THE EU CYBERSECURITY AGENCY

ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗ ΣΤΟ ΠΛΑΙΣΙΟ ΤΟΥ ΓΚΠΔ: ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΤΕΧΝΙΚΕΣ

Αθηνά Μπούρκα,
Ευρωπαϊκός Οργανισμός για την Κυβερνοασφάλεια (ENISA)

13 | 12 | 2019

ΠΕΡΙΕΧΟΜΕΝΑ


enisa
GDPR
Recommendations on shaping technology according to GDPR provisions
An overview on data pseudonymisation
NOVEMBER 2018
www.enisa.europa.eu
European Union Agency For Network and Information Security


enisa
EUROPEAN UNION AGENCY FOR CYBERSECURITY
Pseudonymisation techniques and best practices
Recommendations on shaping technology according to data protection and privacy provisions
NOVEMBER 2019

Η ΕΝΝΟΙΑ ΤΗΣ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

ΤΙ ΕΙΝΑΙ Η ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗ;

«Αποσύνδεση» ή «διαχωρισμός» της ταυτότητας των υποκειμένων των δεδομένων από τα υπόλοιπα δεδομένα.

ISO/TS 25237:2017
Health Informatics — Pseudonymization

- Ένας ιδιαίτερος τύπος **απο-ταυτοποίησης** (de-identification) που αφαιρεί τη σύνδεση με το υποκείμενο και προσθέτει μια συσχέτιση με ένα ειδικό σύνολο χαρακτηριστικών που σχετίζεται με το υποκείμενο και ένα ή περισσότερα **ψευδώνυμα**.
- Ψευδώνυμο (pseudonym): αντικαθιστά τα συνήθη **αναγνωριστικά** (identifiers) και συνδέει τα ψευδώνυμα δεδομένα με το υποκείμενο χωρίς αποκάλυψη της ταυτότητας.

ΤΙ ΕΙΝΑΙ Η ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗ;

ΓΚΠΔ (αρ. 4(5))

Η επεξεργασία δεδομένων προσωπικού χαρακτήρα κατά τρόπο ώστε τα δεδομένα να μην μπορούν πλέον να αποδοθούν σε συγκεκριμένο υποκείμενο των δεδομένων **χωρίς τη χρήση συμπληρωματικών πληροφοριών**, εφόσον οι εν λόγω συμπληρωματικές πληροφορίες διατηρούνται χωριστά και υπόκεινται σε τεχνικά και οργανωτικά μέτρα προκειμένου να διασφαλιστεί ότι (τα δεδομένα) δεν μπορούν να αποδοθούν σε **ταυτοποιημένο ή ταυτοποιήσιμο** φυσικό πρόσωπο.

ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗ ≠ ΑΝΩΝΥΜΟΠΟΙΗΣΗ

- Ψευδώνυμα δεδομένα = προσωπικά δεδόμενα.
- Ανώνυμα δεδομένα: δεν επιτρέπουν με κανένα δυνατό τρόπο τον προσδριορισμό της ταυτότητας (≠ προσωπικά δεδομένα).
- **Συνήθης σύγχηση:** «Αν αφαιρεθούν στοιχεία άμεσης ταυτοποίησης (π.χ. όνομα, διεύθυνση, ΑΜΚΑ, κλπ.), τα δεδομένα είναι ανώνυμα..».

ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗ ≠ ΑΝΩΝΥΜΟΠΟΙΗΣΗ

The New York Times

TECHNOLOGY

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr. AUG. 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."

Researchers reverse Netflix anonymization

Robert Lemos, SecurityFocus 2007-12-04

In a dramatic demonstration of the privacy dangers of databases that collect consumer habits, two researchers from the University of Texas at Austin have shown that a handful of movie ratings can identify a person as easily as a Social Security number.

The researchers -- graduate student Arvind Narayanan and professor Vitaly Shmatikov, both from the Department of Computer Sciences at the University of Texas at Austin -- claim to have identified two people out of the nearly half million anonymized users whose movie ratings were released by online rental company Netflix last year. The company published the large database as part of its $1 million Netflix Prize, a challenge to the world's researchers to improve the rental firm's movie-recommendation engine.

"Releasing the data and just removing the names does nothing for privacy," Shmatikov told SecurityFocus. "If you know their name and a few records, then you can identify that person in the other (private) database."

" Releasing the data and just removing the names does nothing for privacy. If you know their name and a few records, then you can identify that person in the other (private) database. "

Vitaly Shmatikov, Professor of Computer Science, University of Texas at Austin

Search International edition

The Guardian

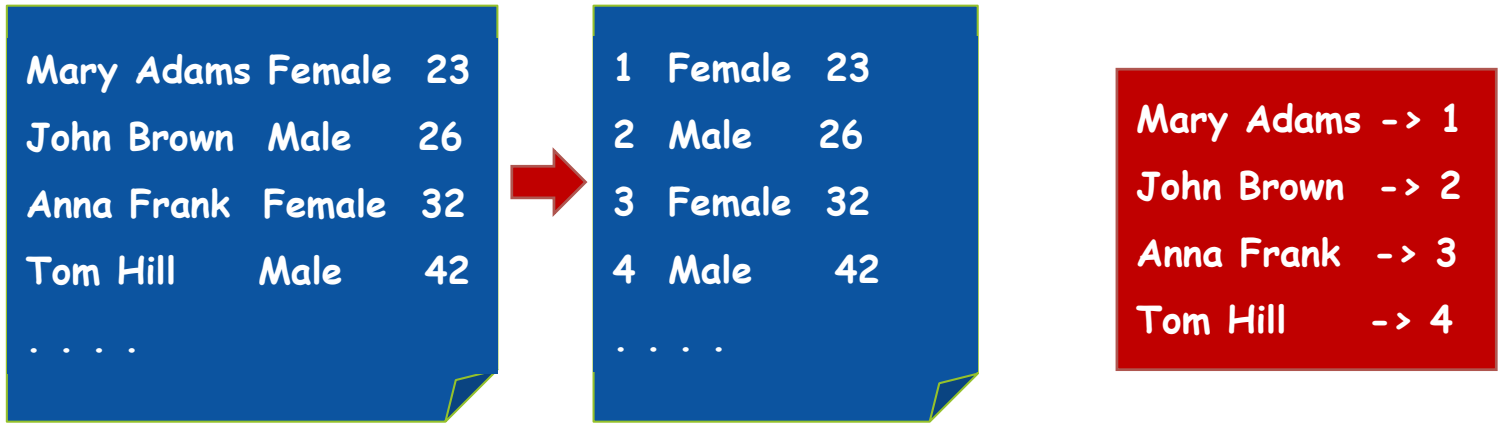
New York taxi details can be extracted from anonymised data, researchers say

FoI request reveals data on 173m individual trips in US city - but could yield more details, such as drivers' addresses and income

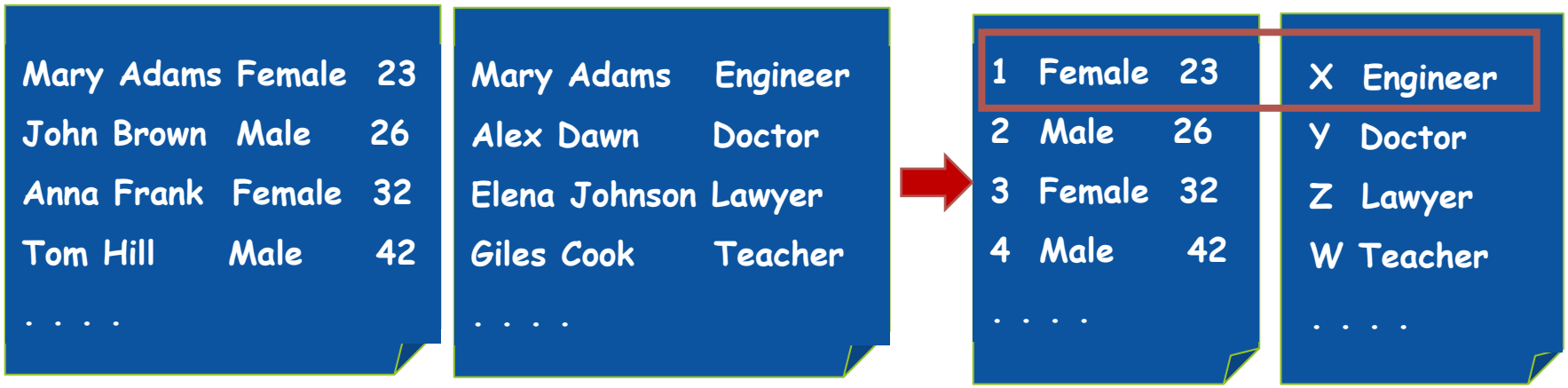
▲ Data about New York city taxi drivers and rides could be de-anonymised, researchers warn. Photograph: Jan Johannessen/Getty Images

ΤΑ ΟΦΕΛΗ ΤΗΣ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

1. Απόκρυψη ταυτότητας σε ένα αρχείο δεδομένων



2. Απο-σύνδεση μεταξύ διαφορετικών αρχείων δεδομένων (unlinkability)



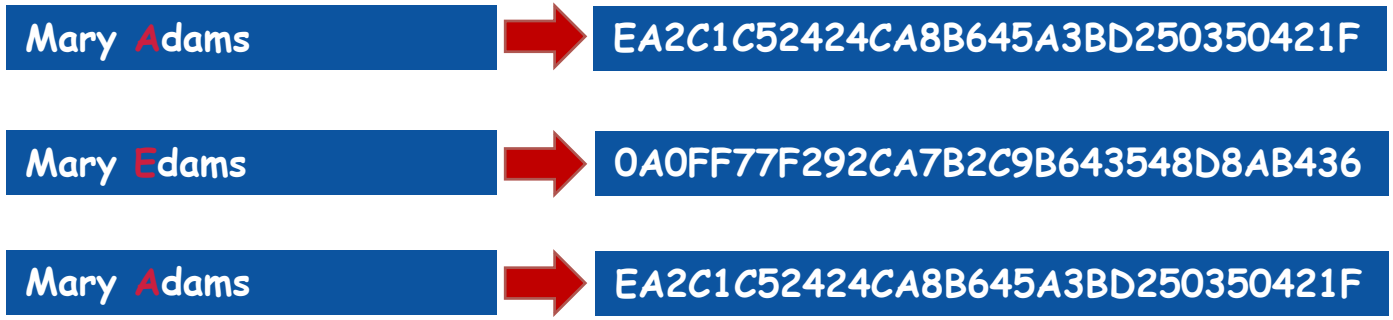
ΤΑ ΟΦΕΛΗ ΤΗΣ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

3. Ελαχιστοποίηση των δεδομένων (data minimisation)

Παράδειγμα:

Χρήση ψευδώνυμων αναγνωριστικών για την παρακολούθηση (tracking) από τον υπεύθυνο χωρίς πρόσβαση στο αρχικό αναγνωριστικό του χρήστη

4. Επαλήθευση της ακρίβειας των δεδομένων



Ο ΡΟΛΟΣ ΤΗΣ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ ΣΤΟΝ ΓΚΠΔ

- Μέτρο προστασίας των δεδομένων ήδη από τον σχεδιασμό (data protection by design) – αρ. 25(1) ΓΚΠΔ.
- Τεχνικό και οργανωτικό μέτρο για την ασφάλεια των δεδομένων – αρ. 32(1) ΓΚΠΔ.
- Δυνατή προϋπόθεση για περαιτέρω επεξεργασία για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον ή σκοπούς επιστημονικής ή ιστορικής έρευνας ή στατιστικούς σκοπούς - αρ. 5(1)(β) και 89(1) ΓΚΠΔ.
- Κριτήριο για εξακρίβωση της συμβατότητας του σκοπού επεξεργασίας – αρ. 6(4) ΓΚΠΔ.
- Κριτήριο εκτίμησης του κινδύνου σε παραβίαση προσωπικών δεδομένων (Ο.Ε. αρ. 29).

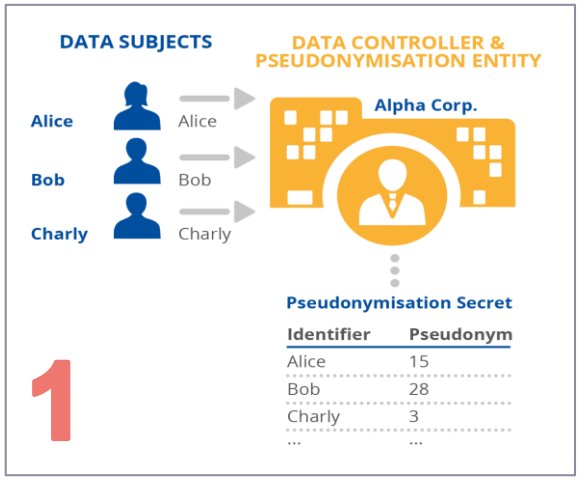
ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗ ΣΤΗΝ ΠΡΑΞΗ

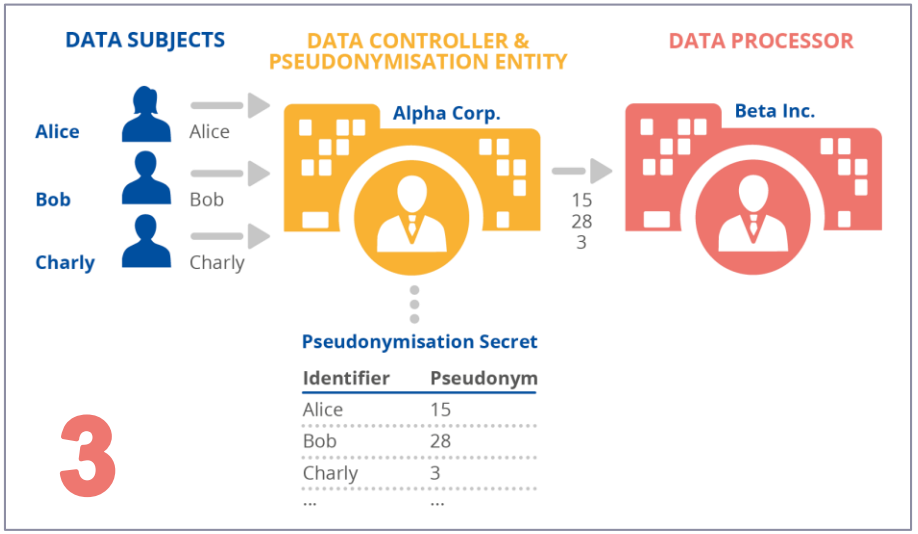
ΒΑΣΙΚΑ ΕΡΩΤΗΜΑΤΑ

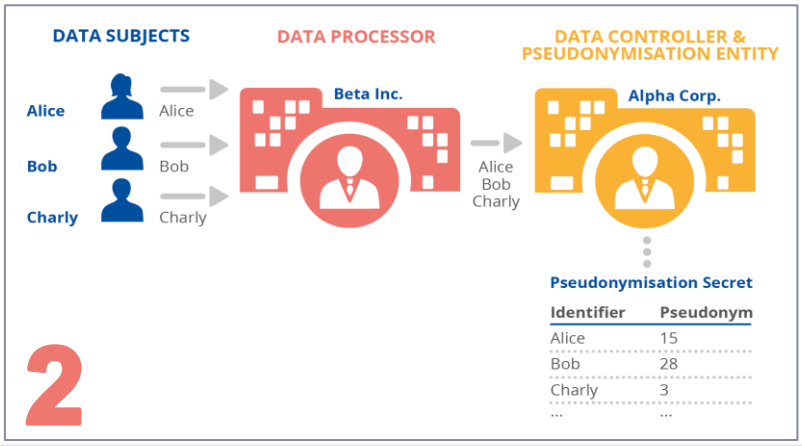
- Ποιος έχει την ευθύνη της ψευδωνυμοποίησης;
- Ποιος πραγματοποιεί την ψευδωνυμοποίηση στην πράξη;
- Για ποιο σκοπό;
- Με ποιόν τρόπο; (τεχνικό, οργανωτικό)
- Πόσο «ψευδώνυμα» χρειάζεται/πρέπει να είναι τα δεδομένα;

ΣΕΝΑΡΙΑ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

Ψευδωνυμοποίηση από τον υπεύθυνο

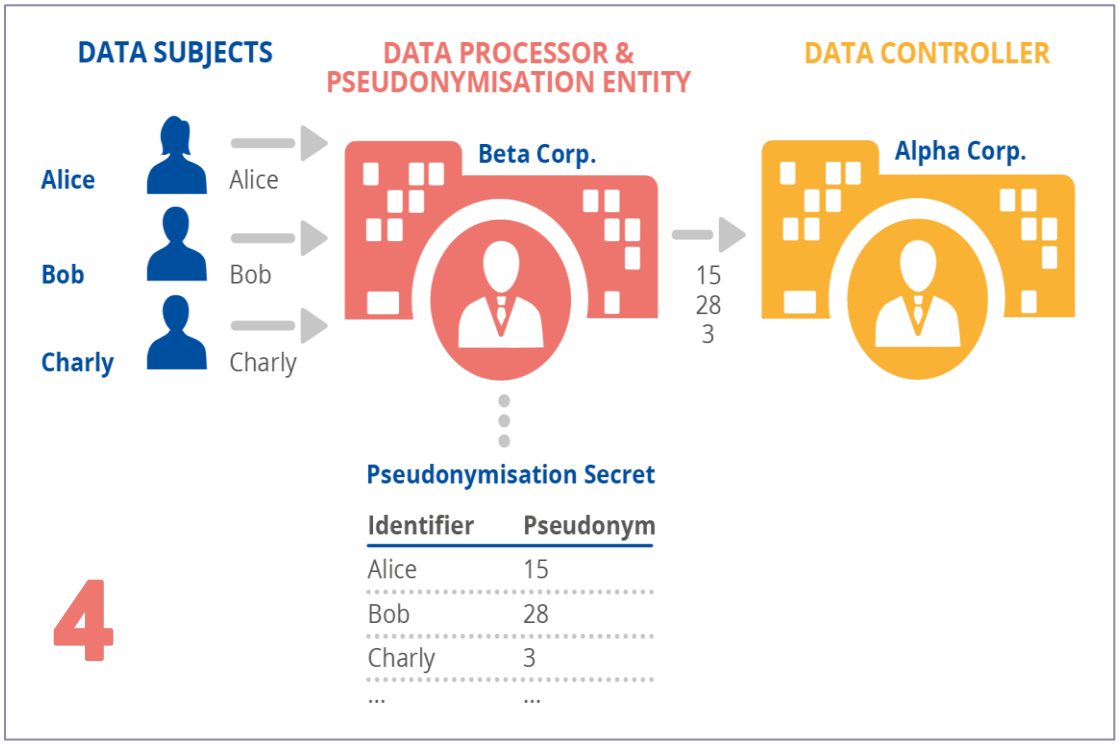






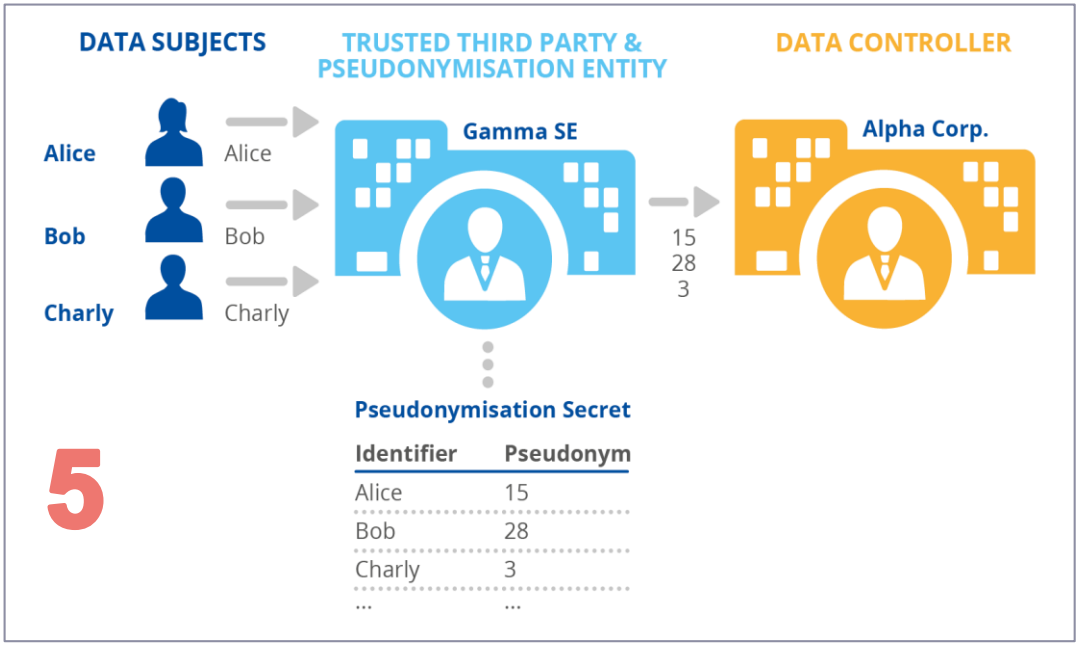
ΣΕΝΑΡΙΑ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

Ψευδωνυμοποίηση από εκτελούντα



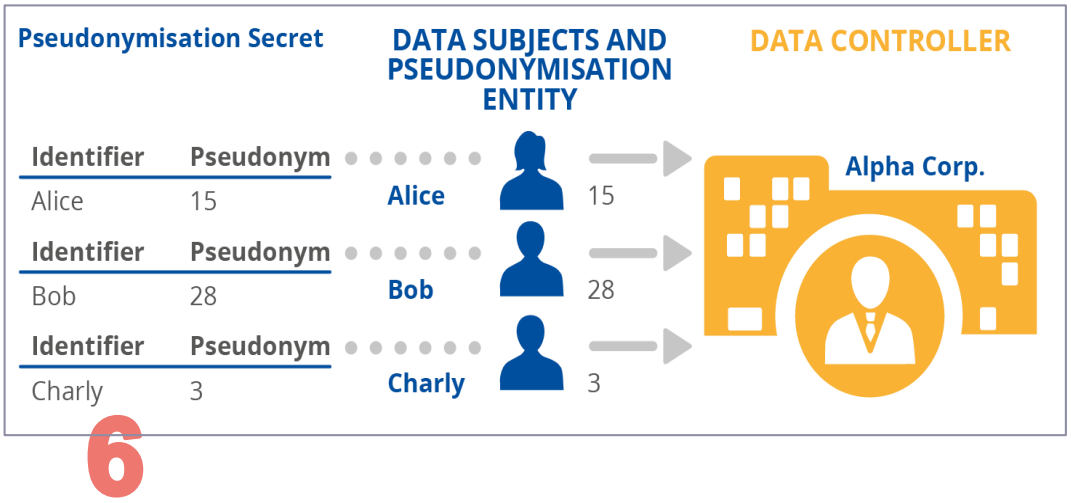
ΣΕΝΑΡΙΑ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

Ψευδωνυμοποίηση από έμπιστο τρίτο μέρος (συνυπεύθυνο?)



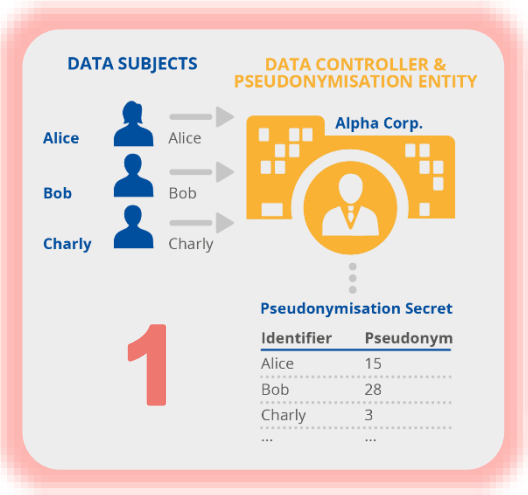
ΣΕΝΑΡΙΑ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

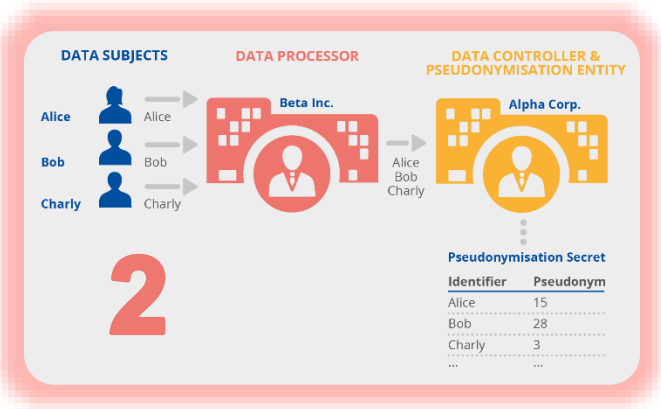
Ψευδωνυμοποίηση από τα υποκείμενα των δεδομένων

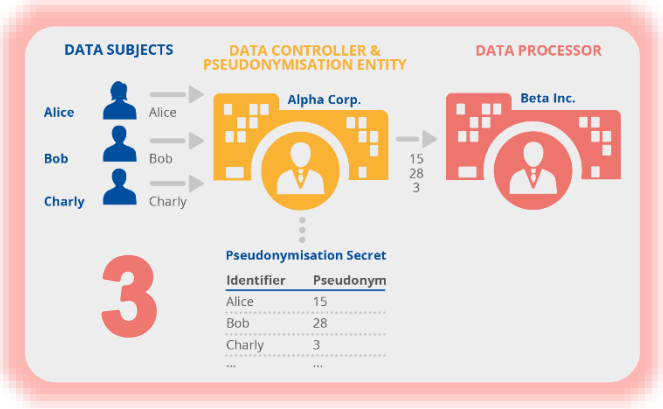


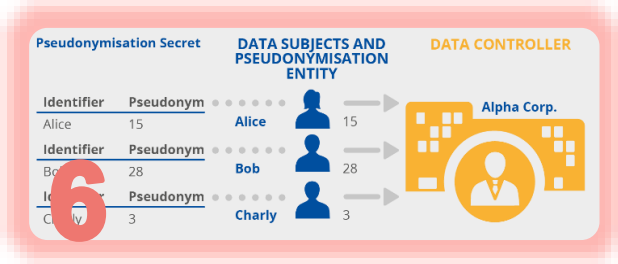
ΕΠΙΘΕΣΕΙΣ ΣΤΗΝ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗ

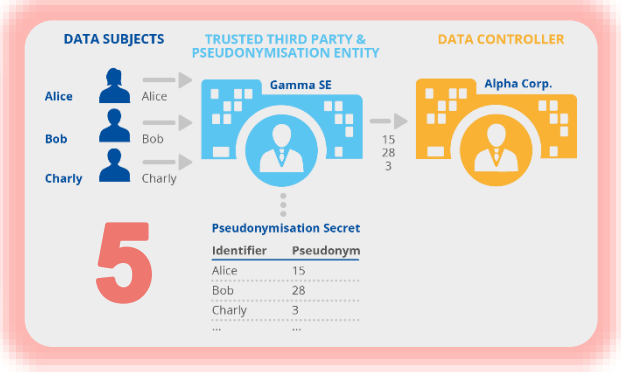
- *Ποια είναι τα κακόβουλα μέρη;*
- *Ποιες είναι οι τεχνικές επίθεσης;*

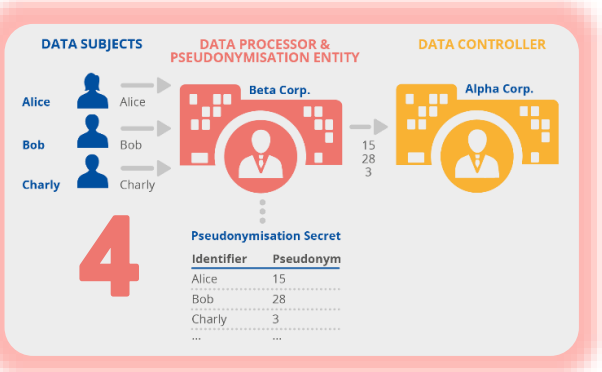












ΤΕΧΝΙΚΕΣ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

- Μετρητής (Counter).
- Γεννήτρια τυχαίων αριθμών (RNG).
- Συνάρτηση κατακερματισμού (χωρίς κλειδί) – Hash function.
- Συνάρτηση κατακερματισμού με κλειδί (HMAC).
- Συμμετρική κρυπτογράφηση.
- Προηγμένες τεχνικές.

ΤΕΧΝΙΚΕΣ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

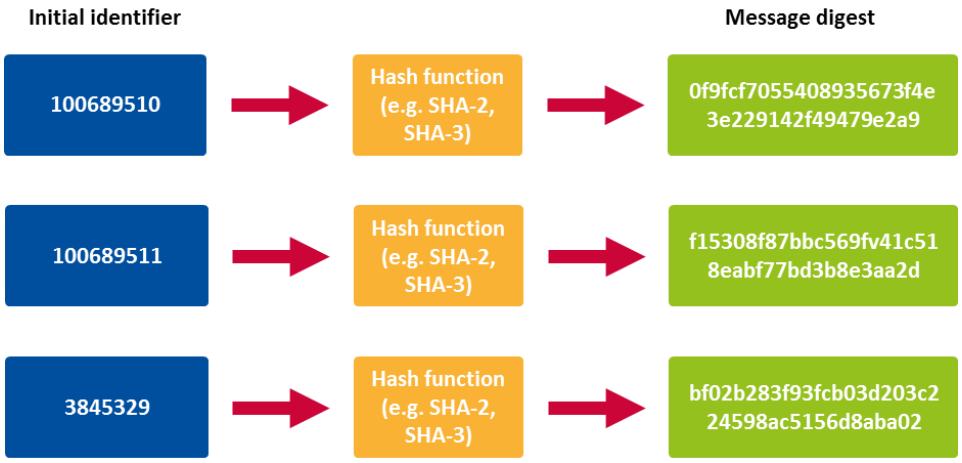
1. Μετρητής (Counter) – RNG (Γεννήτρια τυχαίων αριθμών)

E-mail address	Pseudonym (Random number generator)	Pseudonym (counter generator)
alice@abc.eu	328	10
bob@wxyz.com	105	11
eve@abc.eu	209	12
john@qed.edu	83	13
alice@wxyz.com	512	14
mary@clm.eu	289	15

+ Απλότητα
+ Ισχυρή προστασία κατά επιθέσεων (RNG)
- Δυσκολία πρακτικής υλοποίησης (π.χ. μεγάλες βάσεις δεδομένων)
- Counter: διαδοχικότητα (μπορεί να αποκαλύψει δεδομένα)

ΤΕΧΝΙΚΕΣ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

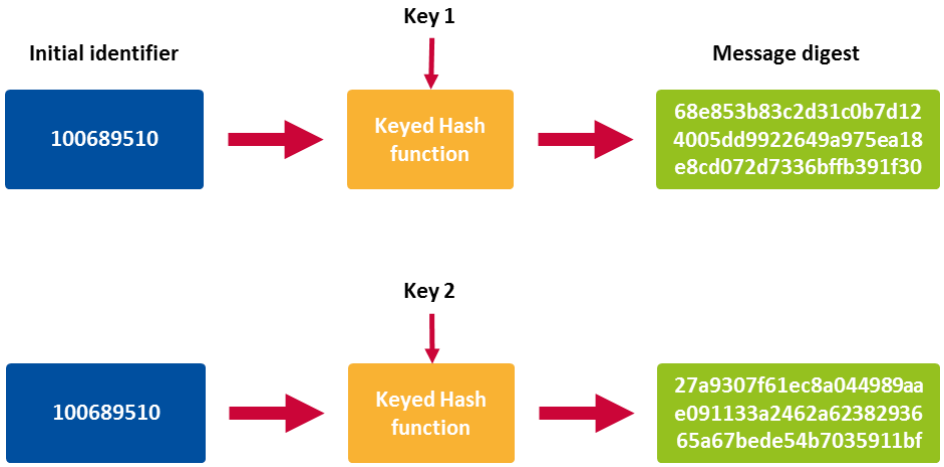
2. Συνάρτηση κατακερματισμού (χωρίς κλειδί)



+ Επαλήθευση ακεραιότητας
- Ιδιαίτερα επιρρεπής σε επιθέσεις (brute force attacks): **ΑΔΥΝΑΜΗ ΤΕΧΝΙΚΗ**

ΤΕΧΝΙΚΕΣ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

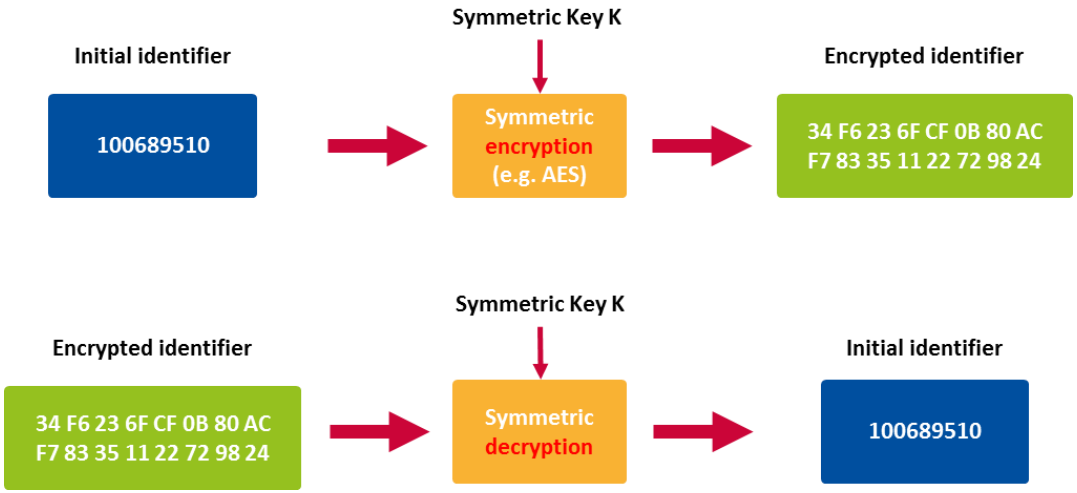
3. Συνάρτηση κατακερματισμού με κλειδί (HMAC)



+ Ισχυρή προστασία κατά επιθέσεων ψευδωνυμοποίησης
+ Πρακτική υλοποίηση - κλιμάκωση
- Επαναφορά (recovery)

ΤΕΧΝΙΚΕΣ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

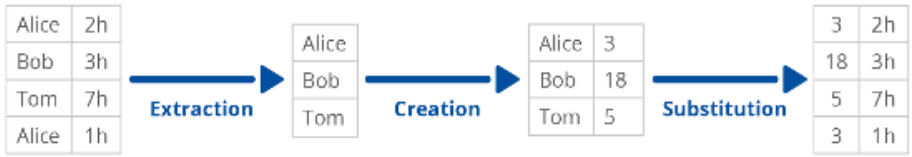
4. Κρυπτογράφηση



+ Ισχυρή προστασία κατά επιθέσεων ψευδωνυμοποίησης
+ Πρακτική υλοποίηση – κλιμάκωση - επαναφορά
- Ο υπεύθυνος πρέπει να έχει πάντα πρόσβαση στα αρχικά αναγνωριστικά

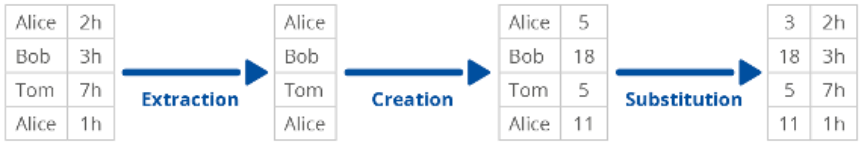
ΠΟΛΙΤΙΚΕΣ ΨΕΥΔΩΝΥΜΟΠΟΙΗΣΗΣ

1. Ντετερμινιστική (deterministic) ψευδωνυμοποίηση



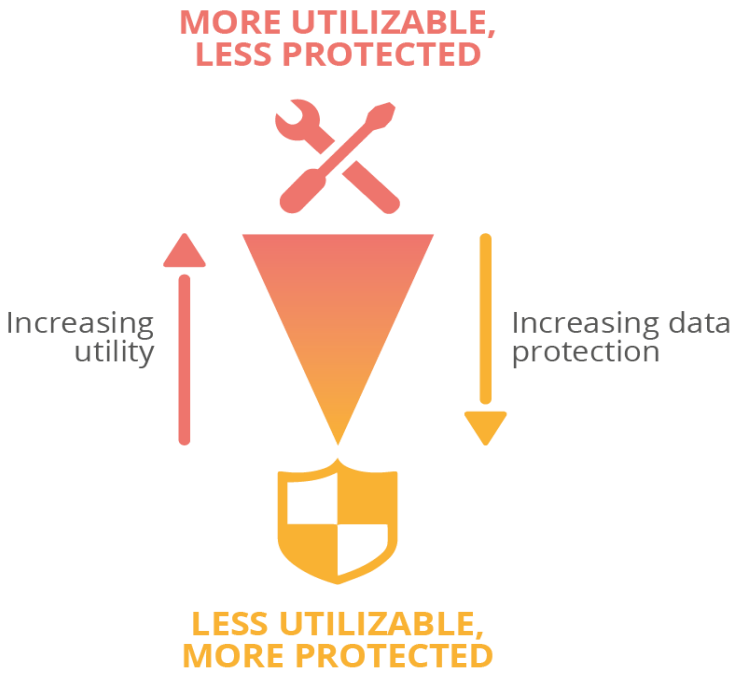
+ Χρηστικότητα των ψευδώνυμων δεδομένων (π.χ. εξαγωγή στατιστικών).
- Χαμηλότερο επίπεδο προστασίας από επιθέσεις

2. Πιθανοτική (randomized) ψευδωνυμοποίηση



+ Υψηλότερο επίπεδο προστασίας από επιθέσεις
- Χαμηλότερη χρηστικότητα των ψευδώνυμων δεδομένων

ΠΡΟΣΤΑΣΙΑ ΚΑΙ ΧΡΗΣΤΙΚΟΤΗΤΑ (UTILITY)



alice@abc.eu →

3281051

328@1051

328@1051.3

328@abc.eu

Each MAC computation utilizes the same (secret) key

alice@abc.eu
MAC(alice) MAC(abc.eu)

eve@abc.eu
MAC(eve) MAC(abc.eu)

ΠΩΣ ΝΑ ΔΙΑΛΕΞΩ ΤΗΝ ΚΑΛΥΤΕΡΗ ΤΕΧΝΙΚΗ?

Σενάριο – εκτίμηση κινδύνου

- Προστασία δεδομένων.
- Απαιτούμενη χρηστικότητα.
- Πρακτική υλοποίηση – κλιμάκωση.
- Ευελιξία – δυνατότητα επαναφοράς αρχικών δεδομένων.

- Συνδυασμός τεχνικών/πολιτικών
- Προηγμένες μέθοδοι (π.χ. βάσει τεχνικών ανωνυμοποίησης)

ΣΥΜΠΕΡΑΣΜΑΤΑ

- Δεν υπάρχει βέλτιστη τεχνική ψευδωνυμοποίησης για όλες τις περιπτώσεις.
- Απαραίτητη η εκτίμηση του κινδύνου και των απαιτήσεων της επεξεργασίας.
- Περαιτέρω ανάπτυξη τεχνικών και συνδυασμών τους.
- Περαιτέρω προβολή παραδειγμάτων και περιπτώσεων χρήσης στην πράξη.

ΕΥΧΑΡΙΣΤΩ ΓΙΑ ΤΗΝ ΠΡΟΣΟΧΗ ΣΑΣ!

Βασ. Σοφίας 1, Μαρούσι 151 24
Αττική, Ελλάδα

+30 28 14 40 9711

info@enisa.europa.eu

www.enisa.europa.eu