Protecting Personal Data

Presented By : Dharak Sanjaybhai Pandadiya
Matriculation Number : 4243201

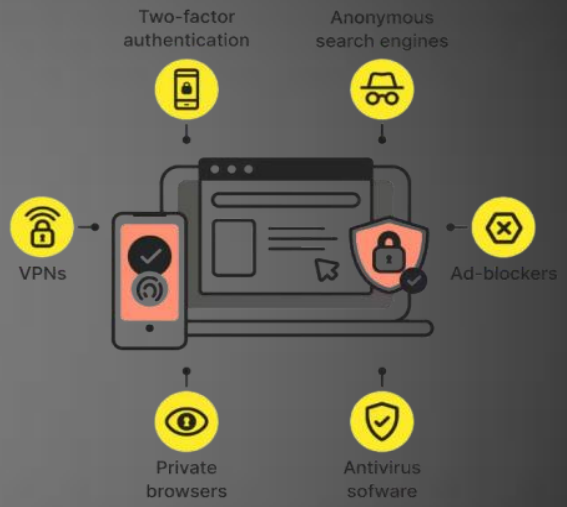
Index

1. Overview of protecting personal data
2. Different technical approaches
3. Advantages and Disadvantages of three approaches
4. Describe for three approaches with one scenario
5. Where it is particularly well suited

Overview of Protecting Personal Data and Its Main Components

➢ In today's digital age, personal data has become one of the most valuable assets, making it a prime target for cyberattacks, identity theft, and unauthorized surveillance.

➢ Protecting personal data is crucial to maintaining privacy and ensuring the security of individuals' sensitive information, such as financial details, health records, and personal identification.



❖ Example of Personal Data : -

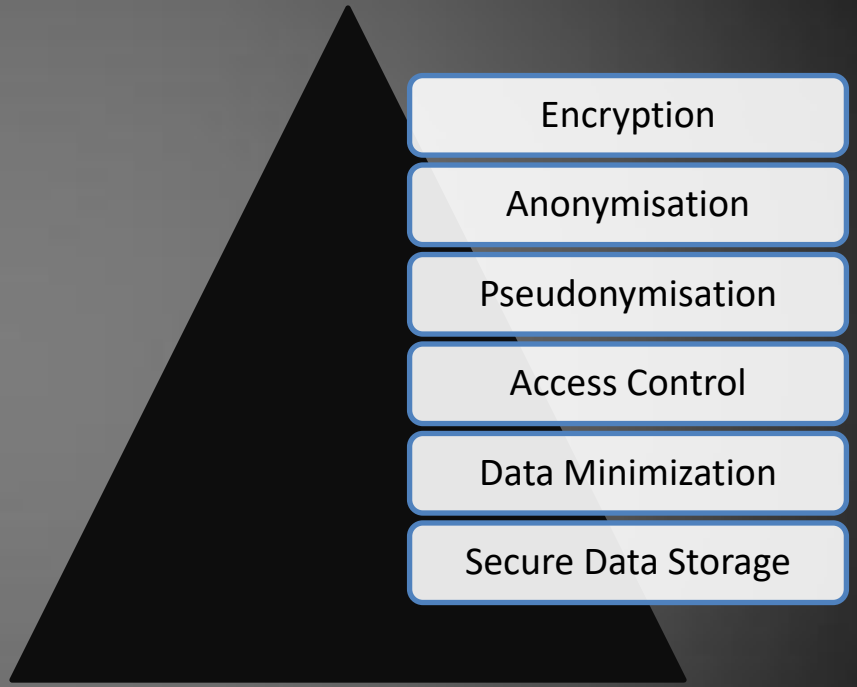
- a name and surname
- a home address
- an email address
- an internet protocol (IP) address
- an identification card number

Different Technical Approaches

There are many different technical approaches that can be used to protect personal data.

1. Encryption
2. Anonymisation
3. Pseudonymisation

There are three main technical approaches that can be used to protect personal data and explain with example.



Encryption

❑ Explain : -

- Encryption is the process of converting data into a coded format that is unreadable to unauthorized users.
- It uses algorithms and encryption keys to transform plaintext into ciphertext, which can only be reverted to its original form using a decryption key.

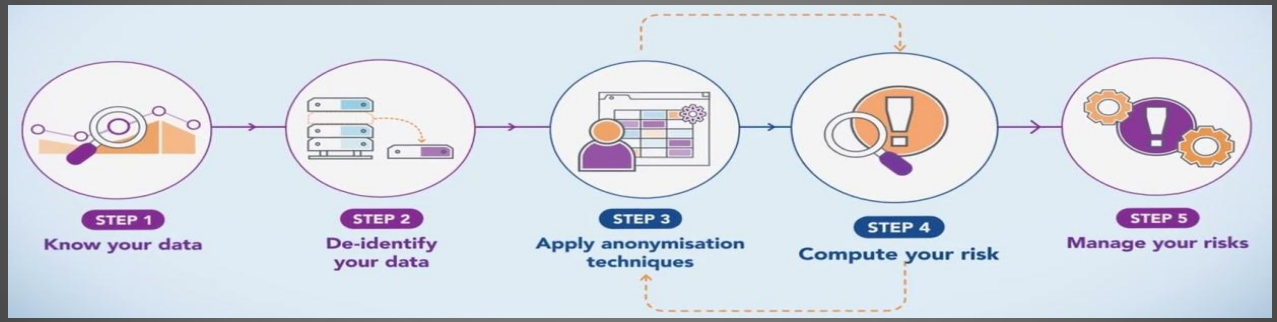
❑ Purpose : -

- The primary purpose of encryption is to protect the confidentiality of data during storage or transmission, ensuring that even if the data is intercepted or accessed by unauthorized parties, it cannot be read or used without the correct decryption key.

Anonymisation

❑ Explain : -

- Anonymisation is the process of permanently removing personal identifiers from data so that individuals cannot be identified, even if the data is combined with other data.
- This often involves altering or removing direct identifiers (e.g., names, addresses) and indirect identifiers (e.g., date of birth, geographical location).



❑ Purpose : -

- The goal of anonymisation is to prevent the identification of individuals within a dataset, enabling the data to be used without concerns about privacy violations.
- Once anonymised, data can no longer be traced back to any specific individual.

Pseudonymisation

❑ Explain : -

- Pseudonymisation involves replacing identifiable information within a dataset with pseudonyms (i.e., artificial identifiers or codes).
- Unlike anonymisation, pseudonymisation allows the re-identification of individuals if necessary by using additional information (e.g., a key or lookup table).

❑ Purpose : -

- The purpose of pseudonymisation is to reduce the risk of personal data exposure while maintaining the ability to reverse the process and re-identify individuals if needed (such as in case of legal requirements or for research purposes).

Compare of the three approaches

Approach	Encryption	Anonymisation	Pseudonymisation
Definition	The process of converting data into a coded format (ciphertext) that can only be decrypted with a key.	The process of permanently removing or altering personal identifiers so that individuals cannot be re-identified.	The process of replacing personally identifiable information (PII) in a dataset with artificial identifiers.
Purpose	To ensure data confidentiality by making it unreadable to unauthorized parties.	Privacy by preventing identification that can be no longer be tracked.	Privacy protection with possible re-identification.
Protection Level	Protects confidentiality, High security.	Strong privacy, Very high security in terms of identification.	Medium privacy but allows re-identification if the mapping key is available.
Data Modification	Data is transformed into unreadable form.	Data is altered or removed to prevent identification.	Data is transformed but can be linked the additional information (Key or mapping) is available.
Example Use Case	Securing sensitive documents, transactions.	Sharing aggregated data for research without identifying individuals.	Medical research, surveys, or anonymized customer data where re-identification is possible but controlled.

Advantages and Disadvantages

1. Encryption : -

➢ Advantages : -

- Encryption offers numerous advantages, primarily centered around ensuring data confidentiality and security.
- It protects sensitive information by transforming it into unreadable ciphertext, ensuring that only authorized parties with the correct decryption key can access the original data.
- This level of protection is crucial for safeguarding personal, financial, and business information, both during storage and transmission.

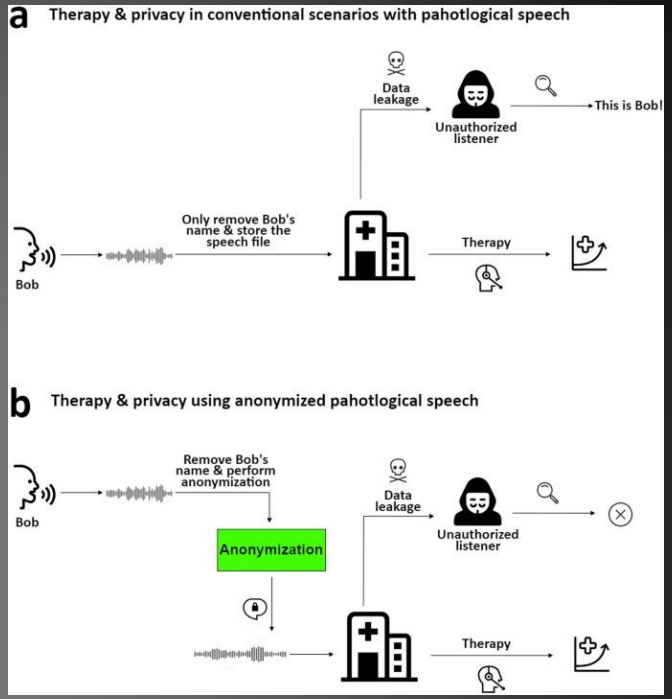
➢ Disadvantages : -

- One of the main challenges is the complexity of key management, as the encryption process relies on the secure storage and handling of encryption keys.
- Additionally, encryption can introduce performance overhead, as the process of encrypting and decrypting data consumes computational resources, potentially slowing down systems, especially when dealing with large volumes of data.

2. Anonymisation : -

➢ Advantages : -

- Provides a high level of privacy, as individuals cannot be re-identified.
- It allows organizations to use and share data for analysis, research, and statistical purposes without compromising the privacy of individuals.
- Anonymisation also reduces the risks associated with data breaches, as exposed anonymised data cannot be traced back to any individual, minimizing potential harm.



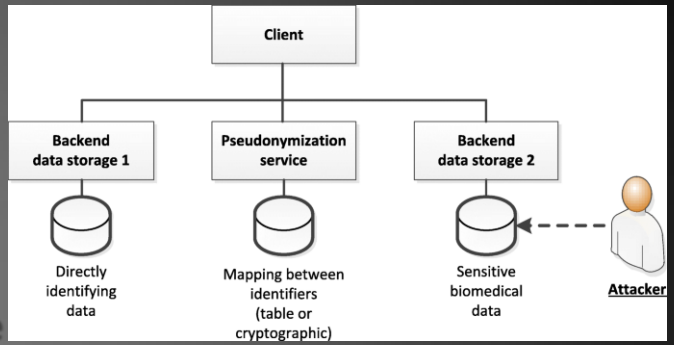
➢ Disadvantages : -

- The process can result in the loss of valuable data.
- May reduce the utility of the data (e.g., detailed analysis may be impossible).
- The anonymisation process is not thorough or if combined with other datasets; This makes it difficult to guarantee complete privacy protection in all cases.

3. Pseudonymisation : -

➢ Advantages : -

- Pseudonymisation offers several key advantages, particularly in balancing data privacy with the ability to retain useful information.
- Reduces the risk of exposure while still allowing for re-identification if necessary.
- One of the main benefits is that it enables the use of sensitive data without directly identifying individuals, while still preserving the ability to re-identify data if necessary, under controlled conditions.



➢ Disadvantages : -

- The process is reversible, so data can still be re-identified if the pseudonymisation key is compromised.
- Requires careful management of the pseudonymisation keys to ensure security.
- Additionally, pseudonymised data can still be vulnerable to re-identification when combined with other datasets or through advanced analytical techniques, especially if the pseudonyms are not sufficiently randomized or protected.

Describe approaches with one scenario

1. **Encryption : -**
   - ➢ **Scenario : - Secure Communication in Online Banking**

- **Use Case : - A user logs into their online banking account, and sensitive data like passwords, account numbers, and transactions are encrypted using TLS (Transport Layer Security).**
- **This ensures that even if a hacker intercepts the data, they cannot read it without the decryption key.**

➢ **Example : -**

- **A visual representation of encrypted communication : -**
- **A sender writes account number “23457889561”**
- **The message is encrypted into unreadable text “X7bnj4f!5s@5s”**
- **The receiver decrypts it back to “23457889561”**

➢ **Best suited for : -**

- **Securing sensitive data in transit, preventing unauthorized access during communication,**

2. Anonymisation : -

- ➢ Scenario : - Traffic Data Analysis

- Use Case : - A city collects GPS locations of cars to analyze traffic patterns.
- Instead of storing actual license plates or owner details, it only records anonymous movement data, ensuring individuals cannot be identified.
- By removing names, addresses, and specific details, the data remains useful for research but cannot be linked to individuals.

- ➢ Example : -

- A city map with multiple anonymous vehicle paths, but no specific identifying information like plate numbers or owner names.

- ➢ Best suited for : -

- Situations where data needs to be analyzed but personal identities must remain completely untraceable.
- Large-Scale Data Sharing Without Privacy Risks.

3. Pseudonymisation : -

- ➢ Scenario : - Online Reviews with Hidden Identities
- Use Case : - A company allows employees to review their workplace anonymously.
- Instead of using real names, employees are assigned random user IDs (e.g., "Reviewer_045") so HR can still identify them if necessary for follow-ups.
- To ensure privacy while allowing HR to follow up if necessary, employees' real names are replaced with pseudonyms like "User_123.“

- ➢ Example : -
- A list of reviews where names are replaced with "User_A12," "User_B34," etc., with a secured database containing the real names separately.

- ➢ Best suited for : -
- Scenarios where data needs to be protected but can be re-identified under strict conditions.
- Protecting Identities While Maintaining Data Usability.

References

(Jonathan Katz Yehuda Lindell, 2020)

Katz, J., & Lindell, Y. pdf

(National Institute of Standards and Technology (NIST) – *Advanced Encryption Standard (AES)*)

NIST AES

(Cynthia Dwork , 2006)

Cynthia Dwork Wikipedia

(The European Data Protection Supervisor (EDPS) – *Anonymisation Techniques)*

EDPS Europe

(Bradley Malin , 2014)

Bradley Malin Wikipedia

(ISO/IEC, 2018 – *Privacy-enhancing data de-identification techniques)*

ISO 2018

(W3schools, 2024)

Data 2024

(Open source, 2024)

Microsoft Copilot in Bing (Just for creating image)

(Stack Overflow)

Stack Overflow - Where Developers Learn, Share, & Build Careers