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
6. Conclusion

Overview of Protecting Personal Data and Its Main Components

➢ In the digital world we live in today, personal data has become one of the most valuable assets. It is also a target (increasing as a target) for cyber attacks, identity theft, and unwarranted monitoring.

➢ Protecting personal data is necessary for privacy and for protecting sensitive information about an individual (i.e., financial records, health records, and forms of personal identification).

❖ Example of Personal Data : -

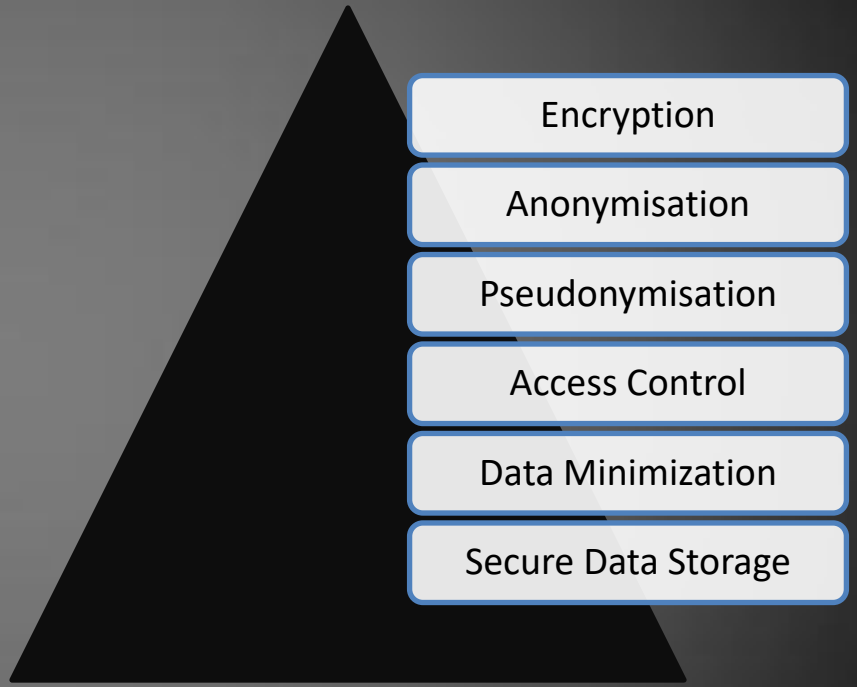
- a name and surname
- a home address
- an email address
- an internet protocol (IP) address
- an identification card number

Different Technical Approaches

There is many technical approaches can be used to protect personal data.

1. Encryption
2. Anonymisation
3. Pseudonymisation

There are three categories we can classify technical approaches to protect personal information and illustrate using example.



Encryption

❑ Explain : -

- Encryption is used to transform data into a code that is unreadable by someone that is not authorized to access that data.
- Using algorithms and encryption keys, encryption takes "plaintext" data and converts it to "ciphertext" data while a decryption key is required to change it back to its original format.

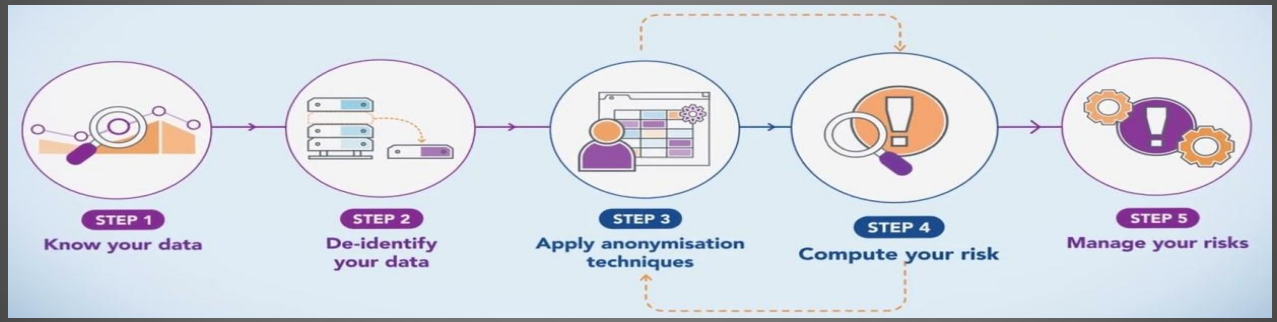
❑ Purpose : -

- Encryption is primarily used to safeguard the confidentiality of data when stored or transmitted. This means that even if data is intercepted or accessed by unauthorized individuals, they are unable to read or utilize it without a valid decryption key.

Anonymisation

❑ Explain : -

- Anonymisation is the act of permanently eliminating personal identifiers from data so that individuals cannot be identified from the data, even when combined with other data.
- This process usually involves removal of direct identifiers (e.g., name, address) and indirect identifiers (e.g., date of birth, geographical location).



❑ Purpose : -

- Anonymisation aims to make it impossible to identify individuals from the data, so that the data can be safely used without privacy concern.
- Permanepresonnalised data has no means of tracing back to any individual.

Pseudonymisation

❑ Explain : -

- Pseudonymisation is the process of replacing identifiable information in a dataset with pseudonyms (i.e., artificial identifiers or codes).
- Unlike anonymisation, pseudonymisation does allow for re-identifying individuals, if deemed necessary, via additional information (e.g., key or lookup table).

❑ Purpose : -

- Pseudonymisation aims to lower the risk of exposure to personal data but keeps the ability to reverse that process and to re-identify a person if required (e.g. legal reason or research).

Compare of the three approaches

Approach	Encryption	Anonymisation	Pseudonymisation
Definition	The method of transforming data into a coded form (ciphertext) readable only with a key.	Removing, or changing, personal identifiers so that a person can't be re-identified.	The procedure of replacing personally identifiable information (PII) in a data set with synthetic identifiers.
Purpose	To protect data confidentiality by rendering it unreadable to those not authorized to read the data.	Privacy through the ability to inhibit identification no longer able to be tracked.	Privacy protection with possible re-identification.
Protection Level	Protects confidentiality, High security.	Strong privacy, very high security where identification is concerned.	Moderate privacy but it may become re-identifiable with the availability to a mapping key.
Data Modification	Data is changed to unreadable form.	Data is changed, data is deleted for the purpose of inhibiting identification.	Data has changed but could be related back if additional information (Key or mapping) is available.
Example Use Case	Protecting sensitive documents, transactions.	Sharing aggregated data in aid of research without identifying individuals.	Medical research, surveys, or de-identified customer data where re-identification could occur but is controlled.

Advantages and Disadvantages

1. Encryption : -

➢ Advantages : -

- Encryption provides many benefits, with the key focus on confidentiality and security of data.
- Encryption protects sensitive information by transforming it to unreadable ciphertext, which means that only a legitimate party with the appropriate decryption key can access the original information.
- The level of protection that encryption provides is vital to protecting sensitive personal, financial, and business information from unwanted access while being stored or transmitted.

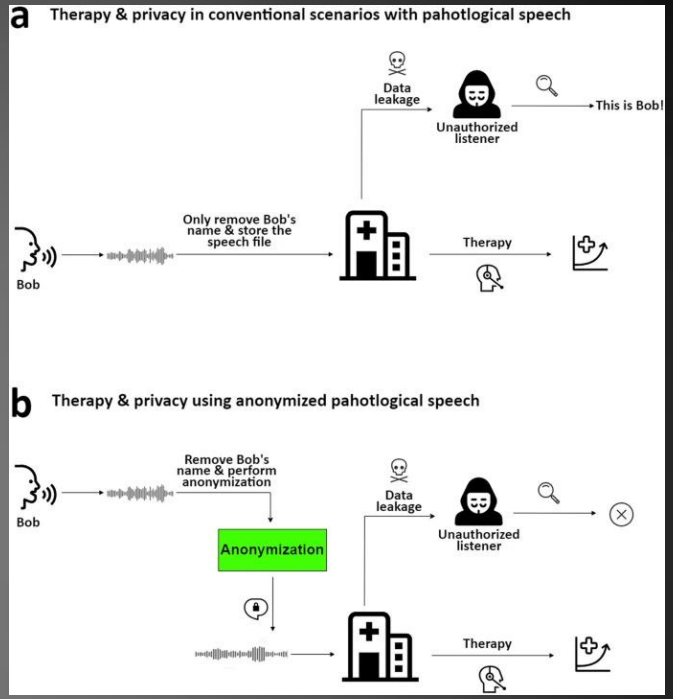
➢ Disadvantages : -

- One of the significant difficulties is how complicated key management is, since the encryption process relies upon the secure storage and handling of encryption keys.
- Encryption can also result in performance overhead since the process of encrypting and decrypting data uses computational resources.

2. Anonymisation : -

➢ Advantages : -

- Offers a great degree of privacy, as they cannot be re-identified.
- Allows organizations to use and share data for analysis, research, and statistical purposes, while protecting the privacy of individuals.
- Anonymisation also reduces the risk from data breaches: if an individual anonymised data is breached, the anonymised data cannot be traced back to any individual, and potential harm is reduced.



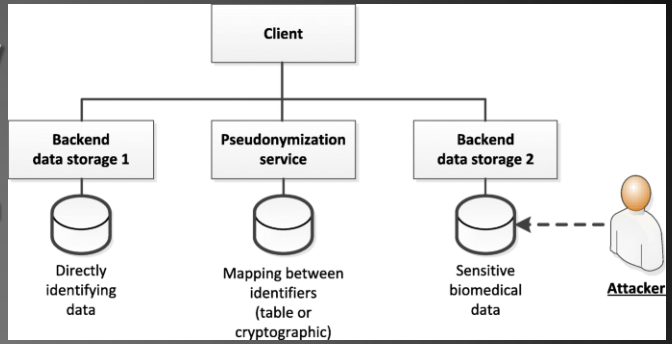
➢ Disadvantages : -

- The process may cause data to be lost, especially the ability to derive value from the data (e.g. intent analysis may not be an option).
- The anonymisation process is weak in terms of thoroughness, or it may be weak if combined with another dataset, There is no way to guarantee complete privacy protection in all situations.

3. Pseudonymisation : -

➢ Advantages : -

- Pseudonymisation has some important advantages, especially when it comes to balancing privacy of data with retaining useful information.
- Lower risk of exposure while still allowing for re-identification if needed.
- One of the main benefits is the ability to use sensitive data without directly identifying an individual, while allowing for the re-identification of the data if needed under controlled conditions.



➢ Disadvantages : -

- The process is reversible, making it possible to re-identify data if the pseudonymisation key is lost.
- Requires management of the pseudonymisation keys to ensure security.
- Additionally, pseudonymised data can still be re-identified by linking to other datasets or other analytical techniques, particularly when the pseudonyms are not sufficiently randomized or protected.

Describe approaches with one scenario

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

- **Use Case : - When a user logs into their online banking account, the user's sensitive information, such as passwords and account numbers are sent with TLS (Transport Layer Security) protection.**
- **This means that if a hacker were to get the data, they could not read it without the decryption key.**

➢ **Example : -**

- **A visual example of encrypted messaging :**
- **A sender types account number “23457889561”**
- **The information is encrypted into unreadable text “X7bnj4f!5s@5s”**
- **The receiver decrypts back with “23457889561”**

➢ **Best suited for : -**

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

2. Anonymisation : -

➢ Scenario : - Traffic Data Analysis

- **Use Case : -** A city is collecting the GPS locations of cars in order to evaluate traffic patterns.
- The city will not store license plates or any ownership information, only the anonymous movements of cars. By storing anonymous movement data, an individual cannot be identified.
- By stripping out the names, addresses, or specifics of the data, it can still be used for research, but won't ever be related to an individual.

➢ Example : -

- A city map with many anonymous vehicle routes, but no identifiable information such as plate numbers or owner names.

➢ Best suited for : -

- Scenarios in which data still needs to be analyzed but the public remains completely untraceable.
- Large-Scale Data Sharing With No Privacy Concern.

3. Pseudonymisation : -

➢ Scenario : - Online Reviews with Hidden Identities

- Use Case : - A company offers their employees a way to review their workplace anonymously.
- Employees use arbitrary user-names (not real names) (example, “Reviewer_045”) to allow the HR department to possibly identify them later if a follow-up is necessary (for example, to investigate an issue or claim).
- Finally, to keep the identity of employees confidential while allowing HR to follow-up if necessary, employees real name is substitute with pseudonym (example, “User_123").

➢ Example : -

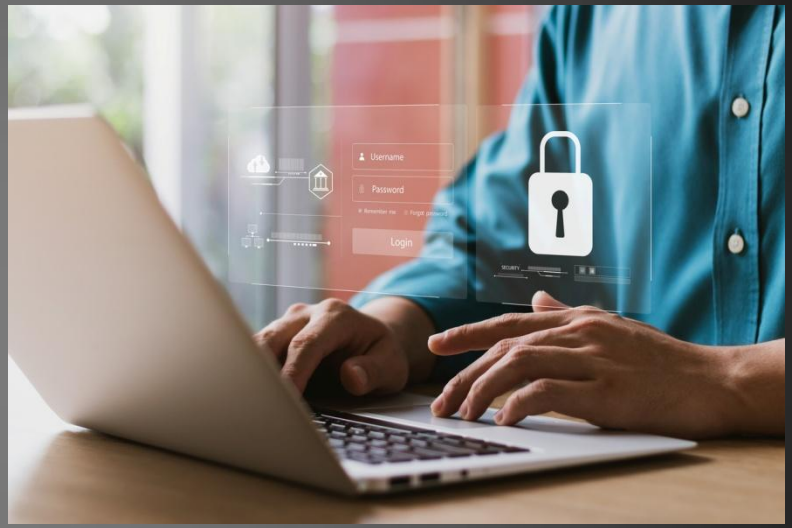
- A list of reviews with the names filled in "User_A12," "User_B34,“ etc., and with the real names stored separately in a secured database.

➢ Best suited for : -

- Scenarios where data can be protected but can be re-identified under strict conditions.
- protecting identities while maintaining usability of data.

Conclusion

- In the digital world, protecting personal data has never been more critical.
- Given the prevalence of privacy risks and security breaches, it is important that individuals make efforts to protect their information (e.g., by using strong passwords, using two-factor authentication, and reducing the amount of information shared through social media or websites) and that organizations implement strong security systems and comply with data protection regulations.



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

(Github)
Github