization.
- **Background Knowledge Attack**: Using external information, attackers can eliminate certain possibilities and narrow down sensitive values.

3.2.4 ℓ-Diversity

Definition

ℓ-diversity addresses k-anonymity limitations by requiring diversity in the sensitive attributes within each equivalence class.

Types of ℓ-Diversity

- **Distinct ℓ-diversity**: Each equivalence class must contain at least ℓ distinct values for the sensitive attribute.
- **Entropy ℓ-diversity**: The entropy of the distribution of sensitive values in each equivalence class must be at least $\log(\ell)$, where entropy is defined as:

$$\text{Entropy}(E) = -\sum_{s \in S} p(s, E) \cdot \log p(s, E)$$

- **Recursive (c, ℓ)-diversity**: The most frequent value does not appear too frequently, and the least frequent values don't appear too rarely.

Implementation

ℓ-diversity is implemented similarly to k-anonymity, using generalization and suppression, but with different termination criteria based on the ℓ-diversity requirement.

3.2.5 Limitations of ℓ-Diversity

ℓ-diversity has several limitations:

- **Complexity**: Achieving ℓ-diversity can be extremely difficult, especially with attributes having skewed distributions.
- **Skewness Attack**: When the distribution of sensitive values in an equivalence class differs significantly from the global distribution.
- **Similarity Attack**: When sensitive values in an equivalence class are distinct but semantically similar (e.g., related diseases).

3.2.6 t-Closeness

Definition

t-closeness addresses ℓ-diversity limitations by requiring the distribution of sensitive attributes in any equivalence class to be close to their distribution in the overall dataset.

Formally: An equivalence class has t-closeness if the distance between the distribution of sensitive attributes in the class and in the whole table is at most t. A table has t-closeness if all equivalence classes have t-closeness.

Earth Mover's Distance

t-closeness uses the Earth Mover's Distance (EMD) to measure the difference between distributions:

- For categorical attributes: Each category is assumed to be at distance 1 from others.
- For numerical attributes: The distance considers the ordering of values.

Properties

t-closeness has useful properties:

- Generalization Property: More general anonymizations maintain t-closeness
- Subset Property: t-closeness for a set of attributes implies t-closeness for any subset

3.2.7 Evaluating Anonymization Quality

Efficiency

Implementing k-anonymity, ℓ-diversity, and t-closeness is NP-hard, making computational efficiency an important consideration.

Equivalence Class Size

Larger equivalence classes reduce re-identification risk but also reduce data utility.

Discernibility

Discernibility measures the degree to which records can be distinguished from each other:

$$\text{discernibility}(T) = \sum_{t \in T} \text{penalty}(t)$$

where:

$$\text{penalty}(t) = \begin{cases} |T| & \text{if } t \text{ is suppressed} \\ |E_t| & \text{otherwise} \end{cases}$$

with E_t being the equivalence class containing t.

Higher discernibility indicates lower data utility but better privacy protection.

3.3 Differential Privacy

3.3.1 Fundamental Law of Information Recovery

The Fundamental Law of Information Recovery states that "overly accurate answers to too many questions will destroy privacy in a spectacular way," with the corollary that "data cannot be fully anonymized and remain useful."

This principle applies to all privacy-preserving data analysis techniques, highlighting a fundamental trade-off between utility and privacy.

3.3.2 Beyond Re-identification

While techniques like k-anonymity focus on preventing re-identification, privacy risks extend beyond identity disclosure:

- Large datasets are vulnerable to differencing attacks
- Summary statistics do not inherently provide privacy
- Seemingly innocuous facts can reveal sensitive information

Traditional approaches like query auditing (refusing to answer potentially revealing queries) or "just a few" policies (sacrificing privacy for outliers) have significant drawbacks.

3.3.3 Differential Privacy Definition

Intuition

Differential privacy aims to ensure that an individual's contribution to a dataset has a negligible impact on the output of analyses performed on that dataset. In essence, the results would be approximately the same whether or not any single individual's data was included.

Formal Definition

A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ε, δ)-differentially private if for all $S \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\Pr[\mathcal{M}(x) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathcal{M}(y) \in S] + \delta$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$. If $\delta = 0$, we say that $\mathcal{M}$ is ε-differentially private.

Key components:

- ε (epsilon): Privacy budget controlling the trade-off between privacy and accuracy
- δ (delta): Probability of the privacy guarantee being broken
- $\|x - y\|_1 \leq 1$: Condition that datasets x and y are neighboring (differ by one record)

3.3.4 Randomized Mechanisms

Differential privacy introduces randomness to protect privacy. A randomized algorithm $\mathcal{M}$ with domain A and discrete range B is associated with a mapping M: A $\rightarrow$ (B), where (B) is the probability simplex over B (vectors of probabilities that sum to 1).

Privacy Loss

The privacy loss incurred by observing outcome o is:

$$L_{\mathcal{M}(x)||\mathcal{M}(y)}(o) = \log\left(\frac{\Pr[\mathcal{M}(x) = o]}{\Pr[\mathcal{M}(y) = o]}\right)$$

For ε-differential privacy, this loss must be bounded by ε for all neighboring datasets and all possible outputs.

3.3.5 Key Properties of Differential Privacy

Composition

Differential privacy guarantees compose gracefully:

- **Sequential Composition**: If mechanisms $\mathcal{M}_1, \mathcal{M}_2, ..., \mathcal{M}_n$ are $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$-differentially private respectively, then their sequential application is $(\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n)$-differentially private.
- **Parallel Composition**: If mechanisms applied to disjoint subsets of the data are $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$-differentially private, then their combined use is $\max(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$-differentially private.

Group Privacy

An ε-differentially private mechanism $\mathcal{M}$ provides $(k\varepsilon)$-differential privacy for groups of size k. This means if datasets differ in k records, the privacy bound increases linearly with k.

Post-processing

Differential privacy is immune to post-processing: any function of a differentially private output remains differentially private with the same parameters. Formally, if $\mathcal{M}$ is (ε, δ)-differentially private and f is any function, then $f \circ \mathcal{M}$ is also (ε, δ)-differentially private.

3.3.6 Mechanisms for Differential Privacy

Randomized Response

One of the earliest privacy-preserving techniques, randomized response protects privacy through plausible deniability:

1. Respondent flips a coin
2. If tails, respond truthfully
3. If heads, flip another coin and respond "Yes" for heads, "No" for tails

This method is ln(3)-differentially private and allows statistical inferences about the true proportion of "Yes" responses through the formula:

$$p = 2Y - \frac{1}{2}$$

where Y is the observed proportion of "Yes" responses and p is the true proportion.

Laplace Mechanism

For numerical queries, the Laplace mechanism adds noise drawn from a Laplace distribution calibrated to the sensitivity of the query:

Given a function $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$, the Laplace mechanism is defined as:

$$\mathcal{M}_L(x, f, \varepsilon) = f(x) + (Y_1, ..., Y_k)$$

where Y_i are drawn independently from the Laplace distribution $\text{Lap}(\Delta f/\varepsilon)$.

The ℓ_1-sensitivity Δf represents the maximum change in the function's output when one record changes:

$$\Delta f = \max_{x,y:\|x-y\|_1 \leq 1} \|f(x) - f(y)\|_1$$

For counting queries (e.g., "How many records satisfy property P?"), the sensitivity is 1, so noise is drawn from $\text{Lap}(1/\varepsilon)$.

Gaussian Mechanism

Similar to the Laplace mechanism but uses Gaussian noise:

$$\mathcal{M}_G(x, f, \varepsilon, \delta) = f(x) + (Y_1, ..., Y_k)$$

where $Y_i \sim \mathcal{N}(0, \sigma^2)$ with $\sigma \geq \frac{\sqrt{2\ln(1.25/\delta)}\Delta_2 f}{\varepsilon}$.

The Gaussian mechanism requires $\delta > 0$ (relaxed differential privacy) but can provide better utility for high-dimensional outputs since it uses ℓ_2-sensitivity.

Exponential Mechanism

For non-numerical outputs, the exponential mechanism selects outputs with probability exponentially proportional to their utility:

Given a utility function $u : \mathbb{N}^{|\mathcal{X}|} \times \mathcal{R} \to \mathbb{R}$, the exponential mechanism $\mathcal{M}_E(x, u, \mathcal{R})$ selects and outputs an element $r \in \mathcal{R}$ with probability proportional to $\exp\left(\frac{\varepsilon u(x,r)}{2\Delta u}\right)$, where Δu is the sensitivity of the utility function.

This mechanism is particularly useful for selection problems (e.g., finding the best threshold, auction pricing) where small perturbations to numeric answers might significantly reduce utility.

3.3.7 Advanced Techniques

Report Noisy Max

Reports the index of the largest noisy count without releasing the counts themselves. Given m count queries, adding $\text{Lap}(1/\varepsilon)$ noise to each and returning only the index of the maximum provides ε-differential privacy.

Sparse Vector Technique

Answers a stream of queries until finding the first one that exceeds a threshold:

This approach allows answering many queries with privacy cost depending only on the number of above-threshold queries, not the total number of queries.

Algorithm 1 Above-Threshold(D, Q, T, ε)

```
Let T_p = T + Lap(2/ε)
for q_i in Q do
  Let v_i = Lap(4/ε)
  if q_i(D) + v_i ≥ T_p then
    Output ⊤
    Break
  else
    Output ⊥
  end if
end for
```

3.3.8 Local Differential Privacy

Central vs. Local DP

- **Central DP**: A trusted central entity holds all data and applies DP mechanisms before releasing results
- **Local DP**: Users apply DP mechanisms to their own data before sharing it with any central entity

Local DP provides stronger privacy guarantees since raw data never leaves the user's device, but typically results in lower utility.

RAPPOR (Google)

RAPPOR (Randomized Aggregatable Privacy-Preserving Ordinal Response) is Google's implementation of local DP:

1. Encode a value using a Bloom filter B
2. Create a permanent randomized version B' (memoized)
3. Further randomize B' to create a response S for transmission

The procedure maintains ε-DP with:

$$\varepsilon = \ln\left(\frac{1 - \frac{1}{2}f}{f/2}\right) + \ln\left(\frac{1-q}{p}\right)$$

where f, p, and q are parameters controlling the randomization.

Count Mean Sketch (Apple)

Apple's approach for local DP:

1. Randomly select a hash function from a set
2. Hash the value and create a vector with all zeros except at the hash position
3. Randomly flip bits with controlled probability
4. Send the result and hash function index to the server

Combined with the Sequence Fragment Puzzle technique, this allows discovering previously unknown strings while maintaining privacy.

Low Communication LDP (Microsoft)

Microsoft's method enables computing population statistics using minimal communication (single bit):

For a value $X_i \in [0, m]$, send a single bit Y_i such that:

$$\Pr[Y_i = 1] = \frac{e^{\varepsilon/2} \cdot X_i/m}{1 + (e^{\varepsilon/2} - 1) \cdot X_i/m}$$

The average value can be estimated as:

$$\bar{X} \approx \frac{m \cdot \frac{\sum Y_i}{n}}{(e^{\varepsilon/2} - 1) \cdot \frac{\sum Y_i}{n} + \frac{1}{e^{\varepsilon/2}}}$$

3.3.9 Real-world Applications

US Census

The US Census Bureau applies differential privacy to protect respondent privacy while maintaining data utility:

- Uses a composition of approaches with combined $\varepsilon = 19.61$
- Maintains exact counts at the state level but perturbs statistics at lower geographic levels

Private ML

Differential privacy is increasingly applied to machine learning:

- DP-SGD (Differentially Private Stochastic Gradient Descent) for training neural networks
- PATE (Private Aggregation of Teacher Ensembles) for knowledge transfer
- Federated Learning with DP for collaborative model training

3.4 Geomasking

3.4.1 Importance of Geographic Data

Geographic data are essential for discovering patterns in various domains:

- Consumer behavior analysis
- Census and demographic studies
- Medical and epidemiological research

A Geographic Information System (GIS) is a specialized database containing geographic information that facilitates spatial analysis and visualization.

3.4.2 Privacy Risks in Geographic Data

Geographic data presents significant privacy risks:

- Precise locations can uniquely identify individuals
- Reverse geocoding can convert map coordinates to street addresses

- Combining location data with other information enables powerful inference attacks

Historical examples demonstrate these risks:

- Identification of Hurricane Katrina victims from newspaper maps
- Identification of crime victims in Vienna from published maps
- Studies showing 68% of probable victims were identifiable using reverse geocoding and online directories

3.4.3 Geographic Masking Techniques

Geomasking alters the coordinates of point location data to limit re-identification risk while preserving general spatial patterns.

Transformation Methods

- **Rotation**: Changing the orientation of point patterns
- **Scaling**: Expanding or contracting distances between points
- **Translation**: Shifting point locations

While these methods can protect individual locations, they often distort spatial relationships and patterns important for analysis.

Randomization Methods

Randomization introduces controlled uncertainty into precise locations:

- **Random Direction, Fixed Radius**: New locations are randomly placed on a circle of fixed radius around original points.
- **Random Direction, Random Radius**: New locations are randomly placed within a disk of fixed radius. This tends to place points toward the periphery since there's more area away from the center.
- **Random Direction, Gaussian Radius**: Direction is uniformly random, but distance follows a Gaussian distribution, concentrating points closer to the original location.
- **Donut Masking**: New locations fall within a ring (area between two concentric circles) with fixed internal and external radii. This enforces a minimum displacement for stronger privacy.
- **Binomial Gaussian Displacement**: Combines aspects of donut masking and Gaussian displacement, with direction uniform and distance following a truncated Gaussian distribution.

3.4.4 Displacement Calibration

The appropriate displacement magnitude varies by context:

- Should be inversely proportional to population density
- Large displacements in sparsely populated areas may be necessary
- Smaller displacements suffice in densely populated areas

3.4.5 Spatial k-Anonymity

Spatial k-anonymity extends traditional k-anonymity to geographic contexts by ensuring a location cannot be distinguished from at least k-1 other locations:

- Often measured using the nth nearest neighbor approach
- Considers how many points could have been masked to the same position as the actual masked point
- The minimum anonymity across all points defines the overall spatial k-anonymity

3.4.6 Metric Differential Privacy for Geomasking

Traditional differential privacy can be too restrictive for spatial data. Metric differential privacy relaxes the requirement by considering the distance between points:

Given a finite set W, a metric $d_W : W \times W \to \mathbb{R}$, and a privacy parameter $\varepsilon > 0$, a mechanism M: W $\to$ W satisfies ε-d_W privacy if for all $w_1, w_2, w \in W$:

$$P[M(w_1) = w] \leq e^{\varepsilon \cdot d_W(w_1, w_2)} \cdot P[M(w_2) = w]$$

This approach allows points that are spatially close to be more likely masked to the same location, respecting the natural spatial relationships in the data.

Chapter 4

Privacy-Preserving Information Access Applications

4.1 Privacy-Preserving IR Systems

4.1.1 Sensitive Documents in Information Retrieval

In many IR contexts, documents have varying levels of sensitivity, creating a tension between finding relevant information and protecting sensitive content:

- **Corporate email search**: Private emails may be inadvertently retrieved
- **Medical record search**: Some documents may contain particularly sensitive information (e.g., related to STDs, abortion, gender affirmation)
- **E-discovery**: Finding relevant documents for legal proceedings while protecting irrelevant sensitive information

4.1.2 Sensitivity Classification

Feature Selection for Sensitivity Classification

For emails and documents, various features can indicate sensitivity:

- **Lexical features**: Term frequencies, document length, subject line content
- **Temporal features**: Day of the week, time of day
- **Metadata**: Sender/receiver information, job roles, departments
- **Attachment information**: Number and types of attachments

Sensitivity Classification Approaches

Sensitivity classification typically combines traditional classification with Learning to Rank (LTR) approaches:

- Features include both standard IR features (BM25, TF-IDF) and sensitivity indicators
- Neural networks or gradient boosting trees often serve as classifiers

4.1.3 Privacy-Preserving Ranking Strategies

Relevance-Only

The simplest approach ranks documents solely based on relevance to the query, ignoring sensitivity considerations.

- **Pros**: Simple implementation, optimal relevance effectiveness
- **Cons**: No privacy protection

Pre-Filtering + Relevance

Train the IR model using only non-sensitive documents. At runtime, filter out sensitive documents before ranking.

- **Pros**: Intuitive approach, modular components
- **Cons**: Requires separate training for two models

Relevance + Post-Filtering

Rank all documents based on relevance, then filter out sensitive documents from the results.

- **Pros**: Uses best relevance ranking
- **Cons**: Computationally inefficient, may leave gaps in results

Demoted Relevance

Train the IR model with reduced relevance scores for sensitive documents.

- **Pros**: Simple implementation
- **Cons**: Challenging to determine appropriate demotion for relevant but sensitive documents

Joint Relevance-Sensitivity

Model relevance and sensitivity together in a unified approach.

- **Pros**: Can leverage multiple signals, potentially optimal balance
- **Cons**: More complex development, balancing competing objectives

4.1.4 Cost-Sensitive Performance Measures

Traditional IR evaluation measures like precision and recall don't account for the harm of exposing sensitive documents. Cost-sensitive measures address this gap.

Simple Ternary Measure (TERN)

TERN assigns one of three values to each retrieved result:

- 1: Document is both relevant and non-sensitive
- 0: Document is non-relevant but also non-sensitive
- -M: Document is sensitive (regardless of relevance)

For a query, TERN is:

- -M: If any sensitive document is shown
- 1: If no sensitive documents are shown and at least one relevant document is shown
- 0: Otherwise

SENS

SENS extends nDCG by:

- Calculating discounted cumulative gain only for non-sensitive documents
- Using an ideal ranking considering only non-sensitive documents
- Assigning -M if any sensitive document is retrieved

Cost-Sensitive DCG (CS-DCG)

CS-DCG allows more nuanced evaluation by incorporating position-dependent costs for sensitive documents:

$$\text{CS-DCG@k} = \sum_{i=1}^{k} \frac{g_i}{\log_2(i+1)} - \sum_{i=1}^{k} c_i \cdot s_i$$

where:

- g_i is the gain (relevance) of document at position i
- c_i is the cost of showing a sensitive document at position i
- s_i is 1 if document at position i is sensitive, 0 otherwise

Normalized CS-DCG (nCS-DCG) is calculated using the best and worst possible rankings:

$$\text{nCS-DCG} = \frac{\text{CS-DCG} - \text{wCS-DCG}}{\text{bCS-DCG} - \text{wCS-DCG}}$$

4.1.5 Test Collections for Privacy-Preserving IR

A privacy-preserving IR test collection requires:

- Documents
- Topics
- Relevance judgments
- Sensitivity judgments

Two approaches for creating such collections:

Simulated Collections

Existing collections can be adapted by defining sensitivity criteria:

- OHSUMED (medical literature): Documents with certain Medical Subject Headings (MeSH) terms (e.g., STDs, urogenital diseases) marked as sensitive

Purpose-Built Collections

- Avocado Email Collection: Contains emails from a defunct IT company with sensitivity judgments made by annotators considering specific personas (representing different privacy concerns)

4.2 Privacy in Recommender Systems

4.2.1 Privacy Concerns in Recommender Systems

Recommender systems rely on user data to generate personalized recommendations, which creates significant privacy implications:

- Collection of explicit feedback (ratings, likes) and implicit feedback (viewing time, clicks)
- Building of detailed user profiles revealing preferences and behaviors
- Potential for inferring sensitive attributes not explicitly shared
- Risks from data breaches or unauthorized access

4.2.2 Privacy Threats in Collaborative Filtering

Collaborative filtering systems face particular privacy challenges:

Memory-based CF

- Requires storing and processing the complete user-item interaction matrix
- User preferences are directly compared, potentially exposing sensitive information
- Often relies on centralized storage of all user data

Model-based CF

- Latent factor models may encode and potentially expose private information
- Model parameters might reveal user preferences through inference attacks
- Training data might be vulnerable to membership inference attacks

4.2.3 Privacy-Preserving Recommendations

Anonymization Approaches

- **k-anonymity**: Ensuring user profiles cannot be distinguished from at least k-1 other profiles
- **Differential Privacy**: Adding calibrated noise to recommendations or model updates
- **Pseudonymization**: Replacing identifiers with pseudonyms while maintaining ability to generate recommendations

Encryption-Based Methods

- **Homomorphic Encryption**: Allowing computations on encrypted data without decryption
- **Secure Multi-Party Computation**: Enabling multiple parties to jointly compute recommendations without revealing inputs
- **Encrypted Matrix Factorization**: Performing matrix factorization while keeping user data encrypted

4.2.4 Federated Learning for Recommender Systems

Federated learning allows training recommendation models without centralizing user data:

- User data remains on local devices
- Devices train local models on their own data
- Only model updates (not raw data) are sent to a central server
- The server aggregates updates to improve the global model
- Updated global model is distributed back to devices

Privacy Enhancements for Federated Learning

- **Differential Privacy**: Adding noise to model updates
- **Secure Aggregation**: Cryptographic protocols ensuring the server can only see the aggregate of updates, not individual contributions
- **Model Compression**: Reducing the information content of model updates

Challenges

- Statistical heterogeneity: User data is non-IID (not independently and identically distributed)
- Systems heterogeneity: Varying computational capabilities across devices
- Communication efficiency: Limited bandwidth and connectivity
- Privacy-utility trade-off: Stronger privacy protection often reduces recommendation quality

4.2.5 Evaluation in Privacy-Preserving Recommender Systems

Evaluating privacy-preserving recommender systems requires balancing multiple objectives:

Recommendation Quality Metrics

- Accuracy measures: RMSE, MAE
- Ranking measures: nDCG, MAP
- Beyond-accuracy measures: diversity, novelty, coverage

Privacy Metrics

- Formal guarantees: values in differential privacy
- Empirical privacy: Success rates of inference attacks
- Information disclosure: Mutual information between private data and model outputs

Efficiency Metrics

- Computational overhead
- Communication costs
- Memory requirements

4.3 Privacy in Databases and Data Publishing

4.3.1 Types of Disclosures in Database Privacy

When publishing or providing access to databases, multiple types of privacy disclosures can occur:

- **Identity disclosure**: Re-identification of individuals through combinations of attributes
- **Attribute disclosure**: Inferring sensitive attributes for known individuals
- **Inferential disclosure**: Deducing information with high probability based on statistical properties

4.3.2 Query Restriction and Statistical Databases

Statistical databases allow only distributional queries while attempting to protect individual records:

- **Query size restriction**: Only allowing queries that apply to a minimum number of records
- **Query overlap control**: Restricting queries that have high overlap with previous queries
- **Auditing**: Tracking queries to identify potential disclosure risks

However, these approaches face limitations:

- Differencing attacks remain possible
- Query denial itself reveals information
- Auditing is computationally expensive
- Restricted queries limit data utility

4.3.3 Differential Privacy in Database Queries

Differential privacy offers a rigorous approach to database privacy:

- **Query perturbation**: Adding calibrated noise to query results
- **Private data release**: Generating differentially private synthetic data or statistics
- **Private SQL**: Systems like PINQ (Privacy Integrated Queries) supporting DP versions of SQL operations

4.3.4 Private Data Publishing

Several approaches enable privacy-preserving data publication:

- **k-anonymity, -diversity, t-closeness**: For microdata protection
- **Cell suppression, rounding, perturbation**: For sensitive cells in tabular data
- **Synthetic data generation**: Creating artificial data that preserves statistical properties without exposing real records

4.3.5 Privacy-Preserving Data Mining

Data mining on sensitive databases requires specialized privacy-preserving techniques:

- **Randomization-based methods**: Adding noise to data before mining
- **Condensation-based approaches**: Creating condensed representations that preserve statistical properties
- **Cryptographic techniques**: Using secure multi-party computation for distributed data mining
- **DP data mining**: Developing differentially private algorithms for common mining tasks (clustering, classification, association rule mining)

Chapter 5

Conclusion and Future Directions

5.1 Summary of Key Concepts

This course has explored the multifaceted challenges of preserving privacy in information access systems:

- Privacy is a fundamental right with significant societal, legal, and personal importance
- Computational privacy techniques provide mathematical frameworks for privacy protection:
  - k-anonymity, -diversity, and t-closeness for protecting microdata
  - Differential privacy for providing rigorous privacy guarantees
  - Geomasking for protecting location information
- Information access systems present unique privacy challenges:
  - Databases must balance detailed data storage with privacy protection
  - Search engines must filter sensitive content while maintaining relevance
  - Recommender systems must generate personalized suggestions without exposing personal preferences

5.2 Emerging Trends

5.2.1 Privacy-Enhancing Technologies (PETs)

Recent advances in privacy-enhancing technologies are reshaping information access:

- **Federated learning**: Training models across multiple devices without centralizing data
- **Homomorphic encryption**: Computing on encrypted data without decryption
- **Secure multi-party computation**: Collaborative computation without revealing inputs
- **Trusted execution environments**: Secure enclaves for sensitive computations
- **Advanced obfuscation techniques**: Sophisticated methods to mask sensitive information

5.2.2 Regulatory Landscape

The regulatory environment continues to evolve, influencing privacy-preserving approaches:

- Expanding data protection regulations globally
- Growing requirements for privacy impact assessments
- Increasing emphasis on privacy by design and default
- Emerging standards for privacy-preserving artificial intelligence

5.2.3 AI and Privacy

The rise of advanced AI systems creates new privacy challenges and opportunities:

- **Challenges**: More powerful inference capabilities, complex data requirements
- **Opportunities**: Privacy-preserving machine learning, AI-based privacy protection

5.3 Future Research Directions

Key areas for future research in privacy-preserving information access include:

- **Usable privacy**: Making privacy protections more understandable and manageable for users
- **Privacy-utility trade-offs**: Developing better approaches to balance privacy protection with information utility
- **Longitudinal privacy**: Addressing privacy erosion over time through multiple data releases
- **Privacy in multimodal systems**: Extending privacy protection to systems handling text, images, audio, and video
- **Privacy in large language models**: Addressing memorization, extraction, and inference concerns in foundation models

5.4 Conclusion

Privacy-preserving information access represents a critical intersection of technical innovation, ethical considerations, and regulatory compliance. As information systems become more pervasive and powerful, the need for robust privacy protections grows accordingly.

The approaches explored in this course—from formal privacy models to practical implementation techniques—provide a foundation for developing information access systems that respect user privacy while delivering valuable functionality. The future of privacy-preserving information access will likely involve increasingly sophisticated combinations of these approaches, tailored to specific contexts and requirements.

By understanding both the theoretical underpinnings and practical applications of privacy protection in information access, we can contribute to developing systems that harness the power of data while respecting fundamental privacy rights.

Bibliography

[1] Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3), 477-564.

[2] European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Union*, L119, 1-88.

[3] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference* (pp. 265-284). Springer.

[4] Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05), 557-570.

[5] Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 3-es.

[6] Li, N., Li, T., & Venkatasubramanian, S. (2007). t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd International Conference on Data Engineering* (pp. 106-115). IEEE.

[7] Armstrong, M. P., Rushton, G., & Zimmerman, D. L. (2008). Geographically masking health data to preserve confidentiality. *Statistics in medicine*, 27(20), 4928-4944.

[8] Konomi, S., & Shang, S. (2017). Foundations of Location Based Services. In *Frontiers of Computing Systems Research* (pp. 1-30). Springer.

[9] Cleverdon, C. W. (1997). The Cranfield tests on index language devices. In *Readings in information retrieval* (pp. 47-59). Morgan Kaufmann.

[10] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

[11] Ricci, F., Rokach, L., & Shapira, B. (2011). *Introduction to recommender systems handbook*. Springer.

[12] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4), 211-407.

[13] Erlingsson, Ú., Pihur, V., & Korolova, A. (2014). RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1054-1067).

[14] McSherry, F. D. (2009). Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data* (pp. 19-30).

[15] Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security* (pp. 1310-1321).

[16] Bagdasaryan, E., Poursaeed, O., & Shmatikov, V. (2019). Differential privacy has disparate impact on model accuracy. In *Advances in Neural Information Processing Systems* (pp. 15479-15488).

[17] McDonald, D. W., & Cranor, L. F. (2008). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society*, 4, 543-568.

[18] Domingo-Ferrer, J., & Torra, V. (2002). Inferring knowledge from privacy protected tabular data. In *Accuracy 2002, International Conference on Knowledge Extraction from Statistical Data Bases*.

[19] Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620.

[20] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30-37.

[21] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 308-318).

[22] Zinn, D., & Wardenga, N. (2021). Semantic vs. syntactic neural networks for detecting sensitive information in text. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 18, pp. 16388-16396).

[23] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Aguera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics* (pp. 1273-1282).

[24] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1-2), 1-210.

[25] Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy* (pp. 111-125). IEEE.

[26] Chatzikokolakis, K., Andrés, M. E., Bordenabe, N. E., & Palamidessi, C. (2013). Broadening the scope of differential privacy using metrics. In *International Symposium on Privacy Enhancing Technologies Symposium* (pp. 82-102). Springer.

[27] Feyisetan, O., Balle, B., Drake, T., & Diethe, T. (2020). Privacy-and utility-preserving textual analysis via calibrated multivariate perturbations. In *Proceedings of the 13th International Conference on Web Search and Data Mining* (pp. 178-186).

[28] Xu, C., Ren, J., Zhang, D., Zhang, Y., Qin, Z., & Ren, K. (2020). A differentially private text perturbation method using regularized mahalanobis metric. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 7101-7112).