Η ψευδωνυμοποίηση στον Γενικό Κανονισμό Προστασίας Δεδομένων

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

Δρ. Κωνσταντίνος Λιμνιώτης
Ε.Ε.Π., Πληροφορικός
klimniotis at dpa.gr

Time to adopt PETs

ENISA - UNIPI Workshop on Privacy Enchancing Technologies
November 22nd, 2018

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Επισκόπηση παρουσίασης

- Εισαγωγή - Βασικές έννοιες
  - Παραδείγματα ανεπαρκούς ανωνυμοποίησης
  - Η εποχή των «μεγάλων δεδομένων» (Big Data)
- Νέος Κανονισμός (ΕΕ) 679/2016 (GDPR)
  - Προσωπικά δεδομένα – Ανώνυμα δεδομένα
  - Ψευδωνυμοποιημένα δεδομένα
- Αναλύοντας την έννοια της ψευδωνυμοποίησης
  - Διαχείριση αναγνωριστικών και ψευδωνύμων
  - Πλεονεκτήματα και μειονεκτήματα γνωστών τεχνικών
- Συμπεράσματα

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Ανώνυμα δεδομένα: το συχνό λάθος

- Το (συχνό) λάθος: Αν δεν είναι «καταφανές» σε ποιον αναφέρονται τα δεδομένα, τότε είναι ανώνυμα

- Το σωστό: Ακόμα και αν δεν είναι προφανής η ταυτότητα του ανθρώπου στον οποίο αναφέρονται τα δεδομένα, πρέπει – προτού χαρακτηριστούν ως ανώνυμα – να εξεταστεί ενδελεχώς αν όντως έχει «εκμηδενιστεί» η δυνατότητα ανακάλυψης της ταυτότητάς του

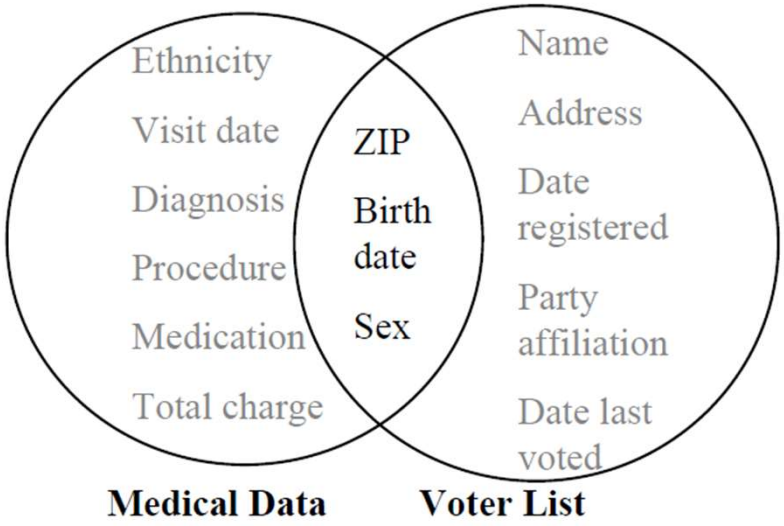
- Οι «λανθασμένες» ανωνυμοποιήσεις είναι μία ιστορία παλιά....

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Επίθεση συσχετίσεων (Linking Attack)

- **[Sweeney, 2002]** : Αντιπαραβολή της λίστας ψηφοφόρων με την (ανωνυμοποιημένη) λίστα νοσηλευομένων δημόσιου νοσοκομείου



- Για μία συγκεκριμένη ημ/νία γέννησης, έξι άτομα είχαν την ίδια,
  - Τρεις εξ αυτών άντρες, μόνο ένας με τον ίδιο ταχυδρομικό κώδικα (ZIP code)
  - Αυτός ήταν ο (τότε) κυβερνήτης της Μασαχουσέτης
- Σύμφωνα με αυτήν την έρευνα, το 87% του πληθυσμού των Η.Π.Α. μπορεί να ταυτοποιηθεί από την τριπλέτα «Ταχ. Κώδικας - ημερομηνία γέννησης – φύλο»

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Το περιστατικό της AOL (2006)

Αύγουστος 2006: research.aol.com

AOL is embarking on a new direction for its business making its content and products freely available to all consumers. To support those goals, AOL is also embracing the vision of an open research community. To get started, we invite you to visit us at http://research.aol.com, where you will find:

- …
- ***Query streams for 500,000 users over 3 months (20 million queries)***
- ….

- Δεν υπήρχαν στη σχετική λίστα των χρηστών άμεσα στοιχεία ταυτοποίησής τους
  - Ένα τυχαίο ID για τον κάθε χρήστη

- Όμως, συνδυάζοντας κανείς τις αναζητήσεις ενός χρήστη (βάσει του ID αυτού) με λοιπές πληροφορίες (π.χ. τηλεφωνικού καταλόγου), μπορούσε τελικά να ταυτοποιήσει το χρήστη αυτόν!!
  - Η εφημερίδα New York Times ταυτοποίησε πολλούς χρήστες
    - Κάποιους εξ αυτών, με την άδειά τους, δημοσιοποίησε τα στοιχεία τους για να καταδείξει το περιστατικό.

Time to adopt PETs

Η περίπτωση του χρήστη #4417749

HOME PAGE | MY TIMES | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS

The New York Times **Technology**

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

CAMCORDERS | CAMERAS | CELLPHONES | COMPUTERS | HANDHELDS | HOME VIDEO | MUSIC | PERIPHERALS

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

SIGN IN TO E-MAIL THIS
PRINT
SINGLE PAGE
REPRINTS
SAVE
ARTICLE TOOLS SPONSORED BY

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.

Erik S. Lesser for The New York Times
Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem.

Multimedia

Graphic: What Revealing Search Data Reveals

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."

It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.

AOL removed the search data from its site over the weekend and apologized for its release, saying it was an unauthorized move by a team that had hoped it would benefit academic researchers.

But the detailed records of searches conducted by Ms. Arnold and 657,000 other Americans, copies of which continue to circulate online, underscore how much people unintentionally reveal about themselves when they use search engines — and how risky it

ΑΡΧΗ ΠΡΟΣΤΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Το περιστατικό της Netflix

- **2006:** Η **Netflix** δημοσιοποίησε τις αξιολογήσεις που έκαναν οι εγγεγραμμένοι σε αυτή χρήστες σε ταινίες
  - Κάθε στοιχείο ταυτοποίησής τους είχε απομακρυνθεί
- Ένα χρόνο μετά (2007), οι ερευνητές **Narayanan and Shmatikov** ταυτοποίησαν σημαντικό ποσοστό των χρηστών της **Netflix**, με βάση τις (δημόσια προσβάσιμες) αξιολογήσεις σε ταινίες που έκαναν στην πλατφόρμα **IMDB**

  "Given a user's public IMDb ratings, which the user posted voluntarily to selectively reveal some of his (or her; but we'll use the male pronoun without loss of generality) movie likes and dislikes, we discover all the ratings that he entered privately into the Netflix system, presumably expecting that they will remain private"
- Με την ταυτοποίηση/συσχέτιση, αποκαλύφθηκαν και ευαίσθητα δεδομένα βάσει συγκεκριμένων αξιολογήσεων ταινιών που έγιναν στη **Netflix** (με την προσδοκία ότι δεν θα δημοσιοποιηθούν).
  - *Π.χ. "Power and Terror: Noam Chomsky in Our Times", "Fahrenheit 9/11", "Jesus of Nazareth"*

Πότε μπορεί να αναγνωριστεί η ταυτότητα ενός προσώπου;

- Εκτός από τα αναγνωριστικά (identifiers), υπάρχουν και τα ψευδο-αναγνωριστικά (quasi-identifiers), που συνδυαστικά δύνανται επίσης να οδηγήσουν σε ταυτοποίηση!

Identifier	Quasi-identifier			Sensitive attribute
Name	DOB	Gender	Zipcode	Disease
Andre	1/21/76	Male	53715	Heart Disease
Beth	4/13/86	Female	53715	Hepatitis
Carol	2/28/76	Male	53703	Brochitis
Dan	1/21/76	Male	53703	Broken Arm
Ellen	4/13/86	Female	53706	Flu
Eric	2/28/76	Female	53706	Hang Nail

☞ Η απαλοιφή των αναγνωριστικών δεν διασφαλίζει εγγυημένα ανωνυμία

Παράδειγμα «κακής ανωνυμοποίησης»

(a) Patient table

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV
Dancer	Female	30	HIV
Dancer	Female	30	HIV

☞ Παράδειγμα: Νοσοκομείο «δημοσιεύει» τον ανωτέρω «ανωνυμοποιημένο» πίνακα

- Με τον όρο «δημοσίευση» εννοούμε οποιαδήποτε διαβίβαση/κοινοποίησή του σε τρίτο (π.χ. σε ιατρικό/ερευνητικό κέντρο)
- Έχει αφαιρέσει κάθε στοιχείο που θα μπορούσε να οδηγήσει στην ταυτοποίηση (ΑΦΜ, ΑΜΚΑ, Αρ. ταυτότητας, ονοματεπώνυμο)

Πόσο ανώνυμος είναι ο πίνακας;

(a) Patient table

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV
Dancer	Female	30	HIV
Dancer	Female	30	HIV

(b) External table

Name	Job	Sex	Age
Alice	Writer	Female	30
Bob	Engineer	Male	35
Cathy	Writer	Female	30
Doug	Lawyer	Male	38
Emily	Dancer	Female	30
Fred	Engineer	Male	38
Gladys	Dancer	Female	30
Henry	Lawyer	Male	39
Irene	Dancer	Female	32

Πηγή: B. Fung et.al., Privacy-Preserving Data Publishing: A Survey of Recent Developments, ACM Computing Surveys, 2010

- Έστω ότι το ιατρικό κέντρο γνωρίζει ότι στη λίστα του νοσοκομείου υπάρχουν κάποια συγκεκριμένα άτομα (π.χ. οι κάτοικοι ενός μικρού χωριού)
- Για αυτά τα άτομα, μπορεί «εύκολα» (π.χ. από δημόσια προσβάσιμες πηγές) να έχει πρόσβαση σε δεδομένα τους (βλ. Πίνακα b)
  - Συνδυάζοντας τους δύο πίνακες, μπορεί να οδηγηθεί σε ταυτοποίηση κάποιων!!
    - Π.χ. από την τριπλέτα (Job, Sex, Age) = (Laywer, Male, 38) εξάγει το συμπέρασμα για τη νόσο HIV στον Doug

Τεχνικές ανωνυμοποίησης - «Γενίκευση» γνωρισμάτων

- Ως τεχνική ανωνυμοποίησης, μεταβάλλουμε κατάλληλα τις τιμές των πεδίων που είναι ψευδο-αναγνωριστικά, μέσω γενίκευσής τους (**generalization**)
  - Π.χ δεν δημοσιεύουμε επακριβώς την ηλικία, αλλά ένα εύρος ηλικιών (π.χ. 30-40)
  - Με αυτόν τον τρόπο, η μονοσήμαντη συσχέτιση μίας καταχώρησης του «ανώνυμου» πίνακα με μία του «επώνυμου» καθίσταται πιο δύσκολη
    - Όσο πιο μεγάλη η γενίκευση, τόσο ενισχύουμε την ανωνυμοποίηση, αλλά από την άλλη πλευρά έχουμε «απώλεια» χρήσιμης πληροφορίας για ερευνητικούς σκοπούς
    - Στόχος η – κατά το δυνατόν – μέγιστη ανωνυμοποίηση, με τη μικρότερη δυνατή απώλεια πληροφορίας

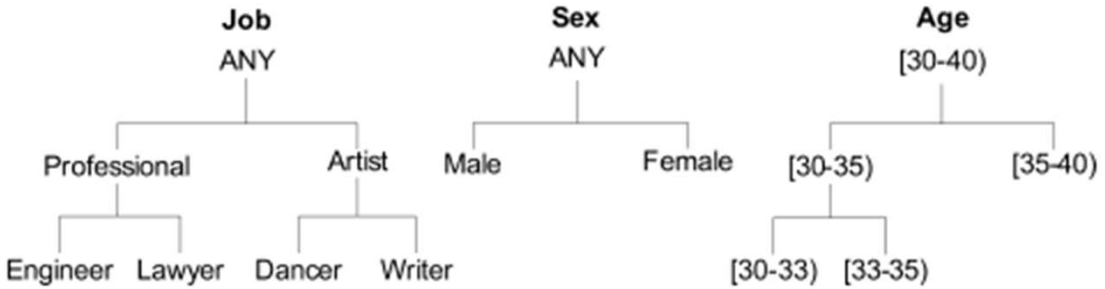
«Γενίκευση» στον προηγούμενο πίνακα

(a) Patient table

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV
Dancer	Female	30	HIV
Dancer	Female	30	HIV

«Γενίκευση» των τιμών των ψευδοαναγνωριστικών

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV



Περιγραφή των τιμών γενίκευσης

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Τι κερδίσαμε;

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

??

(b) External table

Name	Job	Sex	Age
Alice	Writer	Female	30
Bob	Engineer	Male	35
Cathy	Writer	Female	30
Doug	Lawyer	Male	38
Emily	Dancer	Female	30
Fred	Engineer	Male	38
Gladys	Dancer	Female	30
Henry	Lawyer	Male	39
Irene	Dancer	Female	32

- Η συσχέτιση δύο εγγραφών των πινάκων, συγκρίνοντας τα ψευδο-αναγνωριστικά, δεν μπορεί να γίνει
- Ανωνυμία k τάξης (k-anonymity) - Samarati-Sweeney, 1998:
  Ικανοποιείται όταν, σε έναν ανώνυμο πίνακα, το σύνολο των εγγραφών (καταχωρήσεων) με τις ίδιες τιμές στα ψευδο-αναγνωριστικά είναι τουλάχιστον k
- Προφανώς, όσο μεγαλύτερο το k, τόσο ενισχύεται η ανωνυμία
  - Ο ανωτέρω πίνακας είναι ανώνυμος 3ης τάξης

Είναι αρκετή η ανωνυμοποίηση k τάξης;

- Έστω ότι ξέρουμε τα εξής:

	Zip	Age	National
Bob →	13053	31	American
Akira →	13068	21	Japanese

- Καθώς επίσης και ότι οι Bob και Akira εμπεριέχονται σε έναν πίνακα που θα δημοσιεύσει ο εκδότης (publisher)

Αρχικά δεδομένα

	Non-Sensitive Data			*"Sensitive" Data*
#	***ZIP***	***Age***	***Nationality***	***Salary***
1	13053	28	Russian	20K
2	13068	29	American	20K
3	13068	21	Japanese	40K
4	13053	23	American	40K
5	14853	50	Indian	60K
6	14853	55	Russian	20K
7	14850	47	American	40K
8	14850	49	American	40K
9	13053	31	American	60K
10	13053	37	Indian	60K
11	13068	36	Japanese	60K
12	13068	35	American	60K

Έχουν απομακρυνθεί οι identifiers (Οι Bob, Akira έχουν «χρωματιστεί»)

Ανωνυμία 4ης τάξης

	Non-Sensitive Data			"Sensitive" Data
	#	**ZIP**	**Age**	**Nationality**	**Salary**
Η Akira ανήκει εδώ	1	130**	< 30	*	20K
	2	130**	< 30	*	20K
	3	130**	< 30	*	40K
	4	130**	< 30	*	40K
	5	*1485**	*> = 40*	*	*60K*
	6	*1485**	*> = 40*	*	*20K*
	7	*1485**	*> = 40*	*	*40K*
	8	*1485**	*> = 40*	*	*40K*
Ο Bob ανήκει εδώ	9	130**	3*	*	60K
	10	130**	3*	*	60K
	11	130**	3*	*	60K
	12	130**	3*	*	60K

Time to adopt PETs

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Ανωνυμία 4ης τάξης

	Non-Sensitive Data			"Sensitive" Data
#	**ZIP**	**Age**	**Nationality**	**Salary**
1	130**	< 30	*	20K
2	130**	< 30	*	20K
3	130**	< 30	*	40K
4	130**	< 30	*	40K
5	*1485**	*> = 40*	*	*60K*
6	*1485**	*> = 40*	*	*20K*
7	*1485**	*> = 40*	*	*40K*
			*	*40K*
			*	60K
			*	60K
11	130**	3*	*	60K
12	130**	3*	*	60K

Η Akira ανήκει εδώ

Ο Bob ανήκει εδώ

Ο Bob έχει μισθό 60K
- Εξαγωγή συμπεράσματος, παρά την «ανωνυμοποίηση»

Time to adopt PETs

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Επιθέσεις εξαγωγής συμπεράσματος (Inference attacks)

- Εφαρμόζονται όταν εξάγεται συμπέρασμα για μία «ευαίσθητη» πληροφορία ενός ατόμου, ακόμα και αν δεν αναγνωρίζεται επακριβώς ποια είναι η καταχώρησή του στον ανωνυμοποιημένο πίνακα
- Οι ανωνυμοποιήσεις τάξης k δεν μπορούν να προστατεύσουν ως προς αυτές τις επιθέσεις (βλ. προηγούμενο παράδειγμα)
- l-diversity (Machanavajjhala et al., 2006): Κάθε κλάση ισοδυναμίας πρέπει να περιέχει τουλάχιστον l «καλά ορισμένες» διακριτές τιμές του ευαίσθητου πεδίου
  - Πιο απλή περίπτωση: Distinct l-diversity
    - Σε κάθε κλάση ισοδυναμίας εμφανίζονται ακριβώς l διακριτές τιμές από το «ευαίσθητο» πεδίο

Time to adopt PETs

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Distinct 3-diversity

	Non-Sensitive Data			*«Sensitive» Data*
#	**ZIP**	**Age**	**Nationality**	**Salary**
1	1305*	<= 40	*	20K
2	1305*	<= 40	*	40K
3	1305*	<= 40	*	60K
4	1305*	<= 40	*	60K
5	*1485**	*>= 40*	*	*60K*
6	*1485**	*>= 40*	*	*20K*
7	*1485**	*>= 40*	*	*40K*
8	*1485**	*>= 40*	*	*40K*
9	1306*	<= 40	*	20K
10	1306*	<= 40	*	40K
11	1306*	<= 40	*	60K
12	1306*	<= 40	*	60K

Οι Bob, Akira ανήκουν εδώ

Time to adopt PETs

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Τεχνικές ανωνυμοποίησης - Συμπέρασμα

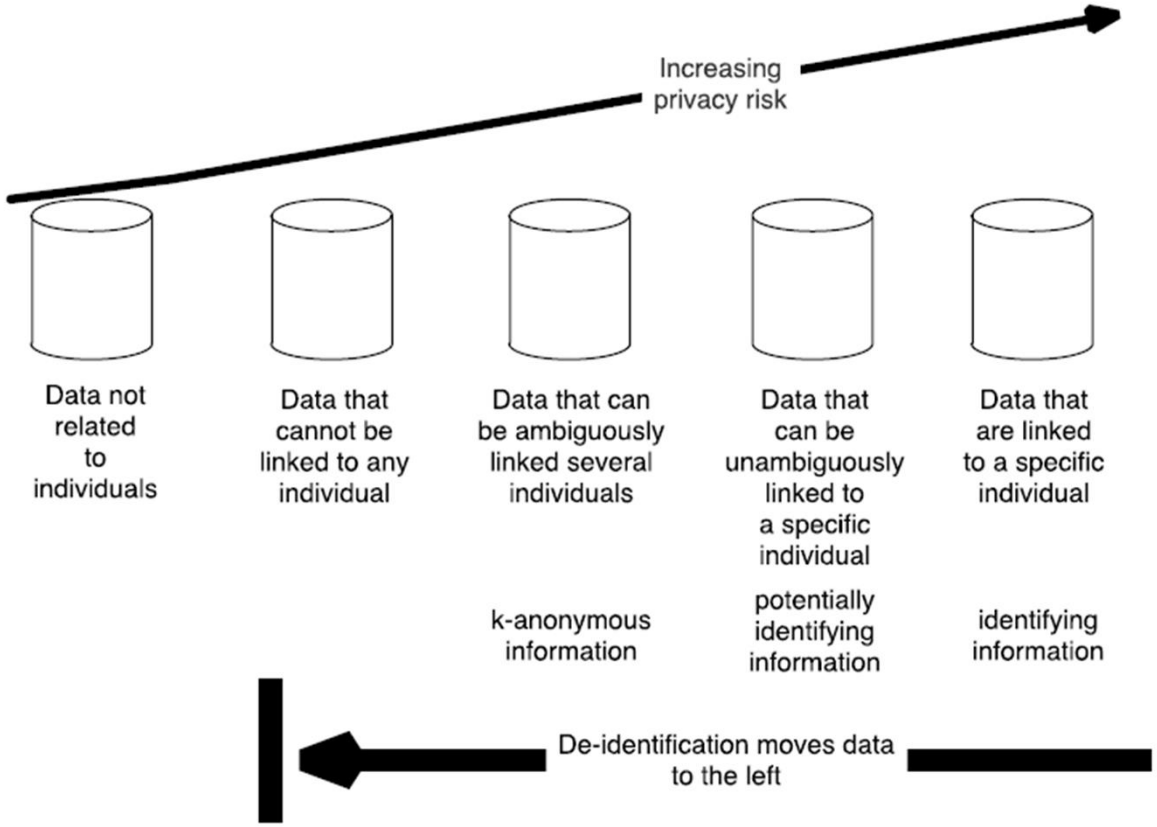
- Η ανωνυμοποίηση είναι εξαιρετικά δύσκολη ως διαδικασία
- Ούτε η τεχνική **L-diversity** μπορεί να τη διασφαλίσει πάντα
  - Π.χ. αν για τον **Bob** είχαμε **L** διαφορετικές μεν τιμές αλλά πολύ «κοντινές» μεταξύ τους, πάλι θα υπήρχε ο κίνδυνος εξαγωγής συμπεράσματος
- Διάφορες άλλες τεχνικές ανωνυμοποίησης μπορούν να αξιοποιηθούν (π.χ. προσθήκη θορύβου)
- Κάθε περίπτωση ανωνυμοποίησης πρέπει να κρίνεται ξεχωριστά, λαμβάνοντας υπόψη τους κινδύνους και την πιθανότητα επέλευσής τους
  - Μία τεχνική μπορεί να είναι κατάλληλη σε μία περίπτωση και όχι κατάλληλη σε κάποια άλλη
- Είναι πιθανό να μην καταστήσουμε τα δεδομένα ανώνυμα, αλλά ωστόσο να έχουμε περιορίσει αποτελεσματικά τους κινδύνους με τις τεχνικές αυτές

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Time to adopt PETs

«Στάδια» ανωνυμοποίησης



S. L. Garfinkel, "**De-Identification of Personal Information",** NIST Internal Report 8053, 2015

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Η εποχή των «μεγάλων δεδομένων» (Big Data)

Ευχερής δυνατότητα συγκέντρωσης και αξιοποίησης πολλών πληροφοριών από πολλές διαφορετικές πηγές μπορεί να επιφέρει:

- Ταυτοποίηση/αναγνώριση προσώπου από φαινομενικά «ανώνυμα» δεδομένα
- Εξαγωγή συμπερασμάτων – δημιουργία προφίλ

Πηγή: J. Domingo-Ferer, "Directions in Big Data Anonymization", 2016

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Ανώνυμα δεδομένα – Νομική διάσταση

Κανονισμός (ΕΕ) 2016/679 (General Data Protection Regulation - GDPR)

- Οι αρχές της προστασίας δεν θα πρέπει να εφαρμόζονται σε ανώνυμες πληροφορίες, δηλ. πληροφορίες που δεν σχετίζονται προς ταυτοποιημένο ή ταυτοποιήσιμο πρόσωπο ή σε δεδομένα που έχουν καταστεί ανώνυμα κατά τρόπο ώστε η ταυτότητα του υποκειμένου να μην μπορεί να εξακριβωθεί

- Για να κριθεί κατά πόσον ένα φυσικό πρόσωπο είναι ταυτοποιήσιμο, θα πρέπει να λαμβάνονται υπόψη όλα τα μέσα τα οποία είναι ευλόγως πιθανό ότι θα χρησιμοποιηθούν, όπως για παράδειγμα ο διαχωρισμός του (singling out), είτε από τον υπεύθυνο επεξεργασίας είτε από τρίτο για την άμεση ή έμμεση εξακρίβωση της ταυτότητας του φυσικού προσώπου.

- Για να διαπιστωθεί κατά πόσον κάποια μέσα είναι ευλόγως πιθανό ότι θα χρησιμοποιηθούν για την εξακρίβωση της ταυτότητας του φυσικού προσώπου, θα πρέπει να λαμβάνονται υπόψη όλοι οι αντικειμενικοί παράγοντες, όπως τα έξοδα και ο χρόνος που απαιτούνται για την ταυτοποίηση, λαμβανομένων υπόψη της τεχνολογίας που είναι διαθέσιμη κατά τον χρόνο της επεξεργασίας και των εξελίξεων της τεχνολογίας.
  - Άρα, και στον GDPR αναγνωρίζεται ότι ο χαρακτηρισμός δεδομένων ως ανώνυμων «χρήζει προσοχής», σε απόλυτη συνάφεια με τους κινδύνους που μόλις είδαμε

Time to adopt PETs

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Time to adopt PETs

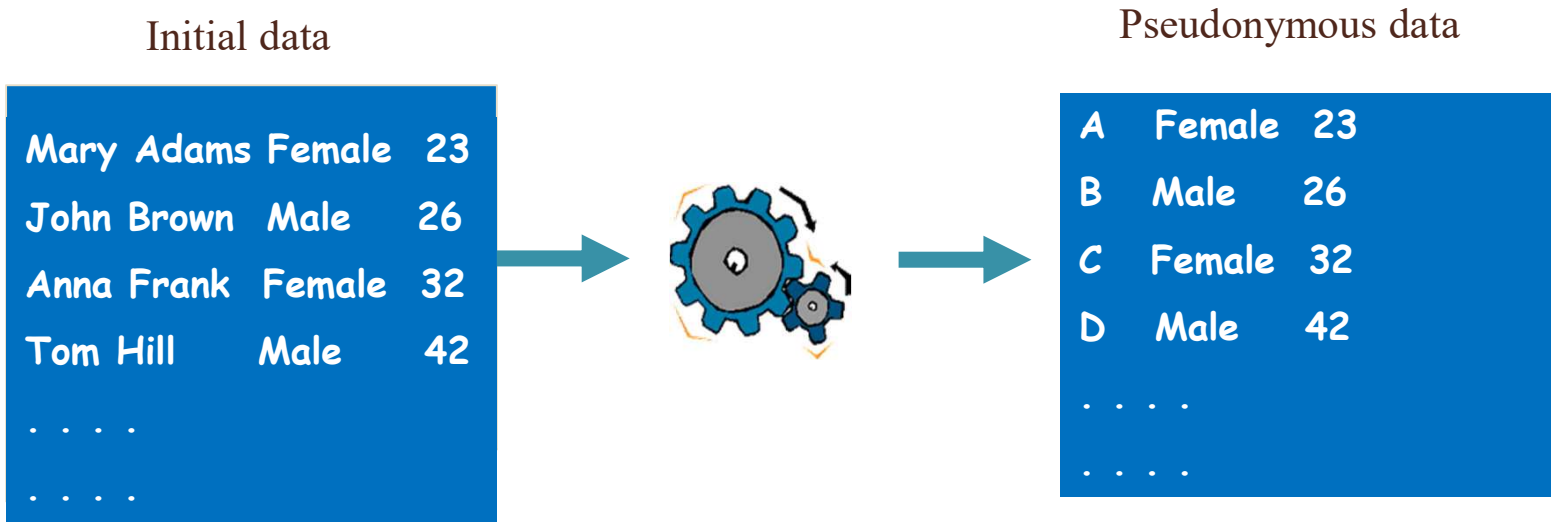
Ψευδωνυμοποιημένα δεδομένα

- Στον GDPR εμφανίζεται ο ορισμός της ψευδωνυμοποίησης:
  - *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 δεν μπορούν να αποδοθούν σε ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο*
- Στον GDPR αναφέρεται ρητώς ότι τα ψευδωνυμοποιημένα δεδομένα δεν πρέπει να θεωρούνται ανώνυμα
  - *Τα δεδομένα προσωπικού χαρακτήρα που έχουν υποστεί ψευδωνυμοποίηση, η οποία θα μπορούσε να αποδοθεί σε φυσικό πρόσωπο με τη χρήση συμπληρωματικών πληροφοριών, θα πρέπει να θεωρούνται πληροφορίες σχετικά με ταυτοποιήσιμο φυσικό πρόσωπο.*
    - ***Π.χ. ας αναλογιστούμε εκ νέου το περιστατικό της AOL – στην ουσία τα δεδομένα ήταν ψευδωνυμοποιημένα***

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Ψευδωνυμοποίηση – Ένα απλό σενάριο



- Οι αντιστοιχίσεις των αναγνωριστικών (Mary Adams, John Brown, …) με τα ψευδώνυμά τους (A, B, …) πρέπει να είναι, κατά κάποιο τρόπο, "προστατευμένες"
  - Τα ψευδώνυμα μπορούν ενδεχομένως να αντικαθιστούν περισσότερα από ένα αναγνωριστικά ή/και ψευδο-αναγνωριστικά

Time to adopt PETs

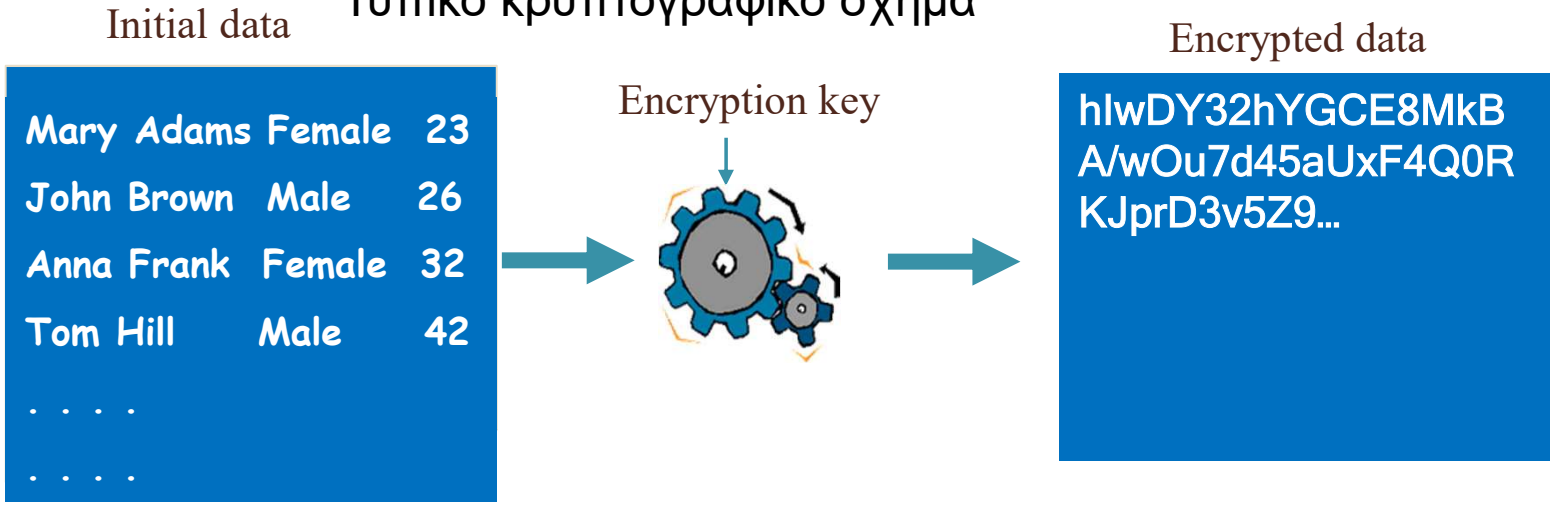
ΑΡΧΗ ΠΡΟΣΤΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Ψευδωνυμοποίηση ≠ Κρυπτογράφηση

Τυπικό κρυπτογραφικό σχήμα



- Η κρυπτογράφηση καθιστά ολόκληρα τα δεδομένα «μη αναγνώσιμα»
- Ουσιαστικά δεν μπορεί να γίνει κάποια στατιστική ανάλυση επί των κρυπτογραφημένων δεδομένων ή να αποκαλυφθεί κάποια άλλη πληροφορία – πρέπει υποχρεωτικά να αποκρυπτογραφηθούν πρώτα (απαιτείται το κλειδί αποκρυπτογράφησης)
  - Αν και υπάρχουν κρυπτογραφικές τεχνικές που «επιτρέπουν» κάποιες πράξεις επί κρυπτογραφημένων μηνυμάτων (π.χ. ομομορφική κρυπτογραφία), εν τούτοις η διαφορά ψευδωνυμοποιημένων με κρυπτογραφημένων δεδομένων είναι καταφανής

Η σημασία της ψευδωνυμοποίησης

- Ο GDPR προκρίνει τη χρήση της ψευδωνυμοποίησης
- Η λέξη «ψευδωνυμοποίηση» εμφανίζεται περί τις 15 φορές εντός του GDPR, σε διάφορες εκφάνσεις
  - Πιθανή «κατάλληλη εγγύηση» για:
    - *Επεξεργασία δεδομένων για άλλο σκοπό από αυτόν για τον οποίο έχουν αρχικώς συλλεγεί, η οποία δεν βασίζεται στη συγκατάθεση του προσώπου, προκειμένου να εξακριβωθεί κατά πόσον η επεξεργασία αυτή είναι συμβατή με τον αρχικό σκοπό (άρ. 6, παρ. 3)*
    - *Διασφάλιση της προστασίας των δεδομένων ήδη από το σχεδιασμό (άρ. 25, παρ. 1)*
    - *Ασφάλεια της επεξεργασίας (άρ. 32, παρ. 1)*
    - *Επεξεργασία για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον ή σκοπούς επιστημονικής ή ιστορικής έρευνας ή στατιστικούς σκοπούς (άρ. 89, παρ. 1)*
  - Ενθάρρυνση για να λαμβάνεται υπόψη στους κώδικες δεοντολογίας *(άρ. 40, παρ. 2)*

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Ειδικότερες περιπτώσεις που άπτονται ανωνυμοποίησης/ψευδωνυμοποίησης

- Ο βαθμός δυσκολίας της δυνατότητας άρσης της ψευδωνυμοποίησης μπορεί να επηρεάσει στον προσδιορισμό του εάν ένα περιστατικό παραβίασης δεδομένων πρέπει να ανακοινωθεί στα υποκείμενα αυτών
  - Άρθρο 34: Μία τέτοια ανακοίνωση γίνεται όταν «η παραβίαση δεδομένων προσωπικού χαρακτήρα ενδέχεται να θέσει σε υψηλό κίνδυνο τα δικαιώματα και τις ελευθερίες των φυσικών προσώπων»

- Άρθρο 11: Όταν ο υπεύθυνος επεξεργασίας μπορεί να αποδείξει ότι δεν είναι σε θέση να εξακριβώσει την ταυτότητα του υποκειμένου των δεδομένων, ο υπεύθυνος επεξεργασίας ενημερώνει σχετικά το υποκείμενο των δεδομένων.
  - Στις περιπτώσεις αυτές, <u>τα άρθρα 15 ως 20 δεν εφαρμόζονται</u>, εκτός εάν το υποκείμενο των δεδομένων (...) παρέχει συμπληρωματικές πληροφορίες που επιτρέπουν την εξακρίβωση της ταυτότητάς του.
    - Δικαιώματα πρόσβασης / διόρθωσης / διαγραφής / περιορισμού επεξεργασίας/ φορητότητας δεδομένων

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Πότε χρησιμοποιείται ψευδωνυμοποίηση τελικά;

- Ασφάλεια: Μπορεί να αποτελέσει ένα μέσο «απόκρυψης» ταυτοτήτων των χρηστών (υποκειμένων των δεδομένων)
  - Οι πραγματικές ταυτότητες των χρηστών τηρούνται ξεχωριστά, με ασφάλεια
- Ιδιωτικότητα: Μπορεί να είναι αναγκαία προϋπόθεση για τη νόμιμη επεξεργασία προσωπικών δεδομένων
  - Π.χ. δημοσίευση ερευνητικών αποτελεσμάτων σε ψευδωνυμοποιημένη μορφή ή περίπτωση όπου ο υπεύθυνος επεξεργασίας δεν χρειάζεται να γνωρίζει τις ταυτότητες των χρηστών
- Σε κάθε περίπτωση, κρίσιμο στοιχείο είναι η σωστή τεχνική ψευδωνυμοποίησης, που να εγγυάται τους ανωτέρω στόχους
  - Σε ορισμένες περιπτώσεις, μία σωστή ψευδωνυμοποίηση μπορεί να είναι αναγκαίο να συνδυαστεί με μία τεχνική ανωνυμοποίησης

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

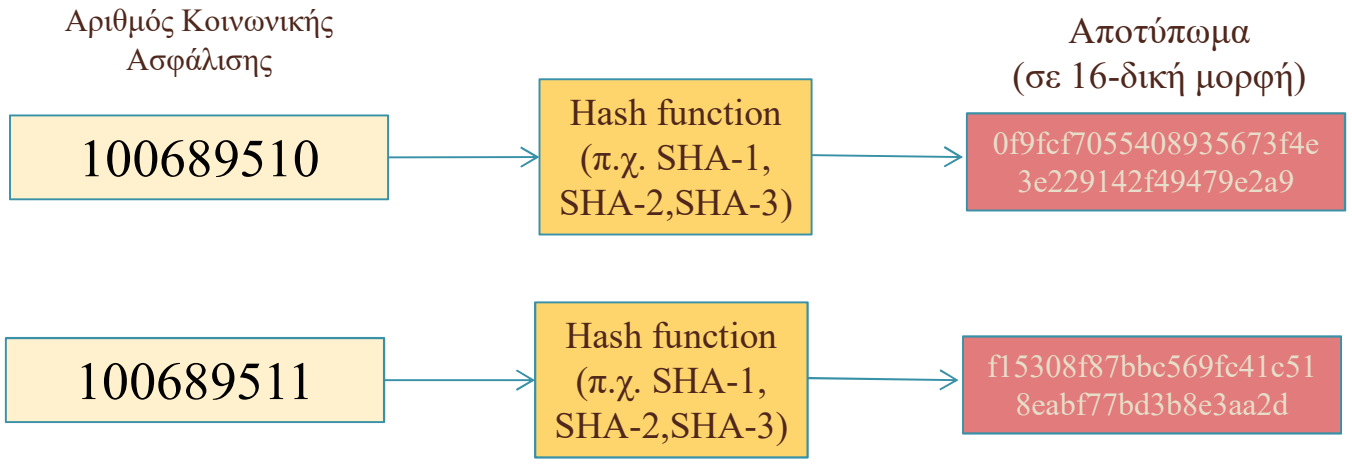
www.dpa.gr

Κρυπτογραφικές συναρτήσεις κατακερματισμού ως τεχνικές ψευδωνυμοποίησης

- Πολλοί θεωρούν ότι μία κρυπτογραφική συνάρτηση κατακερματισμού, επειδή μαθηματικά είναι μη αντιστρεπτή, μπορεί να χρησιμοποιηθεί για ψευδωνυμοποίηση προσωπικών δεδομένων
  - Π.χ. το αποτύπωμα ενός αριθμού ταυτότητας είναι μη αντιστρέψιμο
- Έχουν το πλεονέκτημα ότι θα αποδίδεται πάντα το ίδιο ψευδώνυμο στον ίδιο χρήστη – σε ορισμένες περιπτώσεις, αυτό είναι αναγκαίο
- Αν όμως είναι κρίσιμο το να μην είναι εφικτή η άρση της ψευδωνυμοποίησης από τρίτους, είναι καλή μέθοδος;

Συναρτήσεις κατακερματισμού ως τεχνικές ψευδωνυμοποίησης

Παράδειγμα: Ερευνητές θέλουν να συλλέξουν (και ενδεχομένως να δημοσιεύσουν) ψευδωνυμοποιημένα δεδομένα. Θέλουν να αποδώσουν ένα μοναδικό ακατάληπτο ψευδώνυμο σε κάθε χρήστη που συμμετείχε στην έρευνα. Θεωρούν ότι μία κρυπτογραφική συνάρτηση κατακερματισμού είναι καλή επιλογή, διότι είναι μη αναστρέψιμη και ταυτόχρονα εξασφαλίζει ότι το ίδιο ψευδώνυμο θα αποδίδεται πάντα στο ίδιο πρόσωπο

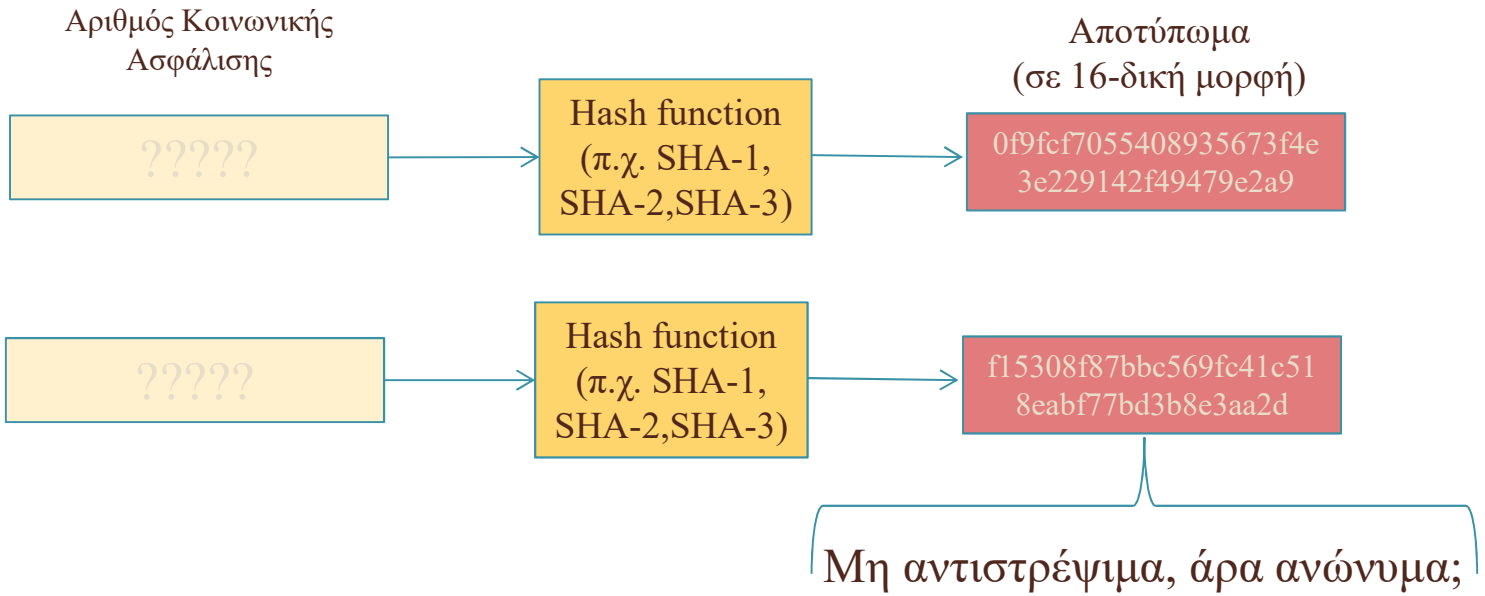


Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Συναρτήσεις κατακερματισμού ως τεχνικές ψευδωνυμοποίησης

Παράδειγμα: Ερευνητές θέλουν να συλλέξουν (και ενδεχομένως να δημοσιεύσουν) ψευδωνυμοποιημένα δεδομένα. Θέλουν να αποδώσουν ένα μοναδικό ακατάληπτο ψευδώνυμο σε κάθε χρήστη που συμμετείχε στην έρευνα. Θεωρούν ότι μία κρυπτογραφική συνάρτηση κατακερματισμού είναι καλή επιλογή, διότι είναι μη αναστρέψιμη και ταυτόχρονα εξασφαλίζει ότι το ίδιο ψευδώνυμο θα αποδίδεται πάντα στο ίδιο πρόσωπο



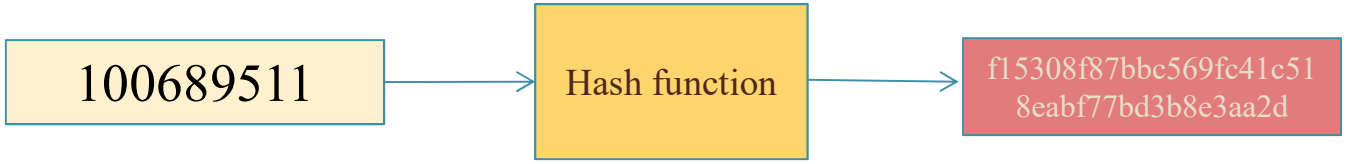
Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

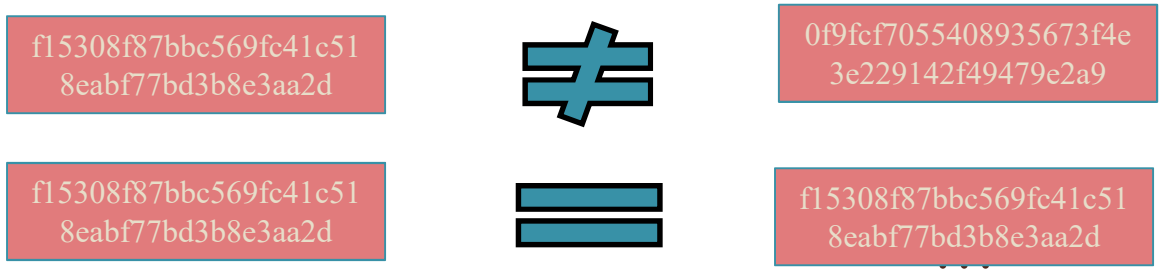
Συναρτήσεις κατακερματισμού ως τεχνικές ψευδωνυμοποίησης

- Μπορούμε να διαπιστώσουμε αν ο Α με Αριθμό Κοινωνικής Ασφάλισης (Α.Κ.Α.) 100689511 βρίσκεται στη λίστα;

1. 1) Υπολογίζουμε το αποτύπωμα του Α.Κ.Α. του χρήστη Α



- 2) Ελέγχουμε αν το αποτύπωμα είναι στην «ανωνυμοποιημένη» λίστα



Άρα, αυτή η καταχώρηση αναφέρεται στο χρήστη Α!

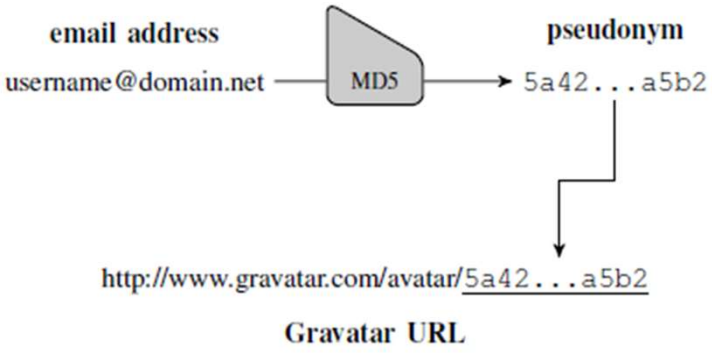
Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Η περίπτωση του Gravatar

- Το Gravatar είναι μία υπηρεσία που επιτρέπει σε χρήστες blogs, forums να έχουν αυτόματα την ίδια εικόνα ("avatar") για το προφίλ τους, αρκεί να χρησιμοποιούν πάντα την ίδια ηλεκτρονική διεύθυνση στα blogs/forums.
- Αυτές οι εικόνες για τον κάθε χρήστη είναι δημόσια προσβάσιμες με το εξής URL: https://www.gravatar.com/digest

όπου ως «digest» είναι το αποτύπωμα (με χρήση της συνάρτησης κατακερματισμού MD5) της ηλεκτρονικής διεύθυνσης του χρήστη.



Source: Demir et. al., 2018

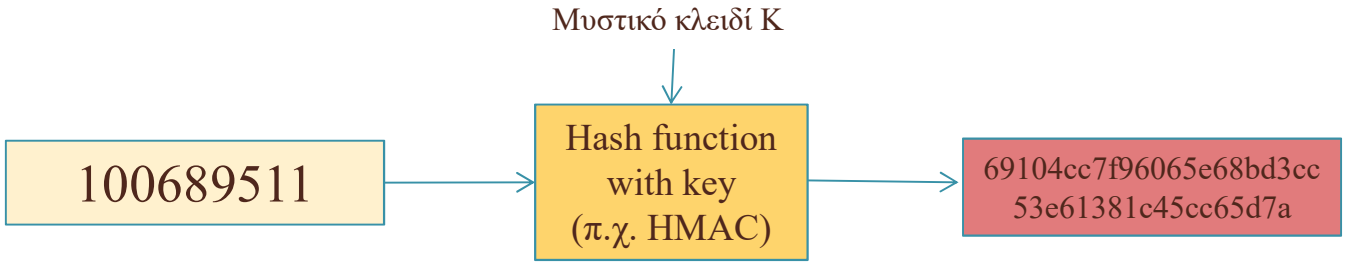
- Ερευνητές κατάφεραν να υπολογίσουν τις ηλεκτρονικές διευθύνσεις χρηστών από τα «αποτυπώματα» αυτά – δηλαδή αντιστοίχισαν e-mails με gravatars (εικόνες), χωρίς να έχουν εγγραφεί στα σχετικά blog/forum

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

«Σωστή» εναλλακτική: Συνάρτηση κατακερματισμού με κλειδί

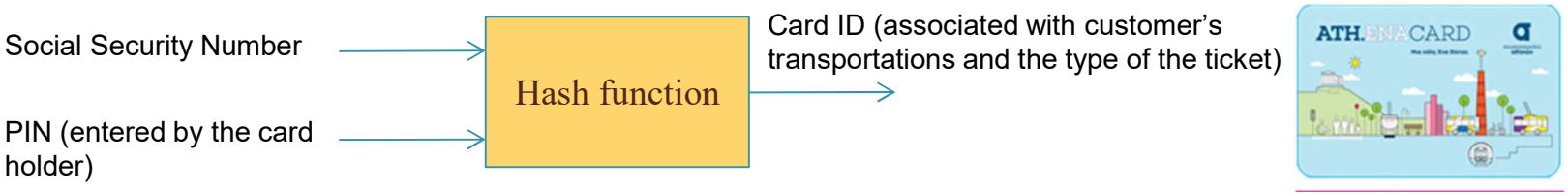
- Χρήση κρυπτογραφικής συνάρτησης με μυστικό κλειδί
  - Π.χ. υπολογισμός ενός Message Authentication Code (MAC)



1. Κανείς τρίτος δεν μπορεί, γνωρίζοντας το ψευδώνυμο, να το «αντιστρέψει» στο αρχικό αναγνωριστικό
2. Κανείς τρίτος, ο οποίος επιτελεί ψευδωνυμοποίηση, δεν θα παράγει το ίδιο ψευδώνυμο για το ίδιο αναγνωριστικό (unlinkability)

Η περίπτωση του ηλεκτρονικού εισιτηρίου στα ΜΜΜ της Αθήνας

- Ο ΟΑΣΑ ακολούθησε, τελικά, μία τεχνική ψευδωνυμοποίησης, έτσι ώστε να επιτυγχάνονται όλοι οι στόχοι του συστήματος, χωρίς όμως να ταυτοποιούνται οι επιβάτες στις διαδρομές τους



- Μοναδικό αναγνωριστικό («ψευδώνυμο») για κάθε εισιτήριο, αλλά το «κλειδί» το γνωρίζει ο κάτοχος του εισιτηρίου μόνο
- Ο ΟΑΣΑ, όπως και κάθε άλλο μέλος που ενδεχομένως αποκτήσει πρόσβαση στην υποκείμενη βάση δεδομένων, δεν θα γνωρίζει τις ταυτότητες των κατόχων των εισιτηρίων
- Εάν κάποιος χρήστης χάσει το εισιτήριό του, μπορεί να αποδείξει ότι το απωλεσθέν εισιτήριο αντιστοιχεί σε αυτόν
- => Γνώμη 4/2017 της Αρχής Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

Time to adopt PETs

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Η κρυπτογράφηση ως τεχνική ψευδωνυμοποίησης

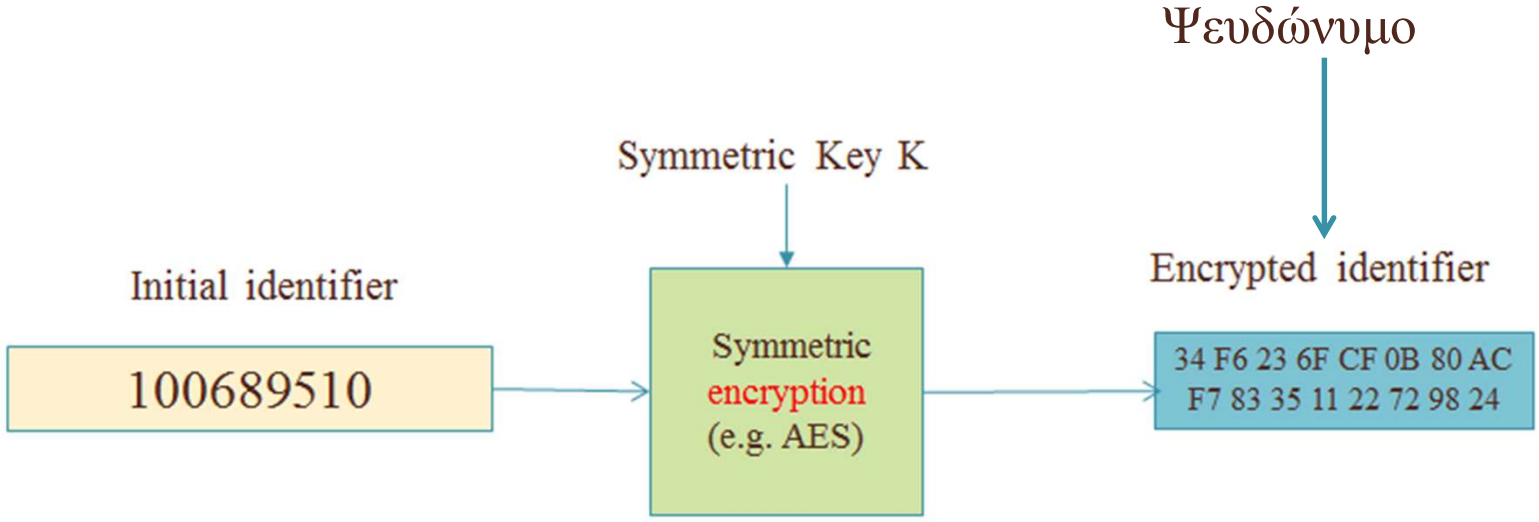
- Για την παραγωγή ψευδωνύμων, μπορούν να χρησιμοποιηθούν - μεταξύ άλλων τεχνικών - και κρυπτογραφικοί αλγόριθμοι
- Η μυστικότητα του κλειδιού διασφαλίζει την "προστασία" των αντιστοιχίσεων «αναγνωριστικών – ψευδωνύμων»
  - Το κλειδί της κρυπτογράφησης, στην περίπτωση αυτή, έχει το ρόλο των «συμπληρωματικών» πληροφοριών που περιγράφονται στον GDPR
  - Ο υπεύθυνος επεξεργασίας μπορεί εύκολα να ανακτήσει, από το ψευδώνυμο, το αρχικό αναγνωριστικό – απλά αποκρυπτογραφεί
- Ντετερμινιστική κρυπτογράφηση: ίδιο ψευδώνυμο για κάθε ίδιο αναγνωριστικό
- Πιθανοτική κρυπτογράφηση: διαφορετικό ψευδώνυμο για το ίδιο κάθε φορά αναγνωριστικό
- Αναλόγως των απαιτήσεων ως προς την προστασία προσωπικών δεδομένων, επιλέγεται κάθε φορά η βέλτιστη προσέγγιση

Time to adopt PETs

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Η ντετερμινιστική κρυπτογράφηση ως τεχνική ψευδωνυμοποίησης

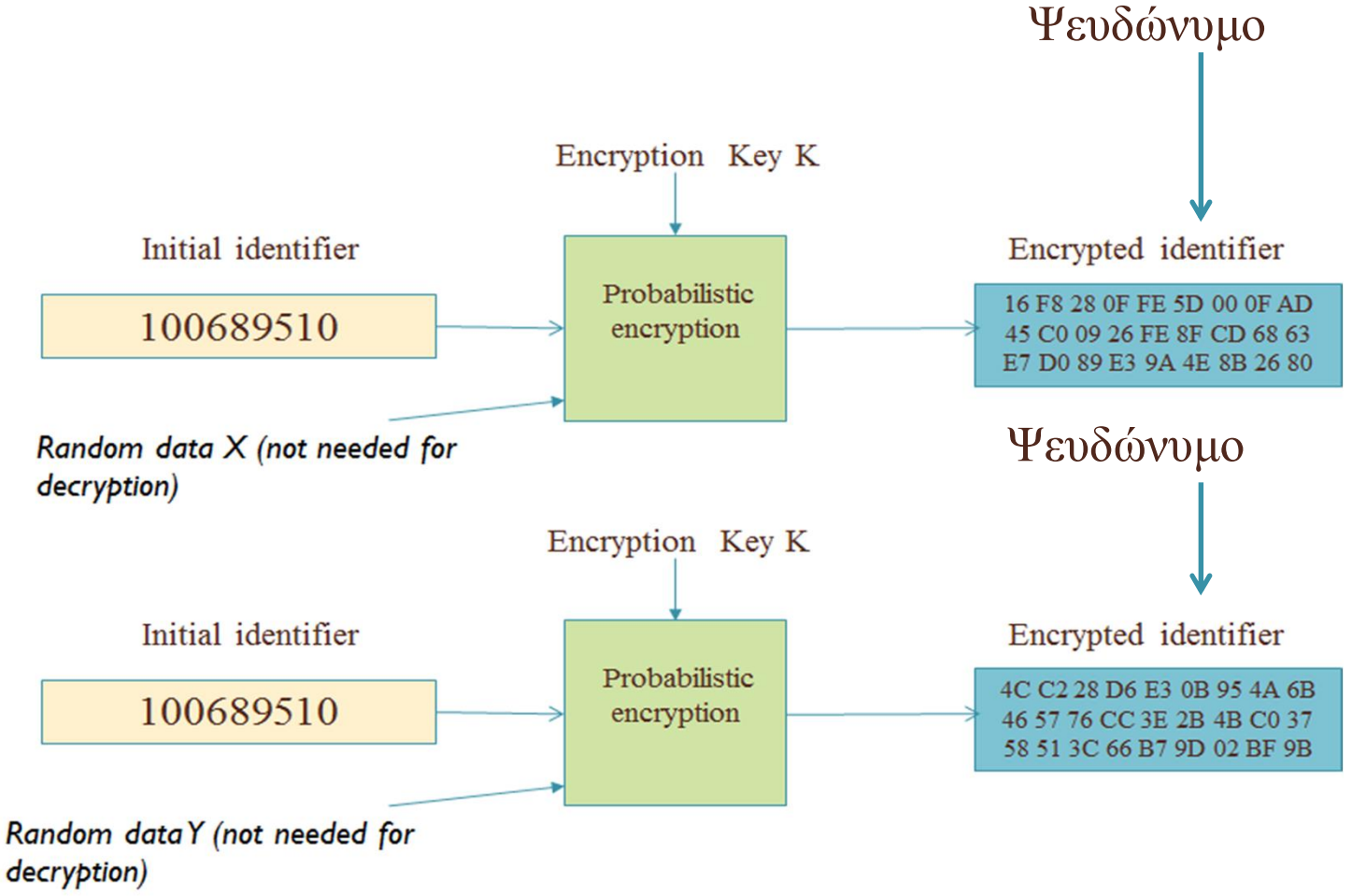


Τόσο το κλειδί, όσο και άλλες παράμετροι που υπεισέρχονται κατά την κρυπτογράφηση (π.χ. διανύσματα αρχικοποίησης σε συγκεκριμένους τρόπους λειτουργίας κρυπτογραφικών αλγορίθμων τμήματος) θα πρέπει να είναι σταθερά, προκειμένου να διασφαλίζεται ότι παράγεται πάντα το ίδιο ψευδώνυμο για το ίδιο αναγνωριστικό

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

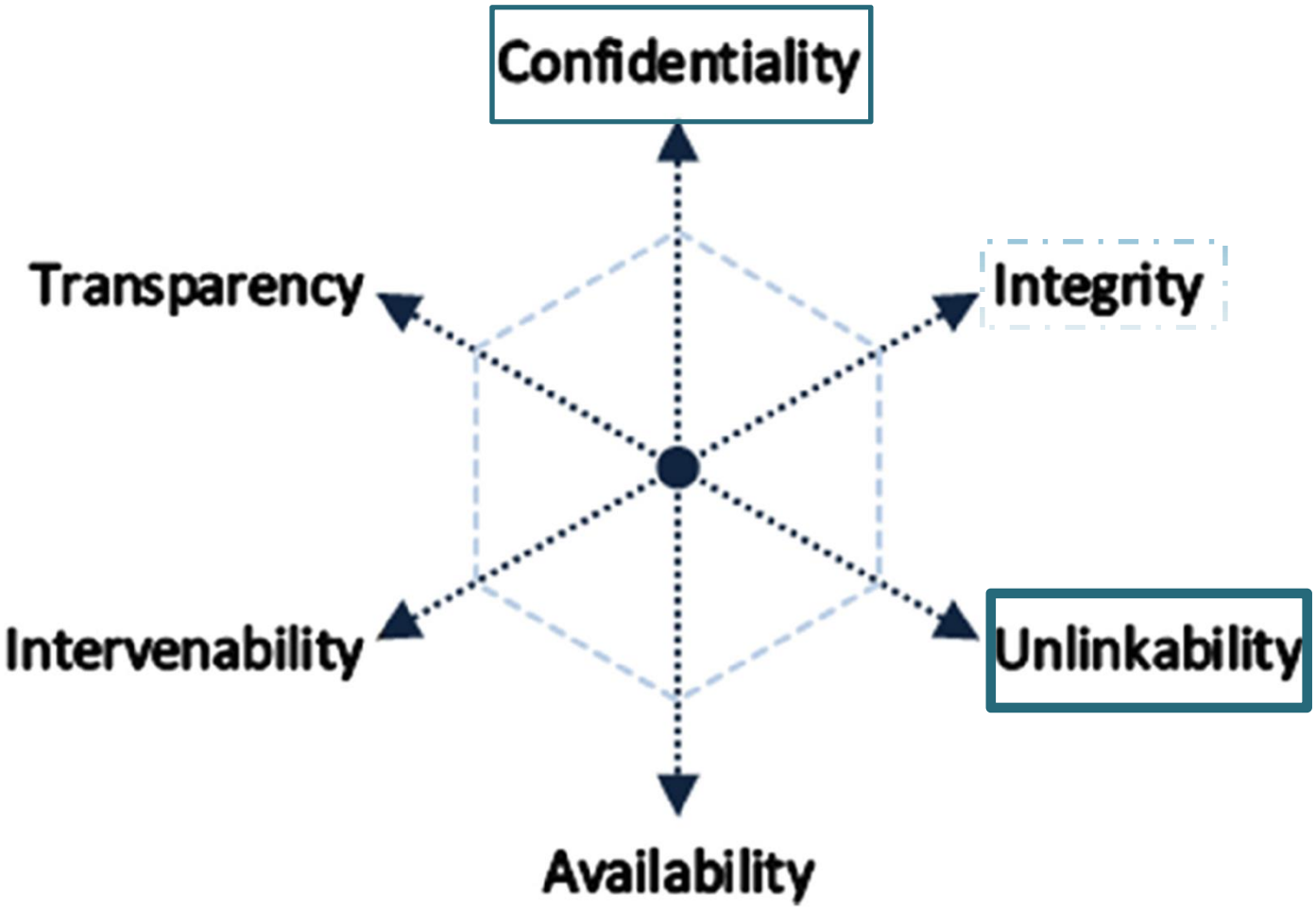
www.dpa.gr

Η πιθανοτική κρυπτογράφηση ως τεχνική ψευδωνυμοποίησης



Κατάλληλη σε περιπτώσεις όπου δεν επιθυμούμε να αποδίδουμε το ίδιο ψευδώνυμο στο ίδιο άτομο (π.χ. για να αποφευχθεί «ιχνηλάτηση» (tracking) του χρήστη).

Η ψευδωνυμοποίηση στην προστασία δεδομένων



Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Συμπεράσματα

- Η ψευδωνυμοποίηση αποτελεί πολύ σημαντικό εργαλείο ενίσχυσης της ιδιωτικότητας
  - Δεν αποτελεί όμως ανωνυμοποίηση!
  - Ενδεχομένως να πρέπει να συνδυαστεί, αναλόγως των κινδύνων από την εκάστοτε επεξεργασία, και με τεχνικές ανωνυμοποίησης
- Αν και προφανώς δεν είναι κατά κανόνα υποχρεωτική, εν τούτοις είναι πιθανό να αποτελεί κάποιες φορές αναγκαία προϋπόθεση για τη νομιμότητα της επεξεργασίας
- Σχεδιαστική πρόκληση η εύρεση, από τη στιγμή που θα κριθεί ότι πρέπει να γίνει ψευδωνυμοποίηση, της βέλτιστης τεχνικής ψευδωνυμοποίησης,
  - Παρά τη σαφή διαφορά κρυπτογράφησης με ψευδωνυμοποίησης, τεχνικές κρυπτογράφησης μπορούν ενδεχομένως να οδηγήσουν σε καλές τεχνικές ψευδωνυμοποίησης
- Μία σωστά εκπονηθείσα εκτίμηση αντικτύπου ως προς την προστασία δεδομένα μπορεί να οδηγήσει σε ασφαλείς επιλογές ως προς την ψευδωνυμοποίηση

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr

Ερωτήσεις?

Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα

www.dpa.gr