Τεχνικές και Διαδικασίες Ψευδωνυμοποίησης βάσει του κανονισμού GDPR

Επαμεινώνδας Μπακούλας

Δεκέμβριος 2025

1 Άσκηση 1

1.1 Ερώτημα 1: Εντοπισμός Απλών Προσωπικών Δεδομένων

- **Άμεσα Αναγνωριστικά:**
  - Ονοματεπώνυμο ασθενούς
  - ΑΜΚΑ
  - Τηλέφωνο
- **Έμμεσα Αναγνωριστικά:**
  - Διεύθυνση κατοικίας

1.2 Ερώτημα 2: Εντοπισμός Ευαίσθητων Προσωπικών Δεδομένων

- Αποτελέσματα εξετάσεων (βιοχημικές & αιματολογικές)
- Ιστορικό θεραπειών
- Κλινική διάγνωση

1.3 Ερώτημα 3: Πρόταση Μεθόδου Ψευδωνυμοποίησης

Για την αντικατάσταση των στοιχείων με κωδικούς (π.χ. "PAT001"), μια μέθοδος είναι η χρήση **Μετρητή (Counter)**.

- **Περιγραφή:** Χρησιμοποιείται ένας μονότονος μετρητής που ξεκινά από μια τιμή και αυξάνεται για κάθε νέα εγγραφή. Τα άμεσα αναγνωριστικά αφαιρούνται πλήρως και αντικαθίστανται από τον κωδικό.
- **Εφαρμογή:** Κατά την εξαγωγή των δεδομένων, το σύστημα αντιστοιχίζει στον πρώτο ασθενή τον κωδικό "PAT001", στον δεύτερο το "PAT002" κ.ο.κ. Ο κωδικός αυτός δεν περιέχει κανένα τμήμα των πραγματικών δεδομένων (π.χ. αρχικά ονόματος), ώστε να μην μπορεί να γίνει αναγνώριση του ασθενούς.

1.4 Ερώτημα 4: Ασφαλής Αποθήκευση Πίνακα Αντιστοίχισης

Σύμφωνα με το άρθρο 32 του GDPR και τις βέλτιστες πρακτικές, ο πίνακας αντιστοίχισης (Association Table) είναι το πιο κρίσιμο αρχείο καθώς αποτελεί το "κλειδί" της αποκρυπτογράφησης της ταυτότητας.

- **Πού φυλάσσεται:** Πρέπει να αποθηκεύεται **χωριστά** από τα ψευδωνυμοποιημένα δεδομένα. Δεν πρέπει ποτέ να βρίσκεται στον ίδιο φάκελο, server ή βάση δεδομένων με τα αρχεία που θα σταλούν στο πανεπιστήμιο.
- **Πώς φυλάσσεται:**
  - Πρέπει να είναι **κρυπτογραφημένος** (Encryption).

– Η πρόσβαση πρέπει να περιορίζεται αυστηρά μόνο σε εξουσιοδοτημένο προσωπικό του νοσοκομείου (Access Control).

Table 1: Παράδειγμα Πίνακα Αντιστοίχισης (Φυλάσσεται Ασφαλώς στο Νοσοκομείο)

Ψευδώνυμο	**Ονοματεπώνυμο**	**ΑΜΚΑ**	**Τηλέφωνο**
PAT001	Νώντας Μπακούλας	15028012345	6971234567
PAT002	Μαρία Δημητρίου	22117554321	6947654321

Table 2: Παράδειγμα Πίνακα προς Διαβίβαση (Στέλνεται στο Πανεπιστήμιο)

Ψευδώνυμο	**Αποτελέσματα Εξετάσεων**	**Διάγνωση**	**Ιστορικό Θεραπειών**
PAT001	Κρεατινίνη: 2.8 mg/dL ...	Χρόνια Νεφρική Νόσος	Αιμοκάθαρση ...
PAT002	Κρεατίνη: 3 mg/dL ...	Σακχαρώδης Διαβήτης	Ινσουλίνη ...

1.5 Ερώτημα 5: Βήματα Διαδικασίας, Ρόλοι και Κίνδυνοι

Βήματα Διαδικασίας

1. **Εξαγωγή (Extraction):** Ανάκτηση των δεδομένων από το Πληροφοριακό Σύστημα του Νοσοκομείου.
2. **Διαχωρισμός (Separation):** Διαχωρισμός των άμεσων και έμμεσων αναγνωριστικών (Όνομα, ΑΜΚΑ, Τηλέφωνο, Διεύθυνση) από τα ιατρικά δεδομένα.
3. **Παραγωγή Ψευδωνύμων:** Δημιουργία των κωδικών (PAT001, PAT002) με τη μέθοδο του μετρητή.
4. **Δημιουργία Πίνακα:** Καταγραφή της σύνδεσης (Κωδικός ↔ Πραγματικά Στοιχεία) στον Πίνακα Αντιστοίχισης.
5. **Αποθήκευση Κλειδιού:** Κρυπτογράφηση και αποθήκευση του Πίνακα Αντιστοίχισης σε ξεχωριστό, ασφαλές σημείο.
6. **Διαβίβαση:** Αποστολή στο πανεπιστήμιο *μόνο* του αρχείου με τους κωδικούς και τα ιατρικά αποτελέσματα.

Ρόλοι

- **Data Protection Officer (DPO):** Επιβλέπει τη διαδικασία, συμβουλεύει για την ορθότητα της μεθόδου ψευδωνυμοποίησης και ελέγχει τη συμμόρφωση με τον GDPR.
- **Εργαζόμενοι (IT/Ιατροί):** Το τμήμα Πληροφορικής υλοποιεί την τεχνική διαδικασία (εξαγωγή, κωδικοποίηση), ενώ οι ιατροί ενδέχεται να επιβεβαιώσουν την εγκυρότητα των κλινικών δεδομένων πριν την ανωνυμοποίηση.

Μείωση Κινδύνου & Περιορισμοί

- **Μείωση Κινδύνου:** Η ψευδωνυμοποίηση μειώνει τον κίνδυνο διότι, εάν τα δεδομένα υποκλαπούν κατά τη μεταφορά ή από το πανεπιστήμιο, είναι άχρηστα για τον επιτιθέμενο χωρίς τον πίνακα αντιστοίχισης. Δεν μπορούν να συνδεθούν άμεσα με φυσικά πρόσωπα.
- **Πότε δεν επαρκεί:** Δεν είναι επαρκής όταν είναι δυνατή η **inference attack**. Αυτό συμβαίνει όταν σπάνιοι συνδυασμοί χαρακτηριστικών στα δεδομένα (π.χ. μια πολύ σπάνια ασθένεια σε συνδυασμό με ηλικία και περιοχή) επιτρέπουν σε κάποιον να αναγνωρίσει τον ασθενή, ακόμη και χωρίς το όνομά του, χρησιμοποιώντας εξωτερικές πηγές πληροφοριών.

2 Άσκηση 2

Ο κώδικας που υλοποιεί την ψευδωνυμοποίηση παρουσιάζεται παρακάτω:

```
import pandas as pd

# Δεδομένα Εισόδου
customers = [
    {
        "name": "΄Αννα",
        "surname": "Παπαδοπούλου",
        "email": "anna@example.com",
        "age": 28,
        "profession": "Ιατρός",
    },
    {
        "name": "Κώστας",
        "surname": "Νικολάου",
        "email": "kostas@example.com",
        "age": 35,
        "profession": "Μηχανικός Η/Υ",
    },
    {
        "name": "Ιωάννα",
        "surname": "Γεωργίου",
        "email": "ioanna@example.com",
        "age": 22,
        "profession": "Καθηγήτρια",
    },
]

# Λίστες για τη διατήρηση των διαχωρισμένων δεδομένων
pseudonymized_data = []
mapping_table = []

# Διαδικασία Ψευδωνυμοποίησης για κάθε πελάτη
for index, person in enumerate(customers, start=1):
    # Δημιουργία του ψευδωνύμου (π.χ., USER1, USER2)
    user_id = f"USER{index}"

    # 1. Δημιουργία της Ψευδωνυμοποιημένης Εγγραφής (Δεδομένα προς κοινοποίηση)
    # Αφαίρεση ονόματος, επωνύμου και email. Διατήρηση Ηλικίας και Επαγγέλματος.
    pseudo_record = {
        "User_ID": user_id,
        "Age": person["age"],
        "Profession": person["profession"],
    }
    pseudonymized_data.append(pseudo_record)

    # 2. Δημιουργία της Εγγραφής Αντιστοίχισης (Το κλειδί για επαναταυτοποίηση)
    # Σύνδεση του ψευδωνύμου με τα πραγματικά αναγνωριστικά.
    map_record = {
        "User_ID": user_id,
        "Name": person["name"],
        "Surname": person["surname"],
        "Email": person["email"],
    }
    mapping_table.append(map_record)

# 3.1 Εξαγωγή πίνακα προς κοινοποίηση
print("--- 1. Ψευδωνυμοποιημένα Δεδομένα (Ασφαλή για Κοινοποίηση) ---")
df_pseudo = pd.DataFrame(pseudonymized_data)
print(df_pseudo.to_string(index=False))
```

```
print("\n")

# 3.2 Εξαγωγή πίνακα αντιστοίχισης
print("--- 2. Πίνακας Αντιστοίχισης (Πρέπει να αποθηκεύεται με ασφάλεια και ξεχωριστά) ---")
df_map = pd.DataFrame(mapping_table)
print(df_map.to_string(index=False))

# 4. Γιατί η ψευδωνυμοποίηση βοηθά στην προστασία των προσωπικών δεδομένων,
# σύμφωνα με τον GDPR
"""
Η ψευδωνυμοποίηση αντικαθιστά τα άμεσα αναγνωριστικά (όπως ονόματα) με τεχνητούς κωδικούς.
Αυτό βοηθά στην προστασία των προσωπικών δεδομένων επειδή τα δεδομένα δεν μπορούν
να αποδοθούν σε ένα συγκεκριμένο άτομο χωρίς τη χρήση πρόσθετων πληροφοριών
(του πίνακα αντιστοίχισης). Σύμφωνα με το GDPR, εφόσον ο πίνακας αντιστοίχισης
διατηρείται ξεχωριστά και με ασφάλεια, ο κίνδυνος για τα υποκείμενα των δεδομένων
μειώνεται σημαντικά σε περίπτωση παραβίασης δεδομένων.
"""
```

2.1 Αποτελέσματα Εκτέλεσης

Αν κατεβάσουμε την βιβλιοθήκη pandas και εκτελέσουμε τον παραπάνω κώδικα, θα λάβουμε τα εξής αποτελέσματα:

```
--- 1. Ψευδωνυμοποιημένα Δεδομένα (Ασφαλή για Κοινοποίηση) ---
User_ID  Age    Profession
  USER1   28        Ιατρός
  USER2   35 Μηχανικός Η/Υ
  USER3   22    Καθηγήτρια


--- 2. Πίνακας Αντιστοίχισης (Πρέπει να αποθηκεύεται με ασφάλεια και ξεχωριστά) ---
User_ID   Name      Surname              Email
  USER1   ΄Αννα Παπαδοπούλου   anna@example.com
  USER2 Κώστας     Νικολάου kostas@example.com
  USER3 Ιωάννα     Γεωργίου ioanna@example.com
```