ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΑΣΦΑΛΕΙΑ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΙΔΙΩΤΙΚΟΤΗΤΑ ΔΙΑ ΤΟΥ ΣΧΕΔΙΑΣΜΟΥ ΣΧΕΣΙΑΚΩΝ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Σπυρίδωνος Νίνου

Επιβλέπων: Αν. Καθηγήτρια Ακριβή Βλάχου, Πανεπιστήμιο Αιγαίου

Μέλη εξεταστικής επιτροπής:

- Αν. Καθηγήτρια Ελισάβετ Κωνσταντίνου, Πανεπιστήμιο Αιγαίου
- Αν. Καθηγητής Θεόδωρος Κωστούλας, Πανεπιστήμιο Αιγαίου

Σάμος, Ιούνιος 2022

Η σελίδα αυτή είναι σκοπίμως κενή

Πρόλογος και ευχαριστίες

Η επιθυμία για να εκπονήσω την παρούσα διπλωματική υπήρχε για πολύ καιρό στο πίσω μέρος του μυαλού μου - πολύ πριν συμμετάσχω στο συγκεκριμένο πρόγραμμα μεταπτυχιακών σπουδών. Η ιδέα ότι θα έπρεπε να υπάρχει ένας απλός και εύκολος τρόπος ώστε να μπορούν να σχεδιαστούν βάσεις δεδομένων που να σέβονται την ιδιωτικότητα των ανθρώπων, ήταν κάτι που με απασχολούσε έντονα στην επαγγελματική μου πορεία. Έτσι, όταν μου δόθηκε η ευκαιρία να εκπονήσω την συγκεκριμένη διπλωματική, ήξερα πως μου δινόταν η ευκαιρία όχι μόνο να εκπληρώσω μια επιθυμία, αλλά και να συνδράμω εν γένει προς την διευκόλυνση του δύσκολου έργου που αναλαμβάνουν οι σχεδιαστές βάσεων δεδομένων. Η περάτωσή της όμως δεν θα ήταν εφικτή, χωρίς την συνδρομή και στήριξη συγκεκριμένων ανθρώπων.

Θα ήθελα, λοιπόν, να ευχαριστήσω θερμά την καθηγήτρια κα Ακριβή Βλάχου, για τον χρόνο και την προσπάθεια που κατέβαλε κατά την επίβλεψη της παρούσας διπλωματικής. Η καθοδήγηση και συμβολή της ήταν καθοριστική για το αποτέλεσμα της προσπάθειάς μου. Επίσης, θα ήθελα να ευχαριστήσω τα μέλη της εξεταστικης επιτροπής, κα Ελισάβετ Κωνσταντίνου και κο Θεόδωρο Κωστούλα, για τις διορθώσεις και πολύτιμες συμβουλές που έδωσαν κατά την εκπόνηση της διπλωματικής. Τέλος, θα ήθελα να ευχαριστήσω την σύζυγο και τα παιδιά μου, στους οποίους αφιερώνω αυτήν την διπλωματική, για την αμέριστη στήριξη, την αστείρευτη υπομονή και κατανόηση που έδειξαν κατά την διάρκεια της συνεχούς απουσίας μου από την καθημερινότητά τους.

ΣΠΥΡΙΔΩΝΟΣ ΝΙΝΟΥ
Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Η σελίδα αυτή είναι σκοπίμως κενή

Πίνακας περιεχομένων

Η σελίδα αυτή είναι σκοπίμως κενή

Κατάλογος Σχημάτων

Η σελίδα αυτή είναι σκοπίμως κενή

Κατάλογος Πινάκων

Η σελίδα αυτή είναι σκοπίμως κενή

Ακρωνύμια

ΒΔ	Βάση Δεδομένων
ΓΚΠΔ	Γενικός Κανονισμός Προστασίας Δεδομένων
ΝΠΔ	Νόμος Προστασίας Δεδομένων (DPA)
ΣΒΔ	Σχεσιακή Βάση Δεδομένων
ΣΔΒΔ	Σύστημα Διαχείρισης Βάσεων Δεδομένων
ΤΕΙ	Τεχνολογίες Ενίσχυσης Ιδιωτικότητας (PET)
CCTV	Closed Circuit TeleVision
DB	DataBase
DPA	Data Protection Act
GDPR	General Data Protection Regulation
IP	Internet Protocol
MTA	Mail Transfer Agent
PET	Privacy Enhancing Technologies
RDBMS	Relational DataBase Management System
SQL	Structured Query Language

Η σελίδα αυτή είναι σκοπίμως κενή

Περίληψη

Η παρούσα διπλωματική εξετάζει το θεμελιώδες ζήτημα της εφαρμογής της ιδιωτικότητας δια του σχεδιασμού, στις σχεσιακές βάσεις δεδομένων. Στην αρχή τίθενται τα όρια εντός των οποίων κινείται η διπλωματική και περιγράφονται τα θέματα με τα οποία δεν ασχολείται. Έπειτα επεξηγούνται οι ορολογίες που χρησιμοποιούνται στην διπλωματική και εξετάζονται οι λόγοι για τους οποίους η ιδιωτικότητα θεωρείται σημαντική για την κοινωνία, αναλύοντας τόσο την κοινωνιολογική όσο και την νομική διάστασή της. Ακολούθως, μελετώνται οι βασικοί τρόποι με τους οποίους παραβιάζεται η ιδιωτικότητα, καθώς και υπάρχουσες μελέτες και έρευνες σχετικά με την εφαρμογή της ιδιωτικότητας δια του σχεδιασμού. Έπειτα, γίνεται μια σύντομη αναφορά στην διαδικασία κανονικοποίησης σχεσιακών βάσεων δεδομένων και αναπτύσσεται το μοντέλο απειλών για τις βάσεις δεδομένων. Στην συνέχεια προτείνεται μια διαδικασία σχεδιασμού, η οποία έχει ως στόχο την εφαρμογή της ιδιωτικότητας δια του σχεδιασμού και αποτελεί επέκταση της κανονικοποίησης των βάσεων. Τέλος, παρατίθενται παραδείγματα εφαρμογής της διαδικασίας, από τα οποία συμπεραίνουμε πως η προτεινόμενη διαδικασία σχεδιασμού πετυχαίνει τον σκοπό της γιατί είναι απλή, κατανοητή και εφαρμόζει την ιδιωτικότητα στον καλύτερο δυνατό βαθμό.

Λέξεις Κλειδιά: *ιδιωτικότητα, ιδιωτικότητα δια του σχεδιασμού, σχεσιακές βάσεις δεδομένων, κανονικοποίηση βάσεων, ΓΚΠΔ*

Η σελίδα αυτή είναι σκοπίμως κενή

Abstract

This thesis examines the fundamental problem of implementing privacy by design on relational databases. Initially we set the boundaries within which this thesis expands and we describe the issues that we won't be dealing with. Next, we explain the terms that are used in the text and we examine the reasons that privacy is considered important for our society, taking a look at its social and legal dimensions. After that, we look at ways that privacy can be violated, as well as existing studies and research on the implementation of privacy by design. Then we make a quick reference to the relational database normalization process and we develop a threat model for relational databases. Following that, we propose a design process that targets the implementation of privacy by design, which expands on the normalization process but without making it difficult to understand or use. Finally, we see a couple of examples on applying the proposed process, which lead us to the conclusion that the proposed design process reaches its goal because it's simple, easy to understand and implements privacy to the best possible level.

Keywords: *privacy, privacy by design, relational databases, database normalization, GDPR*

1 *Εισαγωγή*

Το πρόβλημα της ιδιωτικότητας στα συστήματα επεξεργασίας

Η ιδιωτικότητα δεδομένων είναι μία πλευρά της προστασίας δεδομένων η οποία έχει προσελκύσει το ενδιαφέρον στο πέρασμα του χρόνου. Ιστορικά, οι σχεδιαστές συστημάτων έδιναν περισσότερη έμφαση στην απόδοση του αποθηκευτικού επιπέδου των συστημάτων, καθώς και στην γενικότερη προστασία του μέσω της εκτίμησης κινδύνου και των μέτρων κυβερνοασφαλείας. Αντιθέτως δεν εστίαζαν στην εφαρμογή της ιδιωτικότητας, κυρίως λόγω της πολυπλοκότητας που εισήγαγε στα συστήματα επεξεργασίας πληροφοριών. Αυτό, σε συνδυασμό με το γεγονός ότι συνέβαιναν όλο και περισσότερες παραβιάσεις ασφαλείας κατά το πέρασμα του χρόνου, είχε ως συνέπεια να αρχίσουν να προτείνονται οδηγίες εφαρμογής της ιδιωτικότητας από εθνικούς και διεθνείς οργανισμούς. Οι οδηγίες αυτές, στην συνέχεια, οδήγησαν στην δημιουργία των τεχνολογιών ενίσχυσης της ιδιωτικότητας και τον κλάδο της μηχανικής της ιδιωτικότητας. Αυτά ακολουθήθηκαν απο συστηματική μελέτη του πως η ιδιωτικότητα μπορεί να υλοποιηθεί, ειδικά στο επίπεδο των αποθηκευμένων δεδομένων ("*δεδομένα σε στάση*").

Τα δεδομένα σε στάση ανήκουν γενικά σε δύο κατηγορίες: μπορεί να είναι είτε επιχειρησιακά δεδομένα τα οποία είναι αποθηκευμένα σε κάποια βάση δεδομένων, ή ανωνυμοποιημένα δεδομένα τα οποία έχουν δημοσιευτεί ώστε να αξιοποιηθούν σε στατιστικές αναλύσεις. Για την εφαρμογή της ιδιωτικότητας στην περίπτωση των βάσεων δεδομένων, εισήχθησαν καινούρια μοντέλα (πχ Ιπποκρατικές βάσεις δεδομένων) και προτάθηκαν ειδικοί έλεγχοι πρόσβασης και μοντέλα ασφαλείας, με σκοπό να επιτευχθεί κάποιο επίπεδο ιδιωτικότητας. Αντιστοίχως, στην περίπτωση των δημοσιευμένων δεδομένων, έγιναν αρκετές μελέτες οι οποίες περιγράφουν ανωνυμοποιητικούς αλγορίθμους για τα δεδομένα. Αυτοί οι αλγόριθμοι στοχεύουν στην προστασία των ιδιοκτητών των δεδομένων (τα υποκείμενα στα οποία αναφέρονται τα δεδομένα) από το να αναγνωριστούν από (κακόβουλους) αναλυτές των δεδομένων. Τέτοιοι αλγόριθμοι είναι οι πολύ γνωστοί *k-anonymity*, *l-diversity*, *t-closeness* και *διαφορικής ιδιωτικότητας*.

Αν και υπάρχει αυξανόμενη βιβλιογραφία στο θέμα της ιδιωτικότητας δεδομένων, η οποία πιθανόν να ενισχυθεί απο την σχετικά πρόσφατη νομοθεσία γύρω από την ιδιωτικότητα σε πολλές χώρες (όπως ο Ευρωπαϊκός ΓΚΠΔ και οι Βρετανικοί ΓΚΠΔ και ΝΠΔ), υπάρχει ακόμα απόσταση

ανάμεσα στην θεωρία που προσφέρεται από μελέτες και του πως οι μηχανικοί λογισμικού προστατεύουν την ιδιωτικότητα στα προϊόντα λογισμικού που δημιουργούν.

Η απόσταση αυτή οφείλεται σε διάφορους λόγους - ένας είναι ότι μερικές προσεγγίσεις απαιτούν ειδικά σχεδιασμένα συστήματα διαχείρισης βάσεων δεδομένων, τα οποία δεν είναι ευρέως διαδεδομένα (πχ οι Ιπποκρατικές βάσεις δεδομένων ή οι γραμμές παραγωγής δεδομένων με διαφορική ιδιωτικότητα). Άλλος είναι ότι, τα προσωπικά δεδομένα προτιμάται να αποθηκεύονται χωρίς κάποια προστασία για λόγους απόδοσης (στην απάντηση ερωτημάτων) της βάσης. Τέλος, ένας λόγος είναι πως δεν υπάρχει εύκολη λύση στην εφαρμογή της ιδιωτικότητας των ανθρώπων όσον αφορά στα αντίγραφα ασφαλείας.

Παραδοχές διπλωματικής

Η εκπόνησή της διπλωματικής βασίστηκε σε κάποιες παραδοχές, οι οποίες αναλύονται παρακάτω:

Επιθέσεις εκ των έσω

Το βασικότερο πρόβλημα των βάσεων δεδομένων που δεν χρησιμοποιούν κρυπτογράφηση καθολικά, είναι πως οποιοσδήποτε έχει πρόσβαση στον αποθηκευτικό χώρο των δεδομένων μπορεί να τα προσπελάσει χωρίς περιορισμούς. Θεωρητικά εκείνοι που έχουν πρόσβαση στον αποθηκευτικό χώρο των βάσεων δεδομένων είναι άνθρωποι εξουσιοδοτημένοι, άρα τον μεγαλύτερο κίνδυνο παρουσιάζουν οι επιθέσεις εκ των έσω[1] (παραβίαση την ιδιωτικότητας από νομίμως εξουσιοδοτημένα άτομα). Στην διπλωματική γίνεται η παραδοχή ότι οι επιτιθέμενοι είναι εξουσιοδοτημένες οντότητες (βλέπε και ενότητα "*Αναγωγή του Προβλήματος Προστασίας"* παρακάτω).

Ακεραιότητα προγραμμάτων

Η ακεραιότητα των προγραμμάτων θεωρείται δεδομένη και απαραβίαστη όσον αφορά στην διπλωματική. Κατά κανόνα τα τεχνικά μέτρα ασφαλείας βασίζονται στην ορθή υλοποίησή τους απο προγράμματα, τα οποία πέρα από ελέγχους εξουσιοδοτήσεως κρατάνε και πληροφορίες στην μνήμη τους (πχ κλειδιά αποκρυπτογραφήσεως). Η ακεραιότητα των προγραμμάτων και τα αντίμετρα επιθέσεων σε αυτά δεν εμπίπτουν στο πεδίο μελέτης της διπλωματικής.

Διαχείριση κλειδιών

Όσον αφορά στην διαχείριση των κλειδιών, γίνεται η παραδοχή πως ο σχεδιαστής ακολουθεί τις βέλτιστες πρακτικές που περιγράφονται γενικά στην διεθνή βιβλιογραφία. Συνεπώς, ο κύκλος ζωής των κλειδιών ("*key lifecycle*") που απαιτούνται για την υλοποίηση κάποιων από τις προτάσεις της παρούσας διπλωματικής, δεν εξετάζονται.

Αναγωγή του προβλήματος προστασίας

Όταν ένας επιτιθέμενος παραβιάσει την ασφάλεια της πρόσβασης σε ένα σύστημα διαχείρισης βάσεων δεδομένων, τότε αποκτά πρόσβαση στην βάση ισοδύναμη με κάποιον νομίμως

[1] Στην παρούσα εργασία, ο όρος "*επιθέσεις εκ των έσω*" ("insider attacks") χρησιμοποιείται με την πιο στενή έννοια του όρου, όπως περιγράφεται στην ενότητα "*Μοντέλο απειλών*"

εξουσιοδοτημένο χρήστη[2]. Συνεπώς, το πρόβλημα της επιθέσεως από παράγοντα εκτός ορίων του συστήματος ανάγεται σε επίθεση εκ των έσω. Δηλαδή το πρόβλημα προσπέλασης από μη εξουσιοδοτημένη οντότητα ανάγεται σε πρόβλημα προσπέλασης από νομίμως εξουσιοδοτημένη οντότητα.

Όσον αφορά στις επιθέσεις εκ των έσω, γίνεται η παραδοχή ότι ο επιτιθέμενος έχει πρόσβαση στο σύνολο της βάσης δεδομένων. Συνεπώς, σε τέτοιες επιθέσεις μπορούμε να θεωρήσουμε πως τα δεδομένα *δημοσιεύονται* κατα κάποιον τρόπο, σκεπτικό που βασίζεται στην δυνατότητα του επιτιθεμένου να αντιγράψει τα δεδομένα και να τα δημοσιοποιήσει κατά βούληση και χωρίς περιορισμούς.

Βάσει της αναγωγής αυτής του προβλήματος σε πρόβλημα ανωνυμοποίησης δεδομένων, γίνεται η παραδοχή πως αν επιτευχθεί προστασία από επιθέσεις σε δημοσιευμένα δεδομένα, επιτυγχάνεται ο σκοπός της ιδιωτικότητας.

Αρχή της "αναγκαίας γνώσης"

Η αναγκαία γνώση ("*need-to-know*") είναι μια αρχή η οποία περιορίζει την πρόσβαση σε πληροφορίες ανάλογα με το ποιές από αυτές είναι αναγκαίες για να τελεστεί μια εργασία. Σύμφωνα με αυτή την αρχή, κάποια οντότητα (άνθρωπος ή πρόγραμμα) θα πρέπει να έχει πρόσβαση μόνο στις πληροφορίες που χρειάζεται για να περατώσει την εργασία που του έχει ανατεθεί, ενώ θα πρέπει να του απαγορεύεται η πρόσβαση στις υπόλοιπες πληροφορίες *ακόμα κι αν έχει την κατάλληλη εξουσιοδότηση για να τις προσπελάσει*.

Ένα παράδειγμα της εφαρμογής αυτής της αρχής δίνεται από την διαχείριση πληροφοριών στο πλαίσιο του στρατεύματος μιας χώρας. Πιο συγκεκριμένα, οι αξιωματικοί του στρατού εξουσιοδοτούνται απο το στράτευμα για την διαχείριση πληροφοριών, βάσει βαθμού. Για παράδειγμα ένας ταγματάρχης μπορεί να διαχειριστεί (λάβει γνώση ή/και διαβιβάσει) πληροφορίες απόρρητες ή άκρως απόρρητες. Όμως, βάσει της αρχής της αναγκαίας γνώσης, ο εν λόγω ταγματάρχης θα πρέπει να έχει πρόσβαση μόνο στις απόρρητες πληροφορίες οι οποίες είναι απαραίτητες για την τέλεση του καθήκοντός του και δεν θα πρέπει να μπορεί να προσπελαύνει απόρρητες πληροφορίες γενικά του στρατεύματος. Με αυτόν τον τρόπο μειώνεται ο κίνδυνος του στρατεύματος στην περίπτωση που ο αξιωματικός αποφασίσει να κατασκοπεύσει εκ μέρους κάποιου άλλου κράτους, καθώς μπορεί να δώσει πληροφορίες μόνο για τις δικές του αρμοδιότητες που, θεωρητικά, είναι περιορισμένες σε σχέση με την γενική στρατηγική της εθνικής άμυνας ολόκληρης της χώρας.

Στην παρούσα διπλωματική γίνεται η παραδοχή πως η ιδιωτικότητα αποτελεί περίπτωση της αρχής της αναγκαίας γνώσης καθώς διαφέρουν μόνο ως προς το ποιός εξουσιοδοτεί την πρόσβαση στις προστατευμένες πληροφορίες. Ενώ, δηλαδή, στην αρχή της αναγκαίας γνώσης η εξουσιοδότηση δίνεται από κάπου κεντρικά, στην ιδιωτικότητα συμμετέχει στην απόφαση της εξουσιοδότησης (είτε μερικώς είτε εξ'ολοκλήρου) το υποκείμενο επεξεργασίας (στον οποίο ανήκουν τα δεδομένα προς επεξεργασία).

Αντίγραφα ασφαλείας

Τα αντίγραφα ασφαλείας μιας βάσης δεδομένων είθισται να μην υπόκεινται σε περιορισμούς προσβάσεως, αντίστοιχους με εκείνους που επιβάλλει ένα σύστημα διαχειρίσεως βάσεων δεδομένων. Μπορεί βεβαίως να υπάρχουν άλλοι περιορισμοί - πχ ελεγχόμενη πρόσβαση στον

[2] Νομίμως εξουσιοδοτημένη οντότητα είναι η οντότητα που εξουσιοδοτείται να επεξεργαστεί δεδομένα από τον ιδιοκτήτη του συστήματος επεξεργασίας. Όταν ένας επιτιθέμενος παραβιάσει την ασφάλεια του συστήματος, αποκτά το επίπεδο εξουσιοδότησης της οντότητας την οποία παραβίασε, αλλά δεν είναι νομίμως εξουσιοδοτημένος

χώρο όπου φυλάσσονται τα αντίγραφα - όμως αρκετοί από αυτούς δεν εφαρμόζονται κατά κανόνα, ειδικά σε μικρές εταιρείες ή οργανισμούς. Αυτό γιατί τα μέτρα προστασίας αντιγράφων ασφαλείας στοχεύουν κυρίως στο να μην καταστραφούν ή κλαπούν τα αντίγραφα (και δεν είναι διαθέσιμα σε περίπτωση που χρειαστούν) και όχι στην ιδιωτικότητα των υποκειμένων επεξεργασίας. Στην παρούσα διπλωματική, λοιπόν, γίνεται η παραδοχή πως τα αντίγραφα ασφαλείας ισοδυναμούν με δημοσιευμένα δεδομένα, οπότε για να προστατευτεί η ιδιωτικότητα στην επεξεργασία αντιγράφων ασφαλείας, θα πρέπει να ληφθούν μέτρα αντιμετώπισης για επεξεργασία δημοσιευμένων δεδομένων.

Σκοπός διπλωματικής

Στην παρούσα διπλωματική εξετάζεται το πως θα τροποποιηθεί η διαδικασία σχεδιασμού των *σχεσιακών βάσεων δεδομένων*, προσφέροντας έτσι έναν τρόπο στους μηχανικούς να προστατεύσουν την ιδιωτικότητα δια του σχεδιασμού. Σκοπός είναι να βρεθεί ένας τρόπος σχεδιασμού ή κάποιο μοντέλο δεδομένων το οποίο θα προστατεύει την ιδιωτικότητα των ιδιοκτητών των δεδομένων, θα μπορεί να χρησιμοποιηθεί από όλα τα διαδεδομένα συστήματα διαχείρισης σχεσιακών βάσεων δεδομένων χωρίς ειδικές μετατροπές (πχ PostgreSQL, MariaDB κτλ), θα αντιμετωπίζει στον καλύτερο βαθμό τις επιθέσεις εκ των έσω και θα καλύπτει την περίπτωση των αντιγράφων ασφαλείας, όλα σύμφωνα με την νομοθεσία περί ιδιωτικότητας και ειδικά τον Ευρωπαϊκό ΓΚΠΔ.

Θεμελιώδης στόχος της διπλωματικής είναι ότι η χρήση της προτεινόμενης διαδικασίας σχεδιασμού θα καθιστά δύσκολη την παραβίαση της ιδιωτικότητας των υποκειμένων επεξεργασίας απο εξουσιοδοτημένες οντότητες[3]. Όπως θα δούμε και στο Κεφάλαιο 4 ("*Παραβιάσεις Ιδιωτικότητας*"), υπάρχουν διάφοροι τρόποι για να παραβιάσει ένας επιτιθέμενος την ιδιωτικότητα των υποκειμένων επεξεργασίας, όμως η διπλωματική εστιάζει στην αντιμετώπιση της παραβίασης *ως προς την ταυτοποίηση προσώπων* και του συσχετισμού με τις υπόλοιπες διαθέσιμες (μη ταυτοποιητικές) πληροφορίες.

Δομή της διπλωματικής

Στο επόμενο κεφάλαιο (Κεφάλαιο 2) παραθέτουμε τους ορισμούς των ορολογιών που χρησιμοποιούνται στην διπλωματική. Στο Κεφάλαιο 3 κάνουμε μια σύντομη ανασκόπηση της έννοιας την ιδιωτικότητας και της νομοθεσίας σχετικά με αυτήν. Έπειτα εξετάζουμε κάποιες επιθέσεις στην ιδιωτικότητα (Κεφάλαιο 4), όπου βλέπουμε τους τρόπους με τους οποίους μπορεί κάποιος να παρακάμψει διάφορους μηχανισμούς εφαρμογής της. Στο Κεφάλαιο 5 εστιάζουμε στην ιδιωτικότητα δια του σχεδιασμού, αναφέροντας τις βασικές αρχές της καθώς και διάφορες μελέτες κι έρευνες που έχουν γίνει πάνω σε αυτήν. Έπειτα, (Κεφάλαιο 6) υπενθυμίζουμε τα βασικά σημεία σχεδιασμού μιας σχεσιακής βάσης δεδομένων, ώστε να μπορέσουμε να κατανοήσουμε πως μπορούμε να τροποποιήσουμε την διαδικασία του σχεδιασμού ωστε να προστατεύει την ιδιωτικότητα των υποκειμένων επεξεργασίας. Στο Κεφάλαιο 7 παραθέτουμε ένα μοντέλο απειλών των σχεσιακών βάσεων δεδομένων και των συστημάτων διαχειρίσεώς τους. Η βασική συνεισφορά της διπλωματικής παρουσιάζεται στο Κεφάλαιο 8, στο οποίο εξετάζουμε την διαδικασία του σχεδιασμού μιας βάσης με στόχο την εφαρμογή της ιδιωτικότητας. Τέλος, στο Κεφάλαιο 9 αναλύουμε τα συμπεράσματα της διπλωματικής.

[3] Νόμιμες και μη

2 *Ορισμοί και ορολογία*

Στο παρόν κεφάλαιο θα προσδιορίσουμε τις σημαντικές ορολογίες και τους ορισμούς που χρησιμοποιούνται στην παρούσα διπλωματική. Σε κάποιους ορισμούς βάζουμε συγκεκριμένους περιορισμούς ώστε να μειώσουμε το εύρος της εφαρμογής τους, δεδομένου ότι απευθυνόμαστε σε επιθέσεις και αντίμετρα σε σχεσιακές βάσεις δεδομένων.

αγαθό (asset): οτιδήποτε έχει αξία για τον ιδιοκτήτη του. Στην παρούσα διπλωματική, αγαθό εννοείται η (σχεσιακή) βάση δεδομένων την οποία στοχεύουμε να προστατεύσουμε

ακτίνα δράσης ευπάθειας (blast radius): το εύρος και πλήθος των αγαθών που επηρεάζονται από μια ευπάθεια. Όταν μια ευπάθεια επηρεάζει πολλά αγαθά, λέμε ότι έχει μεγάλη ακτίνα δράσης

απειλή (threat): είναι *η δυνατότητα* να ζημιωθεί ένα αγαθό

διάνυσμα ή διαδρομή επιθέσεως (attack vector): είναι η διαδρομή, ή το σύνολο των βημάτων, που ακολουθείται κατά την διάρκεια μιας επιθέσεως ώστε να διεισδύσει ένας εισβολέας (πράκτορας) στο υπο επίθεση σύστημα. Ονομάζεται διάνυσμα γιατί στα διαγράμματα συστημάτων η επίθεση απεικονίζεται με ένα βέλος, το οποίο επίσης χρησιμοποιείται για να απεικονίσει διανύσματα. Ο όρος συναντάται και ως *διαδρομή επιθέσεως (attack path)*

διεύρυνση εξουσιοδότησης ή αύξηση προνομίων (privilege escalation): η απόκτηση, εκ μέρους μιας οντοτητας, μεγαλύτερου εύρους εξουσιοδότησης (περισσότερα προνόμια, δηλαδή) σε σχέση με την νόμιμη και εγκεκριμένη που έχει προαποφασιστεί για εκείνην. Στο πλαίσιο της κυβερνοασφαλείας, ο όρος χρησιμοποιείται για την παράνομη απόκτηση αυτών των

εξουσιοδοτήσεων, δηλαδή είτε με εκμετάλλευση ευπαθειών συστημάτων είτε με χρήση υποκλαπέντων διαπιστευτηρίων χρηστών που νομίμως διαθέτουν τις επιθυμητές εξουσιοδοτήσεις

δυνατότητα (possibility): είναι η κατάσταση η οποία μας επιτρέπει να κάνουμε κάτι. Είναι απόλυτο μέγεθος, δηλαδή η δυνατότητα υπάρχει ή όχι. Για παράδειγμα ο άνθρωπος δεν μπορεί να πετάξει (χωρίς βοηθήματα) γιατί δεν διαθέτει φτερά. Λέμε οτι *ο άνθρωπος δεν έχει την δυνατότητα να πετάξει*. Αντιθέτως, *ένα πτηνό έχει την δυνατότητα να πετάξει* (εάν το επιθυμεί) γιατί είναι δομικά κατασκευασμένο για να το κάνει

εκτίμηση κινδύνου (risk assessment): είναι η διαδικασία με την οποία αξιολογούμε (ποσοτικοποιούμε) τις απειλές κατά των αγαθών μας, τον κίνδυνο επιθέσεων, καθώς και τις συνέπειες που προκύπτουν από τις επιτυχείς επιθέσεις εναντίων τους

επίθεση (attack): η ενεργός προσβολή ενός συστήματος, δηλαδή όλες οι πράξεις από την πρώτη προσπάθεια απόκτησης πρόσβασης στο σύστημα έως την επίτευξη κάποιου στόχου (πχ δολιοφθορά, εξαγωγή δεδομένων, παρακολούθηση/υποκλοπή κτλ). **Διευκρίνιση:** στην παρούσα διπλωματική, ο όρος αυτός αναφέρεται σε περιορισμένη έννοια της λέξης, δηλαδή εννοείται πως είναι η *ενεργός πράξη* προσβολής ενός συστήματος. Η ανίχνευση ευπαθειών (*vulnerability scanning*) και η χαρτογράφηση της τοπολογίας ενός δικτύου (*network scanning*) θεωρούμε ότι *δεν* αποτελούν μέρος της επίθεσης. Θεωρούμε ότι είναι πράξεις που προετοιμάζουν την επίθεση, γιατί συλλέγουν πληροφορίες ώστε εκείνη να σχεδιαστεί και πραγματοποιηθεί. Ανήκουν δηλαδή στην φάση της συλλογής πληροφοριών, που *προηγείται* της επίθεσης. Αντιθέτως, στην βιβλιογραφία συναντάται συνήθως η διευρυμένη έννοια της επίθεσης, η οποία περιλαμβάνει και την φάση της συλλογής πληροφοριών

επιτιθέμενος (attacker): οντότητα (άνθρωπος ή πρόγραμμα) που επιτίθεται σε κάποιο σύστημα

επίθεση εκ των έσω (insider attack): η επίθεση που πραγματοποιείται από νομίμως εξουσιοδοτημένη οντότητα. *Σημείωση*: στην βιβλιογραφία (πχ [10]) συναντάται ο όρος "*εσωτερική απειλή*" (insider threat) ο οποίος χρησιμοποιείται στην θέση του όρου "επίθεση εκ των έσω"[4]. Συνεπώς, η έννοια του όρου είναι αρκετά διευρυμένη και περιλαμβάνει εκτός από την έννοια της απειλής και την έννοια της επίθεσης, η οποία μπορεί να είναι αποτέλεσμα λάθους ή αμέλειας εξουσιοδοτημένων χρηστών (πχ απώλεια υπολογιστών με κρίσιμα δεδομένα). Δεν πρέπει όμως να αγνοείται *η διαφορά μεταξύ απειλής και επίθεσης* (βλέπε ορισμούς στην παρούσα ενότητα), συνεπώς δεν είναι ακριβές να εναλλάσσονται οι δύο όροι. Αυτό γιατί ενώ οι απειλές που οφείλονται σε εξουσιοδοτημένες οντότητες μπορεί να είναι *εσκεμμένες ή ακούσιες*, οι επιθέσεις εκ των έσω είναι πιο συγκεκριμένες και οφείλονται σε *εσκεμμένες* ενέργειες απο τις εξουσιοδοτημένες οντότητες. Στην παρούσα εργασία, για τις επιθέσεις εκ των έσω θεωρουμε πως δεν συμπεριλαμβάνονται οντότητες που απέκτησαν με παράνομο τρόπο νόμιμη εξουσιοδότηση

επιφάνεια επίθεσης (attack surface): το σύνολο όλων των σημείων εισόδου σε ένα σύστημα ή όλων των τρόπων πρόσβασης σε ένα αγαθό

[4] Στο [10], ενότητα 1.1 ("definition of insider threat") αναφέρεται: "a malicious insider threat is a current or former employee, [...] who has or had authorized access [...] and intentionally exceeded or misused that access [...]"

ευπάθεια (vulnerability): είναι *η δυνατότητα* να χρησιμοποιηθεί ένα αγαθό με τρόπο για τον οποίο δεν προορίζεται και είναι βλαπτικός κι ανεπιθύμητος

κίνδυνος (risk): είναι *η πιθανότητα* να ζημιωθεί κάποιο αγαθό, δηλαδή να χάσει μερικώς ή ολικώς την αξία του

μοντέλο απειλών (threat model): είναι το υπόδειγμα (μοντέλο) που προκύπτει απο την συστηματική απαρίθμηση των απειλών σε ένα σύστημα, καθώς και τα μέτρα αντιμετώπισής τους. Δηλαδή, στο μοντέλο αναγνωρίζονται τα αγαθά που θέλουμε να προστατεύσουμε, οι απειλές εναντίων των αγαθών καθώς και τα αντίμετρα των απειλών για κάθε αγαθό

μοντέλο ασφαλείας (security model): είναι η πολιτική ή διαδικασία που ορίζει τις οντότητες, τους τύπους επεξεργασίας και τους τύπους δεδομένων ενός συστήματος, καθως και τις σχέσεις μεταξύ τους. Πιο πρακτικά, το μοντέλο ασφαλείας ορίζει το *ποιός μπορεί να κάνει τί, και σε ποιά δεδομένα*. Το μοντέλο ασφαλείας βασίζεται στο μοντέλο απειλών, καθώς αξιοποιεί τα μέτρα αντιμετώπισης των απειλών που ορίζονται στο μοντέλο απειλών

περίοδος ευκαιρίας (window of opportunity): η χρονική περίοδος κατά την οποία μας δίνεται η δυνατότητα να εκμεταλλευτούμε κάποια ευπάθεια σε ένα σύστημα και σχετίζεται άμεσα με την συγκεκριμένη ευπάθεια του συστήματος. Κάθε ευπάθεια συνδέεται με διαφορετική περίοδο ευκαιρίας ανά σύστημα στο οποίο υπάρχει. Η κάθε περίοδος ευκαιρίας ορίζεται από τον σχεδιασμό του συστήματος. Για παράδειγμα, έστω ότι υπάρχει μια ευπάθεια σε ένα σύστημα μεταφοράς μηνυμάτων ηλεκτρονικού ταχυδρομείου (MTA). Έστω επίσης πως σε ένα σύστημα που χρησιμοποιείται το συγκεκριμένο σύστημα μεταφοράς μηνυμάτων, δίνεται η δυνατότητα να αποστέλλονται μηνύματα καθ'όλο το 24ωρο και όλες τις ημέρες της εβδομάδας. Έστω, τέλος, πως σε ένα άλλο σύστημα που χρησιμοποιεί το συγκεκριμένο σύστημα μεταφοράς μηνυμάτων, δίνεται η δυνατότητα να αποστέλλονται μηνύματα μόνο τις πρωινές ώρες των εργάσιμων ημερών. Απο το παράδειγμα φαίνεται πως στα δύο συστήματα υπάρχει η ίδια ευπάθεια, όμως υπάρχει διαφορετική περίοδος ευκαιρίας για την εκμετάλλευσή της - η περίοδος ευκαιρίας για το πρώτο σύστημα είναι πολύ μεγαλύτερη από την περίοδο ευκαιρίας του δεύτερου συστήματος

πιθανότητα (probability): είναι η προσδοκία μας, το πόσο πολύ περιμένουμε δηλαδή, να συμβεί κάτι. Είναι ποσοστό που κυμαίνεται ανάμεσα σε δύο άκρα, την βεβαιότητα ότι *κάτι θα συμβεί* και την βεβαιότητα ότι *κάτι δεν θα συμβεί*. Η πιθανότητα εξαρτάται από *την ύπαρξη ή όχι της δυνατότητας* να συμβεί κάτι. Δηλαδή, αν δεν υπάρχει η δυνατότητα, τότε η πιθανότητα είναι *μηδενική (απίθανο και αδύνατο)*. Ενώ, αν υπάρχει η δυνατότητα (*δυνατό*) τότε η πιθανότητα είναι *εγγυημένη (πιθανό)*, αλλά δεν είναι βέβαιο ότι θα συμβεί αυτό που περιμένουμε. Η διαφορά με την δυνατότητα (*possibility*) είναι πως, η δυνατότητα αναφέρεται γενικά σε μια κατάσταση για το αν μας επιτρέπει να κάνουμε κάτι, αν δηλαδή από την φύση της η κατάσταση μας επιτρέπει. Η πιθανότητα, από την άλλη, εξαρτάται από ειδικές συνθήκες οι οποίες επηρεάζουν την κατάσταση ως προς το αν τελικά μας επιτρέπουν να κάνουμε αυτό το κάτι. Για παράδειγμα, έστω ότι έχουμε έναν υπολογιστή ο οποίος εκτελεί ένα πρόγραμμα εξυπηρετητή διαδικτύου (*web server*). Ο υπολογιστής βρίσκεται απομονωμένος σε τοπικό δίκτυο, όμως είναι εκτεθειμένος στο διαδίκτυο (δηλαδή μπορεί κάποιος να συνδεθεί σε αυτόν μέσω διαδικτύου). Έστω ότι ο εξυπηρετητής χρησιμοποιείται αποκλειστικά για να δημοσιεύει ("*σερβίρει*") στατικές σελίδες σχετικά με την θερμοκρασία του δωματίου στον οποίο βρίσκεται ο υπολογιστής. Έστω, τέλος, πως ο

εξυπηρετητής είναι ενημερωμένος στην τελευταία έκδοση και δεν φαίνεται να υπάρχουν (γνωστές) ευπάθειες για την έκδοσή του. Βλέπουμε λοιπόν πως *υπάρχει η δυνατότητα* να επιτεθεί κάποιος στον εξυπηρετητή αφού η έκθεση του εξυπηρετητή στο διαδίκτυο *επιτρέπει απο την φύση της* τις απομακρυσμένες συνδέσεις. Η *πιθανότητα* οπως να επιτεθεί κάποιος είναι *μικρή ή ανύπαρκτη* γιατί ο επιτιθέμενος δεν μπορεί να χρησιμοποιήσει τον εξυπηρετητή ως σημείο εισόδου (*point of entrance*) λόγω έλλειψης ευπαθειών, ή να αλλάξει τομέα ασφαλείας χρησιμοποιώντας τον ως σημείο περιστροφής (*pivoting*) ή, έστω, να τον χρησιμοποιήσει ως μέσο για πλευρική κίνηση (*lateral movement*)

πλευρική κίνηση (lateral movement): η διαδικασία κατά την οποία κάποιος επιτιθέμενος, ο οποίος έχει ήδη πρόσβαση σε έναν τομέα ασφαλείας (ενός ή περισσοτέρων δικτύων ή συστημάτων), αποκτά πρόσβαση σε άλλες συσκευές (πχ υπολογιστές, δρομολογητές, τείχη προστασίας, σημεία ασύρματης προσβάσεως κτλ) *που ανήκουν στον ίδιο τομέα ασφαλείας*, με σκοπό την απόκτηση πληροφοριών

σημείο εισόδου επίθεσης (attack entry point / point of entrance): το σημείο (δικτυακή συσκευή, λειτουργικό σύστημα ή πρόγραμμα), του υπο επίθεση συστήματος ή δικτύου, το οποίο χρησιμοποιεί ο επιτιθέμενος ώστε να αποκτήσει την αρχική πρόσβαση στο σύστημα ή δίκτυο. Εννοείται, δηλαδή, το σημείο εισόδου του επιτιθέμενου στο σύστημα ή δίκτυο

τομέας εμπιστοσύνης (trust domain): το τμήμα του συστήματος στο οποίο οι πληροφορίες που διακινούνται θεωρούνται έγκυρες σε κάποιον συγκεκριμένο βαθμό. Οι πληροφορίες που θεωρούνται έγκυρες *στον ίδιο βαθμό* σχηματίζουν *έναν τομέα εμπιστοσύνης*. Για παράδειγμα, όλες οι πληροφορίες που λαμβάνονται από χρήστες δεν θεωρούνται έμπιστες. Όλα τα σημεία που επεξεργάζονται τέτοιου είδους πληροφορίες, καθώς και οι ίδιες πληροφορίες, ανήκουν στον ίδιο τομέα εμπιστοσύνης (πχ στον τομέα που δεν εμπιστευόμαστε τις πληροφορίες). Τα σημεία στα οποία ελέγχονται οι πληροφορίες του προηγούμενου τομέα, τα σημεία επεξεργασίας των εγκεκριμένων πληροφοριών και οι εγκεκριμένες πληροφορίες αποτελούν έναν άλλον τομέα εμπιστοσύνης. Ένα άλλο παράδειγμα είναι πως, έστω ότι χωρίζουμε τον βαθμό εγκυρότητας των πληροφοριών σε 3 επίπεδα: στο πρώτο είναι οι πληροφορίες που δεν εμπιστευόμαστε. Αυτές μπορεί να είναι οι πληροφορίες που δίνουν οι χρήστες και που είναι εκτός ελέγχου του συστήματος. Στο δεύτερο επίπεδο εγκυρότητας βρίσκονται οι πληροφορίες που έχουν ελεγχθεί για την σωστή δομή τους. Στο τρίτο επίπεδο εγκυρότητας βρίσκονται οι πληροφορίες που έχουν ελεγχθεί για την σωστή δομή τους, καθώς επίσης και για το περιεχόμενό τους - οτι δηλαδή δεν περιέχουν δεδομένα τα οποία είναι ύποπτα για εκμετάλλευση κάποιας ευπάθειας. Τα τρία επίπεδα εγκυρότητας του παραδείγματος αποτελούν τρεις διαφορετικούς τομείς εμπιστοσύνης του συστήματος

τομέας ασφαλείας (security domain): είναι το τμήμα του συστήματος στο οποίο *οι πληροφορίες διακινούνται μεταξύ οντοτήτων που έχουν το ίδιο επίπεδο εξουσιοδότησης* για την επεξεργασία πληροφοριών. Διαφορετικοί τομείς ασφαλείας απαιτούν κατά κανόνα διαφορετικού τύπου εξουσιοδότηση. Συνεπώς, όταν μια οντότητα ανήκει σε έναν τομέα ασφαλείας (έχει δηλαδή εξουσιοδοτηθεί να επεξεργάζεται πληροφορίες ενός τομέα) και θέλει να λειτουργήσει σε άλλον τομέα ασφαλείας (δηλαδή θέλει να επεξεργαστεί πληροφορίες που ανήκουν σε άλλον τομέα) τότε κατά κανόνα πρέπει να *εξουσιοδοτηθεί* για τον καινούριο τομέα ασφαλείας. Να σημειωθεί πως η εξουσιοδότηση μπορεί να συμβεί με πολλούς τρόπους, απο το να επιτρέπεται η μετάβαση από τον έναν τομέα στον άλλον βάσει διευθύνσεων διαδικτύου (IP) έως το να ζητώνται διαφορετικά

διαπιστευτήρια ανά τομέα ασφαλείας από την κάθε οντότητα. **Σε σχέση με τους τομεις εμπιστοσύνης:** ένας τομέας ασφαλείας μπορεί να περιλαμβάνει πολλούς τομείς εμπιστοσύνης, ή ένας τομέας εμπιστοσύνης να επεκτείνεται σε πολλούς τομείς ασφαλείας. Αυτό συμβαίνει γιατί ο τομέας ασφαλείας έχει να κάνει με εξουσιοδότηση για την επεξεργασία πληροφοριών ενώ ο τομέας εμπιστοσύνης έχει να κάνει με το πόσο εμπιστευόμαστε ότι οι πληροφορίες που επεξεργαζόμαστε είναι έγκυρες ή όχι

περιστροφή (pivoting): τακτική που χρησιμοποιείται για να συνδεθεί κάποιος επιτιθέμενος σε σημείο του συστήματος (στόχος) που δεν επιτρέπει την απ'ευθείας σύνδεση από το τμήμα του δικτύου στο οποίο βρίσκεται ήδη ο επιτιθέμενος (πηγή). Σύμφωνα με αυτήν την τακτική, ο επιτιθέμενος συνδέεται πρώτα σε συσκευή (πχ υπολογιστής, δρομολογητής κτλ) η οποία ονομάζεται ***σημείο περιστροφής*** (*pivoter*) και είναι εξουσιοδοτημένη να συνδέεται στο επιθυμητό σημείο. Έπειτα, μέσω της ενδιάμεσης συσκευής, συνδέεται στο επιθυμητό σημείο του συστήματος ([1], [2], [3]). Επειδή τα δύο σημεία δικτύου τα οποία εμπλέκονται στην περιστροφή (πηγή και στόχος) δεν επιτρέπεται να συνδεθούν άμεσα αλλά απαιτείται η σύνδεση μέσω άλλου σημείου (περιστροφής), συνήθως το σημείο πηγή βρίσκεται σε *διαφορετικό τομέα ασφαλείας* από το σημείο στόχο. Επίσης, επειδή το σημείο περιστροφής εξουσιοδοτεί τον επιτιθέμενο να συνδεθεί στον στόχο, το σημείο περιστροφής ανήκει *στον τομέα ασφαλείας του στόχου*, ή σε τομέα ασφαλείας ο οποίος βρίσκεται *ιεραρχικά υψηλότερα* από τον τομέα ασφαλείας του στόχου και μπορεί να δώσει διαπιστευτήρια για τον τομέα ασφαλείας του στόχου στον επιτιθέμενο

Αντίστροφο λεξικό ορισμών

asset: αγαθό
attack: επίθεση
attack entry point / point of entrance: σημείο εισόδου της επίθεσης
attack surface: επιφάνεια επιθέσεων
attack vector: διάνυσμα ή διαδρομη επιθέσεως
attacker: επιτιθέμενος
blast radius: ακτίνα δράσης ευπάθειας
insider attack: επίθεση εκ των έσω
lateral movement: πλευρική κίνηση
pivoting: περιστροφή
possibility: δυνατότητα
privilege escalation: διεύρυνση εξουσιοδότησης ή αύξηση προνομίων
probability: πιθανότητα
risk: κίνδυνος
risk assessment: εκτίμηση κινδύνου
security domain: τομέας ασφαλείας
security model: μοντέλο ασφαλείας
threat: απειλή
threat model: μοντέλο απειλών
trust domain: τομέας εμπιστοσύνης

vulnerability: ευπάθεια
window of opportunity: περίοδος ευκαιρίας

3 *Ιδιωτικότητα*

Εισαγωγή

Για να μπορέσουμε να κατανοήσουμε την αρχή της ιδιωτικότητας δια του σχεδιασμού, πρέπει πρώτα να γίνει αντιληπτή η έννοια της ιδιωτικότητας. Επίσης, είναι σημαντικό να κατανοήσουμε τον λόγο για τον οποίο η ιδιωτικότητα θεωρείται θεμελιώδες ανθρώπινο δικαίωμα και, βάσει αυτού, νομοθετήθηκε η προστασία της. Η θεώρηση της ιδιωτικότητας γίνεται βασιζόμενοι στην θεσμοθέτησή της ως δικαίωμα από την νομοθεσία. Ο βασικός νόμος στον οποίο στηρίχθηκε η ανάλυση της διπλωματικής είναι ο Γενικός Κανονισμός Προστασίας Δεδομένων ("ΓΚΠΔ") και ειδικά το Άρθρο 25 το οποίο θεσπίζει την υποχρέωση της τήρησης της αρχής της ιδιωτικότητας δια του σχεδιασμού των συστημάτων επεξεργασίας πληροφοριών.

Θεμελίωση ως δικαίωμα

Η ιδιωτικότητα προκύπτει ως θεμελιώδες ανθρώπινο δικαίωμα από το Άρθρο 8 της Ευρωπαϊκής Σύμβασης για τα Ανθρώπινα Δικαιώματα, το οποίο εγγυάται το δικαίωμα στον σεβασμό της ιδιωτικής και οικογενειακής ζωής, της οικίας και της αλληλογραφίας των ανθρώπων ([35]). Αν και αποτελεί σύμβαση μεταξύ των κρατών μελών του Συμβουλίου της Ευρώπης, εν τούτοις έχει εναρμονιστεί στις εθνικές νομοθεσίες των μελών (πχ για την Ελλάδα, το δεύτερο μέρος του Συντάγματος: "Ατομικά και κοινωνικά δικαιώματα", [24], και ειδικά για την ιδιωτικότητα, το Άρθρο 9Α).

Όμως, παρά την νομοθέτηση της ιδιωτικότητας ως δικαίωμα, η έννοιά της δεν αποσαφηνίζεται ή αποτυπώνεται σε κάποιο νομοθετικό κείμενο, με αποτέλεσμα το τι σημαίνει ιδιωτικότητα και τί αυτή περιλαμβάνει να αφήνεται προς ερμηνεία στους μελετητές του νόμου.

Προσδιορισμός της έννοιας

Για τον λόγο του οτι η έννοια της ιδιωτικότητας δεν είναι αποσαφηνισμένη, ούτε οι ιδιότητές της ή η έκτασή της γίνονται αντιληπτές από όλους στον ίδιο βαθμό, είναι σημαντικό να εξετάσουμε τις επικρατούσες απόψεις για το τι είναι η ιδιωτικότητα, και το πως αυτές επηρέασαν το σκεπτικό του νομοθέτη κατά την νομοθέτηση της προστασίας της. Επικρατούν, λοιπόν, εν γένει τρεις θεωρίες σχετικα με το τι είναι η ιδιωτικότητα ([25]):

Η πρώτη έρχεται από τον πολιτικό επιστήμονα και νομικό **Alan Westin**, ο οποίος θεωρεί πως η ιδιωτικότητα είναι *η δυνατότητα ενός ανθρώπου να περιορίσει προσωρινά την πρόσβαση άλλων στον εαυτό του*. Όπως διευκρινίζει, πέρα από την δυνατότητα να μπορεί ο κάθε άνθρωπος να ελέγχει το ποιές πληροφορίες που τον αφορούν θα μπορούν να επικοινωνηθούν σε άλλους, θέτει και την δυνατότητα κάθε ανθρώπου να μπορεί να αποστασιοποιηθεί φυσικά και ψυχολογικά από την υπόλοιπη κοινωνία. Ο Westin παραθέτει διάφορα κριτήρια και επιχειρήματα για την θεμελίωση της θεωρίας του, τα οποία προέρχονται από τον χώρο της πολιτικής και νομικής επιστήμης.

Η επόμενη θεωρία έρχεται από τον κοινωνικό ψυχολόγο **Irwin Altman**, ο οποίος θεωρεί (αντιστοίχως με τον Westin) πως *η ιδιωτικότητα ενός ανθρώπου είναι ο επιλεκτικός έλεγχος της προσβάσεως στον εαυτό του*. Η διαφορά όμως με την θεωρία του Westin είναι πως ο Altman θεωρεί την ιδιωτικότητα ως κοινωνικό φαινόμενο (κι όχι ατομική επιλογή), το οποίο βασίζεται στην κοινωνική συμπεριφορά και την αμοιβαία αντιμετώπιση. Τα διάφορα κριτήρια και επιχειρήματα που χρησιμοποίησε ο Altman για να θεμελιώσει την δική του θεωρία έρχονται από την κοινωνιολογία και την "*συμπεριφορική*" ψυχολογία. Είναι σημαντικό εδώ να επισημάνουμε πως η θεωρία του Altman υποστηρίζεται απο εμπειρικά δεδομένα.

Τέλος, πιο πρόσφατη είναι η θεωρία της *Διαχείρισης της Ιδιωτικότητας της Επικοινωνίας* (Communication Privacy Management) της κοινωνιολόγου και ψυχολόγου **Sandra Petronio**. Η θεωρία αυτή βασίζεται σε εκείνην του Altman (άρα αποδέχεται τον ορισμό της ιδιωτικότητας κατά τον ίδιο τρόπο), όμως επαναπροσδιορίζει το πως τίθενται τα όρια προσβάσεως στις προσωπικές πληροφορίες του κάθε ανθρώπου. Διατυπώνει πως ο κάθε ένας μπορεί να κινηθεί μεταξύ δύο άκρων: το ένα άκρο είναι να δώσει απεριόριστη πρόσβαση στις προσωπικές του πληροφορίες και σε απεριόριστο αριθμό ανθρώπων, το άλλο είναι να απαγορεύσει εξ'ολοκλήρου την πρόσβαση στις προσωπικές του πληροφορίες στους πάντες, και φυσικά μπορεί να επιλέξει οποιαδήποτε άλλη διαβάθμιση προσβάσεως από ορισμένο σύνολο ανθρώπων. Η θεωρία της Petronio θεωρείται η επικρατέστερη θεωρία ως προς το τι είναι η ιδιωτικότητα και πως αυτή εκφράζεται (ή θα έπρεπε να εκφράζεται) στην κοινωνία.

Σύμφωνα λοιπόν με τα παραπάνω, μπορούμε να πούμε πως η ιδιωτικότητα ορίζεται ως η εν γένει δυνατότητα ενός ανθρώπου να ελέγξει την πρόσβαση άλλων στις πληροφορίες που τον αφορούν. Ο ορισμός αυτός προβλέπει είτε τον πλήρη αποκλεισμό προσβάσεως, είτε μερικό αποκλεισμό (αντίστοιχα, μερική επίτρεψη) προσβάσεως ή την πλήρη επίτρεψη προσβάσεως στις διάφορες κατηγορίες πληροφοριών που τον αφορούν, απο πεπερασμένο ή απεριόριστο αριθμό ανθρώπων.

Φαίνεται πως ο παραπάνω ορισμός είναι ο κοινά αποδεκτός μεταξύ των νομοθετών, καθώς οι νόμοι που θεσπίζουν την προάσπιση της ιδιωτικότητας κατά την επεξεργασία δεδομένων προσωπικού χαρακτήρα κινούνται γύρω από αυτόν. Επίσης, αξίζει να σημειώσουμε πως στους ορισμούς της ιδιωτικότητας που προκύπτουν από τις προαναφερθείσες θεωρίες, η έκφραση "πρόσβαση στον εαυτό τους" φαίνεται να ερμηνεύεται (και να απεικονίζεται νομικά) όχι ως η φυσική πρόσβαση ή εγγύτητα στον άνθρωπο, αλλά ως η πρόσβαση στις πληροφορίες που τον αφορούν. Αυτή φαίνεται να θεωρείται πλέον η αποδεκτή προσέγγιση ως προς τις ιδιότητες της

ιδιωτικότητας, ακόμη και σε ό,τι αφορά στην εφαρμογή της σε νέες τεχνολογίες (πχ διαδίκτυο των πραγμάτων, [26]).

Σύγκριση με την προστασία δεδομένων

Σε αυτό το σημείο οφείλουμε να αποσαφηνίσουμε την διαφορά μεταξύ της ιδιωτικότητας και της προστασίας δεδομένων. Κατ'αρχήν, η αναφορά σε δεδομένα στο πλαίσιο της ιδιωτικότητας εννοεί τα δεδομένα προσωπικού χαρακτήρα ([13]), άρα η έκφραση "προστασία δεδομένων" έχει την έννοια της "προστασίας δεδομένων προσωπικού χαρακτήρα".

Η κατανόηση της διαφοράς των δύο εννοιών (προστασία δεδομένων και ιδιωτικότητα) έχει σημασία επειδή ο ΓΚΠΔ στοχεύει στην προστασία των προσωπικών δεδομένων στο πλαίσιο της έννομης επεξεργασίας τους απο τον υπεύθυνο επεξεργασίας. Υπο αυτό το πρίσμα, η εν γένει προστασία δεδομένων αφορά κυρίως σε εξωτερικές απειλές, ενώ *η ιδιωτικότητα αφορά σε προστασία από εσωτερικές απειλές*.

Η προστασία δεδομένων, λοιπόν, συνδέεται άρρηκτα με την έννοια της ασφαλείας πληροφοριών (εν προκειμένω, οι πληροφορίες είναι τα προσωπικά δεδομένα). Ο στόχος της ασφαλείας πληροφοριών είναι η προστασία της εμπιστευτικότητας, ακεραιότητας και διαθεσιμότητας[5] των πληροφοριών. Συνεπώς, όταν ο ΓΚΠΔ αναφέρεται στην προστασία προσωπικών δεδομένων, αναφέρεται εν γένει στην εξασφάλιση των τριών αυτών συνθηκών. Ο λόγος που απαιτείται η προστασία των προσωπικών δεδομένων είναι για να μπορεί το υποκείμενο της επεξεργασίας να ασκήσει τα δικαιώματα που νομοθετούνται απο τον κανονισμό (πχ δικαίωμα στην λήθη, στον περιορισμό της επεξεργασίας κτλ).

Η προστασία όμως των προσωπικών δεδομένων μέσω της εξασφάλισης των τριών συνθηκών ασφαλείας δεν συνεπάγεται και την εξασφάλιση της ιδιωτικότητας, παρ'ότι η ιδιωτικότητα είναι άρρηκτα συνδεδεμένη με την εμπιστευτικότητα των δεδομένων. Προκύπτει το ζήτημα πως ενώ η προστασία δεδομένων αναφέρεται γενικά στην εμπιστευτικότητα των δεδομένων προσωπικού χαρακτήρα ενός φυσικού προσώπου (στο πλαίσιο του ΓΚΠΔ), η ιδιωτικότητα αναφέρεται στην παροχή της δυνατότητας στο υποκείμενο επεξεργασίας να επιλέξει με μεγαλύτερη ευελιξία το ποιόν θα εξουσιοδοτήσει, για ποιού είδους επεξεργασία θα δώσει την εξουσιοδότηση και για ποιά δεδομένα θα την δώσει. Η ιδιωτικότητα δηλαδή έρχεται *να περιορίσει την προσέγγιση της "όλα ή τίποτα" επεξεργασίας δεδομένων*, επιβάλλοντας στον υπεύθυνο επεξεργασίας να δώσει τις προαναφερθείσες επιλογές επεξεργασίας στο υποκείμενό της. Υπο αυτό το πρίσμα, το Άρθρο 25 του ΓΚΠΔ αναφέρεται συγκεκριμένα στην υποχρέωση των υπευθύνων επεξεργασίας (αλλά και κατ'επέκταση των παρόχων τεχνολογιών σε αυτούς, βλ. [30]) να παρέχουν την δυνατότητα της ιδιωτικότητας στο υποκείμενο της επεξεργασίας. Επίσης προϋποθέτει ότι η σωστή εφαρμογή της συγκεκριμένης απαιτήσεως επιτυγχάνεται δια του σχεδιασμού των συστημάτων και εκ της καταλλήλου προεπιλογής ρυθμίσεων της λειτουργίας του συστήματος επεξεργασίας. Αντιθέτως, δεν θέτει παρόμοιους περιορισμούς στην επιλογή και εφαρμογή των μέτρων προστασίας των δεδομένων.

Τέλος, θα πρέπει να επισημανθεί πως η προστασία δεδομένων και η ιδιωτικότητα είναι έννοιες διαφορετικές από την προστασία ιδιωτικότητας. Η τελευταία είναι *η νομική κατοχύρωση της ιδιωτικότητας* κατά την επεξεργασία δεδομένων προσωπικού χαρακτήρα.

[5] Βλέπε **Κεφάλαιο 4: Παραβιάσεις Ιδιωτικότητας** για επεξήγηση των εννοιών

Προβλήματα που εγείρονται από την εφαρμογή της

Το ζήτημα της ιδιωτικότητας έχει αποτελέσει βασικό θέμα συζητήσεων μεταξύ του νομοθέτη και των τεσσάρων βασικών παραγόντων της επεξεργασίας δεδομένων προσωπικού χαρακτήρα ([23]): τις κυβερνήσεις, τις εταιρείες, τους ειδικούς ασφαλείας και τα υποκείμενα επεξεργασίας.

Οι κυβερνήσεις έχουν έννομο συμφέρον στο να υπάρχει η δυνατότητα να επεξεργάζονται άνευ περιορισμών τα δεδομένα φυσικών προσώπων, για λόγους αστυνόμευσης αλλά και αποκατάστασης της έννομης τάξεως. Επισης, απο την άνευ περιορισμού επεξεργασία προσωπικων δεδομένων ωφελείται ο κλάδος της υγείας, ο οποίος αξιοποιεί ευαίσθητα προσωπικά δεδομένα, σχετιζόμενα με την υγεία των υποκειμένων, για να διεξάγει έρευνες και μελέτες σχετικές με την υγεία του κοινωνικού συνόλου. Γι'αυτή την κατηγορία παραγόντων όμως, ο ΓΚΠΔ προβλέπει εξαιρέσεις που στοχεύουν στην διευκόλυνση του έργου τους. Δίνεται ιδιαίτερη βαρύτητα στην διατύπωση των εξαιρέσεων, βεβαίως, έτσι ώστε να περιοριστεί το ενδεχόμενο της καταχρήσεως των ελευθεριών που προκύπτουν από αυτές ([13]).

Η επόμενη κατηγορία παραγόντων είναι οι ιδιωτικές εταιρείες τεχνολογίας (ή οι δημόσιες, που δεν δραστηριοποιούνται στους τομείς της αστυνόμευσης ή της υγείας) οι οποίες έχουν τον ρόλο των παρόχων τεχνολογιών ή του υπευθύνου επεξεργασίας. Οι εταιρείες τεχνολογίας έχουν ασκήσει την μεγαλύτερη, ίσως, πίεση στον νομοθέτη για την *μη νομοθέτηση* ή την νομοθέτηση *χαλαρών κανόνων* προστασίας της ιδιωτικότητας των υποκειμένων. Τα επιχειρήματα που χρησιμοποιούν οι εταιρείες είναι κατά βάση το κόστος υλοποίησης τέτοιων συστημάτων, η απώλεια "ανταγωνιστικού πλεονεκτήματος" λόγω της μη δυνατότητας ελεύθερης επεξεργασίας προσωπικών δεδομένων και η δυσκολία εφαρμογής τέτοιων μέτρων σε καινοτόμες τεχνολογίες ([23]).

Θεμελιώδες παράδειγμα περιπτώσεως κατά την οποία επιχειρηματολογείται πως είναι δύσκολο να εφαρμοστεί η ιδιωτικότητα (ειδικά δια του σχεδιασμού) είναι η περίπτωση της ανίχνευσης προσώπων μέσω τεχνολογίας ("*face detection*"). Μια χρήση αυτής της τεχνολογίας είναι η αυτόματη ανίχνευση προσώπων σε κέντρα αγοράς και η κατάταξή τους σε ηλικιακές, φυλετικές και άλλου τύπου ομάδες, ώστε να γίνει στοχευμένη διαφήμιση προϊόντων κατά την παραμονή τους στα κέντρα της αγοράς, με σκοπό την αύξηση των πωλήσεων ([27]). Εταιρείες που έκαναν χρήση αυτής της τεχνολογίας και εφήρμοσαν όλα τα μέτρα ιδιωτικότητας που ορίζονται από την νομοθεσία, εν τέλει αναγκάστηκαν να αποσύρουν τα προϊόντα τους γιατί η συγκεκριμένη χρήση δεν έδινε την δυνατότητα στα υποκείμενα επεξεργασίας να μην δώσουν την συγκατάθεσή τους για την επεξεργασία των προσωπικών τους δεδομένων (ανίχνευση του προσώπου τους). Αντίστοιχο πρόβλημα δημιουργείται από την αναγνώριση προτύπων (σχημάτων) μέσω κάμερας στις εφαρμογές της αυτόνομης οδήγησης. Ενώ τα αυτόνομα οχήματα χρησιμοποιούν κάμερες για τον προσανατολισμό και την κίνησή τους, δεν υπάρχει εγγύηση πως δεν θα λάβει χώρα αναγνώριση προσώπων μέσω των καμερών. Αυτό συνεπάγεται παραβίαση της ελευθερίας των υποκειμένων να μην δώσουν έγκριση για επεξεργασία των προσωπικών τους δεδομένων και, συνεπώς, καθιστά σχεδόν παράνομη την χρήση καμερών γι'αυτόν τον σκοπό.

Η επόμενη κατηγορία η οποία αντιτίθεται εν γένει στην ιδιωτικότητα των υποκειμένων είναι οι ειδικοί ασφαλείας πληροφοριών. Ο λόγος γι'αυτούς είναι αρκετά βασικός, γιατί επηρεάζει στο μέγιστο την άσκηση των καθηκόντων τους: για τον έλεγχο ασφαλείας των υπολογιστικών συστημάτων χρειάζεται να γνωρίζουν τις συνήθειες και τις ενέργειες των χρηστών των συστημάτων. Αυτές οι πληροφορίες χρησιμοποιούνται πολύ συχνά για να φτιάξουν το "*ψυχογράφημα*" ("*προφίλ*") του κάθε χρήστη, έτσι ώστε σε περίπτωση παραβίασης της ασφαλείας των συστημάτων στα οποία είναι υπεύθυνοι, να μπορούν να ιχνηλατήσουν την παραβίαση για να υπολογίσουν την έκταση της πιθανής ζημίας ή και τον υπαίτιο της παραβίασης. Κατά τα

φαινόμενα, με την εφαρμογή της ιδιωτικότητας αλλά και των δικαιωμάτων που απορρέουν από αυτήν, το έργο των ειδικών ασφαλείας δυσχεραίνεται (ή και καθίσταται εξ'ολοκλήρου αδύνατο).

Στον αντίποδα όλων αυτών βρίσκονται τα υποκείμενα επεξεργασίας, τα οποία όμως δεν φαίνεται να έχουν ενιαία στάση όσον αφορά στην επεξεργασία των προσωπικών τους δεδομένων. Απο την μία πλευρά, ως καταναλωτές προϊόντων, δεν επιθυμούν τον αποκλεισμό από οικονομικές ευκαιρίες οι οποίες προκύπτουν από επεξεργασία των προσωπικών τους δεδομένων (στοχευμένη διαφήμιση). Απο την άλλη, πάλι ως καταναλωτές προϊόντων, δεν επιθυμούν τα δεδομένα τους να πωλούνται σε τρίτους και να γίνονται στόχοι διαφημίσεων που δεν τους αφορουν (πχ μέσω ανεπιθύμητης αλληλογραφίας). Επίσης, ως άτομα με αυξημένο το αίσθημα της φυσικής ασφαλείας, υπάρχουν αρκετοί που κατανοούν τους κινδύνους που υπάρχουν όταν άγνωστοι (πχ υπάλληλοι εταιρειών) έχουν πρόσβαση σε στοιχεία όπως διεύθυνση κατοικίας, τραπεζικές καταθέσεις ή θρησκευτικές πεποιθήσεις (πόσο μάλλον ο συνδυασμός αυτών).

Όπως βλέπουμε, οι διάφορες κατηγορίες ενδιαφερομένων κι εμπλεκομένων βλέπουν το ζήτημα της ιδιωτικότητας από διαφορετικές οπτικές γωνίες. Αυτό έχει οδηγήσει στην κατάσταση στην οποία η ιδιωτικότητα δεν φαίνεται να επικρατεί ως κοινωνικό φαινόμενο (όπως είδαμε στο αντίστοιχο εδάφιο), λόγω συγκρούσεως συμφερόντων. Γι'αυτό κατέστη σαφές πως η ιδιωτικότητα έπρεπε να ρυθμιστεί νομικά, παρά να αυτορυθμιστεί αφημένη στις πιέσεις της κοινωνικής δυναμικής.

Περι ιδιωτικότητας δια του σχεδιασμού

Η αρχική ανάγκη να προστατευτεί το υποκείμενο επεξεργασίας προσωπικών δεδομένων μέσω της ιδιωτικότητας προκάλεσε, όπως είδαμε, αντιδράσεις και προβλήματα στην εφαρμογή της. Λόγω της πολυπλοκότητας των λύσεων και του κόστους που απαιτούνταν, αλλά και της απώλειας των πλεονεκτημάτων από πλευράς εταιρειών, των πιέσεων από τις κυβερνήσεις και τους ειδικούς ασφαλείας (και σε μερικές περιπτώσεις η αδιαφορία από την πλευρά των υποκειμένων, εξαιτίας του φαινομενικά απλού επιχειρήματος "δεν έχω κάτι να κρύψω" – [23]), πρακτικά η ιδιωτικότητα δεν εφαρμοζόταν, ούτε εξ'ολοκλήρου ούτε ουσιαστικά ([23], [28]).

Έτσι, κατέστη σαφές πως το δικαίωμα στην ιδιωτικότητα θα έπρεπε να νομοθετηθεί. Επιπλέον αυτού, αποφασίστηκε ότι θα έπρεπε η ιδιωτικότητα, ως δυνατότητα ενός συστήματος, να λαμβάνεται υπ'οψιν από τα αρχικά στάδια του σχεδιασμού του συστήματος. Θα έπρεπε δηλαδή να νομοθετηθεί η χρήση της αρχής της ιδιωτικότητας δια του σχεδιασμού. Αυτό συνέβει με την εισαγωγή του ΓΚΠΔ, μέσω του Άρθρου 25 ([13], [29], [30]).

Βάσει λοιπόν του Άρθρου 25 του ΓΚΠΔ, η εφαρμογή της αρχής της ιδιωτικότητας δια του σχεδιασμού καθίσταται υποχρεωτική για όλα τα συστήματα επεξεργασίας προσωπικών δεδομένων. Αν και στο Άρθρο 25 δεν αναφέρεται, ούτε αποσαφηνίζεται, τι εννοείται με τον όρο "ιδιωτικότητα δια του σχεδιασμού", υπήρξε διεθνής βιβλιογραφία η οποία φαίνεται να επηρέασε το σκεπτικό και τις προθέσεις του νομοθέτη (με αξιοσημείωτη την [11])

Η αρχή αυτή της ιδιωτικότητας φαίνεται να είχε μελετηθεί αρκετά πριν συνταχθεί το Άρθρο 25 του ΓΚΠΔ, πράγμα που σημαίνει πως ο νομοθέτης γνώριζε οτι η αρχή μπορεί να εφαρμοστεί σε ευρύ φάσμα αναγκών της κοινωνίας.

Για παράδειγμα, βλέπουμε πως η αρχή είχε μελετηθεί για την προστασία ιατρικών δεδομένων υποκειμένων, κατά την πρόσβασή τους απο απεριόριστο αριθμό ανθρώπων και τον διαμοιρασμό των δεδομένων μεταξύ τους, χωρίς να προκύπτει θέμα παραβίασης της ιδιωτικότητας των υποκειμένων ([31]).

Άλλο παράδειγμα είναι οι περιπτώσεις των μικροφώνων και μεγαφώνων κλειστών κυκλωμάτων τηλεοράσεως ("CCTV"), ανιχνευτές σώματος, καθώς και άλλων τεχνολογιών που εισβάλλουν στην ιδιωτική ζωή των ανθρώπων. Εκεί επίσης έχουμε μελέτες που προτείνουν την χρήση της ιδιωτικότητας δια του σχεδιασμού για την αποφυγή παραβιάσεων ([32]).

Γενικά, λοιπόν, η διεθνής βιβλιογραφία παρέθεσε αρκετές μελέτες και παρατηρήσεις σχετικά με την εφαρμογή της αρχής στα χρόνια πριν την νομοθέτηση του Άρθρου 25, τόσο για να γίνει κατανοητή από τον νομοθέτη η αρχή της ιδιωτικότητας δια του σχεδιασμού όσον αφορά νέες τεχνολογίες (πχ [33]), όσο και για να υπάρξει μια προσέγγιση προς τις εταιρείες, δείχνοντας τον δρόμο του οικονομικού και καινοτομικού πλεονεκτήματος της χρήσεως της αρχής (πχ [26]).

Έτσι, βάσει της βιβλιογραφίας και της αυξανόμενης εμπειρίας που αποκομίζεται από την εφαρμογή της αρχής στην πράξη, φαίνεται ότι η στάση της αγοράς άρχισε να αλλάζει προς την αποδοχή της. Χαρακτηριστικό παράδειγμα είναι η προσπάθεια να προτυποποιηθούν οι ενέργειες για την εφαρμογή της αρχής στον σχεδιασμό των συστημάτων, παράγοντας διεθνή πρότυπα ασφαλείας ([34]).

Τέλος, είναι σημαντικό να τονίσουμε πως, ανατρέχοντας στην διεθνή βιβλιογραφία που αφορά στην ιδιωτικότητα δια του σχεδιασμού, φαίνεται πως το μεγαλύτερο μέρος της επηρεάστηκε από τις επτά (7) αρχές της ιδιωτικότητας δια του σχεδιασμού που όρισε η **Ann Cavoukian**, στην αναφορά της "*Privacy by Design: The 7 Foundational Principles*" ([11]) . Επειδή η συμβολή αυτή είναι από τις πιο κρίσιμες στην διεθνή βιβλιογραφία, θα εξεταστεί ξεχωριστά στο αντίστοιχο κεφάλαιο (Κεφάλαιο 5).

4 *Παραβιάσεις ιδιωτικότητας*

Εισαγωγή

Η ασφάλεια πληροφοριών, τμήμα της οποίας είναι και η ιδιωτικότητα, βασίζεται στο τρίπτυχο της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας (στην διεθνή βιβλιογραφία αναφέρονται συλλογικά ως *CIA – Confidentiality, Integrity* και *Availability*). Με αυτά εννοούμε:

- **εμπιστευτικότητα (confidentiality)**: είναι η ιδιότητα που έχουν τα δεδομένα να μην μπορούν να προσπελαστούν απο οντότητες (προγράμματα ή πρόσωπα) οι οποίες δεν έχουν την κατάλληλη εξουσιοδότηση

- **ακεραιότητα (integrity)**: είναι η ιδιότητα που έχουν τα δεδομένα να μην μπορούν να τροποποιηθούν εκουσίως ή ακουσίως χωρίς την εξουσιοδότηση του ιδιοκτήτη ή του διαχειριστή τους. Παράδειγμα μη εξουσιοδοτημένης ακούσιας τροποποιήσεως είναι η εισαγωγή σφαλμάτων από το περιβάλλον (πχ αστοχία υλικού), αντικείμενο με το οποίο ασχολείται κυρίως ο κλάδος της διόρθωσης σφαλμάτων των τηλεπικοινωνιών. Παράδειγμα μη εξουσιοδοτημένης εκούσιας τροποποιήσεως είναι η μεταβολή (εισαγωγή, ενημέρωση ή διαγραφή) των δεδομένων απο επιτιθέμενο

- **διαθεσιμότητα (availability)**: είναι η ιδιότητα που έχουν τα δεδομένα να μην μπορούν να καταστούν απροσπέλαστα χωρίς την εξουσιοδότηση του ιδιοκτήτη ή του διαχειριστή τους. Κύριος τρόπος άρσης της διαθεσιμότητας των δεδομένων μιας βάσης είναι η διαγραφή τους ή η καταστροφή του αποθηκευτικού τους χώρου

Εδώ οφείλουμε να διευκρινίσουμε πως η διαφορά της διαγραφής όσον αφορά στην ακεραιότητα και την διαθεσιμότητα είναι, οτι στην μεν *ακεραιότητα* η διαγραφή γίνεται σε *τμήμα δεδομένων* ώστε τα δεδομένα να μείνουν *ανολοκλήρωτα* (αλλά υπάρχει ακόμα πρόσβαση σε αυτά), ενώ η διαγραφή στην *διαθεσιμότητα* αφορά στην *καθολική διαγραφή* των δεδομένων έτσι ώστε ο εξουσιοδοτημένος χρήστης να απωλέσει εξ'ολοκλήρου την πρόσβαση σε αυτά.

Η ιδιωτικότητα αποτελεί ειδική περίπτωση της εμπιστευτικότητας δεδομένων. Υπάρχει όμως μία ιδιαιτερότητα όσον αφορά στην ιδιωτικότητα - ενώ και οι δύο (εμπιστευτικότητα και ιδιωτικότητα) αφορούν στο να μην μπορούν να προσπελαστούν οι πληροφορίες από μη εξουσιοδοτημένες οντότητες, η ιδιωτικότητα επιβάλλει επιπλέον τον περιορισμό οτι ακόμα και οι εξουσιοδοτημένες οντότητες *έχουν περιορισμό στον τύπο της επεξεργασίας* που μπορούν να τελέσουν στις πληροφορίες. Δηλαδή, ενώ η εμπιστευτικότητα αναφέρεται στην επίτρεψη προσβάσεως στα δεδομένα αποκλειστικά στις εξουσιοδοτημένες οντότητες, η ιδιωτικότητα αναφέρεται τόσο στον περιορισμό προσβάσεως όσο και τον περιορισμό της επεξεργασίας από μέρους των εξουσιοδοτημένων οντοτήτων.

Η τελευταία διαφορά είναι πολύ σημαντική κι επισημαίνεται σαφώς στην διεθνή νομοθεσία (πχ ΓΚΠΔ) διότι πλέον αλλάζει το πρίσμα υπό το οποίο εφαρμόζεται ένα σύστημα επεξεργασίας δεδομένων με τον περιορισμό της τήρησης της ιδιωτικότητας. Πιο συγκεκριμένα, η εμπιστευτικότητα είθισται να αφορά στην προστασία των πληροφοριών με τεχνολογικά μέσα (πχ μέσω κρυπτογράφησής τους). Όμως, όταν μια εξουσιοδοτημένη οντότητα αποκτήσει πρόσβαση στις πληροφορίες, μπορεί να τις διαχειριστεί κατά βούληση, χωρίς περιορισμούς. Αντιθέτως, όταν τηρείται ο περιορισμός της ιδιωτικότητας, τότε οι τύποι της επεξεργασίας στις οποίες μπορεί να υπόκεινται οι πληροφορίες *θα πρέπει να είναι προσυμφωνημένες*. Συνεπώς, δεν θα πρέπει να μπορεί κάποια εξουσιοδοτημένη οντότητα να επεξεργαστεί τις πληροφορίες με τρόπο που δεν έχει προσυμφωνηθεί, καθώς και με τρόπο για τον οποίο δεν συναινεί το υποκείμενο της επεξεργασίας.

Η τελευταία πρόταση είναι πάρα πολύ σημαντική γιατί έχει μέγιστο αντίκτυπο στον σχεδιασμό των συστημάτων, καθώς θα *πρέπει τα συστήματα να περιορίζουν όχι μόνο το ποιός έχει πρόσβαση σε ποιά δεδομένα, αλλά και το τι μπορεί να κάνει ο κάθε εξουσιοδοτημένος με τα δεδομένα αυτά*. Ακόμα μεγαλύτερος προβληματισμός υπάρχει και στην (νομοθετική κυρίως) απαίτηση να πρέπει να δίνεται η δυνατότητα σε κάθε υποκείμενο επεξεργασίας να άρει την συναίνεση προς επεξεργασία. Η άρση της συναινέσεως μπορεί να γίνει τόσο για την εξ ολοκλήρου επεξεργασία των δεδομένων του υποκειμένου (πχ μέσω διαγραφής των δεδομένων του), όσο και την κατά περίπτωση επεξεργασία των δεδομένων του - δηλαδή θα πρέπει να μπορεί να άρει την συναίνεση σε κάποιους απο τους τρόπους επεξεργασίας των δεδομένων του ενώ συναινεί σε άλλους τρόπους επεξεργασίας. Αυτή η άρση της συναινέσεως θα πρέπει να γίνεται σεβαστή απο το κάθε σύστημα επεξεργασίας σε εύλογο χρονικό διάστημα (συνεπώς, αυτοματοποιημένα). Αυτός ο περιορισμός αυξάνει αρκετά την πολυπλοκότητα της υλοποιήσεως των συστημάτων επεξεργασίας προσωπικών πληροφοριών και, γι'αυτό άλλωστε, πολλά συστήματα δεν εφαρμόζουν επαρκώς τον περιορισμό της ιδιωτικότητας όπως απαιτείται και απο τον νόμο.

Η ελλιπής εφαρμογή της ιδιωτικότητας έχει ως συνέπεια να καθίσταται εύκολη η παραβίασή της. Οι δύο βασικοί τρόποι παραβίασης της ιδιωτικότητας είναι εκείνη ως προς το εύρος των προσπελάσιμων πληροφοριών κι εκείνη ως προς τον τρόπο επεξεργασίας. Οι επιτιθέμενοι βασιζόμενοι στους δύο αυτούς τρόπους παραβίασης της ιδιωτικότητας μπορούν να ταυτοποιήσουν, κατηγοριοποιήσουν, εξάγουν συμπεράσματα και γενικά καταχραστούν την υπό επίθεση βάση, ακόμα κι αν εκείνη προστατεύεται μερικώς (πχ με την χρήση κρυπτογραφίας). Οι δύο αυτοί τρόποι αναλύονται στις επόμενες ενότητες.

Παραβίαση ως προς το εύρος προσπελάσιμων πληροφοριών

Ο συγκεκριμένος τύπος παραβίασης της ιδιωτικότητας βασίζεται στην δυνατότητα που έχουν οι εξουσιοδοτημένες οντότητες να επεξεργάζονται χωρίς περιορισμούς όλο το εύρος των διαθέσιμων

πληροφοριών. Δηλαδή, η επίθεση καθίσταται εφικτή *όταν δεν τηρείται η αρχή της αναγκαίας γνώσης* για τις εξουσιοδοτημένες οντότητες.

Για παράδειγμα, τέτοιου είδους παραβίαση μπορεί να συμβεί όταν μια εξουσιοδοτημένη οντότητα μπορεί να κάνει ερωτήματα σε όλους τους διαθέσιμους πινακες (σχέσεις) μιας σχεσιακής βάσης δεδομένων. Αν, λοιπόν, η συγκεκριμένη οντότητα χρειάζεται πρόσβαση μόνο σε ένα μέρος των δεδομένων για να τελέσει την λειτουργία της, με το να έχει πρόσβαση σε όλους τους πίνακες της βάσης, δίνουμε την δυνατότητα να κάνει ερωτήματα σε δεδομένα στα οποία εξ αρχής δεν θα έπρεπε.

Ένας συνήθης λόγος, για τον οποίον συμβαίνει το να δίνεται πρόσβαση σε όλες τις διαθέσιμες πληροφορίες σε μια οντότητα, είναι επειδή κατ'αυτόν τον τρόπο είναι πιο εύκολη η ρύθμιση και διαχείριση της βάσης, σε σχέση με το να ρυθμιστεί με λεπτομέρεια η πρόσβαση σε κάθε πίνακα ξεχωριστα. Αυτό όμως καθιστά τον συγκεκριμένο τύπο παραβίασης αρκετά εύκολο σε επιθέσεις εκ των έσω.

Παραβίαση ως προς τον τρόπο επεξεργασίας

Αυτός ο τύπος της παραβίασης οφείλεται στην δυνατότητα που έχουν οι εξουσιοδοτημένες οντότητες να κάνουν απεριόριστα σε μορφή ερωτήματα σε πίνακες με δεδομένα. Αν, δηλαδή, η μορφή (όχι ο αριθμός) των ερωτημάτων που κάνει η οντότητα στα δεδομένα δεν είναι προσυμφωνημένη, τότε *η οντότητα μπορεί να κάνει ερωτήματα τα οποία να ενεργούν σε δεδομένα με τρόπο για τον οποίον δεν έχει συμφωνήσει εξ'αρχής το υποκείμενο επεξεργασίας*.

Τυπικό παράδειγμα τέτοιας παραβίασης είναι πως υποκείμενα επεξεργασίας συναινούν ώστε τα δεδομένα τους να χρησιμοποιηθούν για την τέλεση ενός συμβολαίου, αλλά δεν συναινούν ώστε τα δεδομένα τους να χρησιμοποιηθούν για να δημιουργηθούν ψυχογραφήματα χρηστών (τα "προφίλ" τους δηλαδή), τα οποία χρησιμοποιούνται για στοχευμένη διαφήμιση. Αν λοιπόν το σύστημα έχει σχεδιαστεί ώστε να μην επιβάλει προσυμφωνημένα ερωτήματα στα δεδομένα μιας βάσης, τότε οποιαδήποτε οντότητα με πρόσβαση στην βάση μπορεί να παραβιάσει την ιδιωτικότητα των υποκειμένων. Και αυτός ο τύπος παραβίασης είναι σχετικά εύκολος να συμβεί σε επιθέσεις εκ των έσω.

5 *Ιδιωτικότητα δια του σχεδιασμού*

Εισαγωγή

Παρά του οτι η ιδιωτικότητα ως έννοια υπάρχει εδώ και αρκετό καιρό, η έννοια της εξασφάλισής της κατά την διάρκεια του σχεδιασμού συστημάτων δεν έχει μελετηθεί εκτενώς. Υπήρχε, βεβαίως, αναφορά σε αυτήν την απαίτηση σε διάφορες νομοθετικές ρυθμίσεις[6], αλλά δεν υπήρχε εκτενής μελέτη για το πως μπορεί η ιδιωτικότητα να εξασφαλιστεί κατά την φάση του σχεδιασμού. Σταθμός στην εξέλιξή της υπήρξε ο ορισμός της ιδιωτικότητας δια του σχεδιασμού απο την Άν Καβούκιαν (Ann Cavoukian), την πρώην επίτροπο Πληροφοριών και Ιδιωτικότητας της καναδικής επαρχίας του Οντάριο. Η Άν Καβούκιαν εξέδωσε το 2009 μια αναφορά ([11]) η οποία περιέχει επτά (7) αρχές τις οποίες οφείλουν να τηρούν τα συστήματα επεξεργασίας δεδομένων προσωπικού χαρακτήρα, ώστε να πληρούται η ιδιωτικότητα δια του σχεδιασμού. Η αναφορά αυτή είχε τόσο μεγάλο αντίκτυπο στην μελέτη της ιδιωτικότητας, ώστε αποτέλεσε βασική πηγή για την σύνταξη συγκεκριμένων άρθρων του ΓΚΠΔ[7].

Στο παρόν κεφάλαιο θα εξετάσουμε τις αρχές της ιδιωτικότητας δια του σχεδιασμού της Άν Καβούκιαν, καθώς και κάποιες μελέτες και έρευνες που έχουν γίνει στον τομέα της εφαρμογής της στις βάσεις δεδομένων.

Οι 7 αρχές της Άν Καβούκιαν

Όπως έχουμε αναφέρει, η Άν Καβούκιαν όρισε επτά (7) αρχές τις οποίες πρέπει να πληρούν τα συστήματα επεξεργασίας δεδομένων προσωπικού χαρακτήρα, ώστε να θεωρείται ότι τα συστήματα αυτά είναι σχεδιασμένα για να προστατεύουν την ιδιωτικότητα των υποκειμένων επεξεργασίας. Οι αρχές αυτές είναι:

[6] Βλέπε το σημείο 46 της οδηγίας 95/46/ΕΚ ([12])
[7] Βλέπε άρθρο 25 του ΓΚΠΔ ([13])

Προληπτικά όχι αντιδραστικά - αποτρεπτικά, όχι θεραπευτικά

Η αρχή αυτή (*Proactive not Reactive; Preventative not Remedial*) δηλώνει πως τα οργανωτικά και τεχνικά μέτρα τα οποία θα υλοποιηθούν για να προστατεύσουν την ιδιωτικότητα, θα πρέπει να εφαρμόζονται με τέτοιον τρόπο ώστε να ελαττώνουν την πιθανότητα ή να αποτρέπουν το ενδεχόμενο να παραβιαστεί η ιδιωτικότητα (*προληπτικά και αποτρεπτικά*). Επίσης, η αρχή ορίζει ότι αφού λάβει χώρα κάποια παραβίαση της ιδιωτικότητας, τα υπάρχοντα μέτρα θα πρέπει να στοχεύουν στο να ελαττώνουν τις συνέπειες της παραβίασης αυτής (*αντιδραστικά και θεραπευτικά*).

Ιδιωτικότητα ως η προκαθορισμένη ρύθμιση

Η αρχή αυτή (*Privacy as the Default Setting*) τονίζει την ανάγκη που υπάρχει του να προστατεύεται η ιδιωτικότητα από το σύστημα επεξεργασίας πληροφοριών ως προκαθορισμένη λειτουργία. Δηλαδή, η αρχική συμπεριφορά ενός συστήματος επεξεργασίας πληροφοριών θα πρέπει να προστατεύει την ιδιωτικότητα των υποκειμένων επεξεργασίας, και θα πρέπει να δίνεται η δυνατότητα στα υποκείμενα να "χαλαρώνουν" τα μέτρα προστασίας κατά το δοκούν. Η αρχή αυτή βασίζεται στην στατιστική παρατήρηση πως τα υποκείμενα επεξεργασίας χρησιμοποιούν τα συστήματα χωρίς να αλλάζουν τις αρχικές ρυθμίσεις. Συνεπώς με αυτές τις (αρχικές) ρυθμίσεις θα πρέπει να προστατεύεται η ιδιωτικότητά τους.

Ιδιωτικότητα ενσωματωμένη στον σχεδιασμό

Στην αρχή αυτή (*Privacy Embedded into Design*) καθοδηγείται ο σχεδιαστής συστημάτων επεξεργασίας δεδομένων προσωπικού χαρακτήρα στο να σχεδιάζει τα συστήματα έχοντας υπ'όψιν του πως πρέπει να προστατεύεται η ιδιωτικότητα των υποκειμένων. Η αρχή αυτή προτείνεται γιατί είθισται να δίνεται μεγαλύτερη βάση στην λειτουργικότητα των συστημάτων και η εφαρμογή της ιδιωτικότητας να προστίθεται εκ των υστέρων. Επειδή όμως η ιδιωτικότητα θέτει περιορισμούς στην επεξεργασία, μετά τον αρχικό σχεδιασμό των συστημάτων καθίσταται δύσκολο να εφαρμοστεί όπως θα έπρεπε (πχ επειδή χρειάζεται να επανασχεδιαστεί μεγάλο τμήμα του συστήματος, λόγω έλλειψης χρόνου κτλ). Αυτό έχει ως συνέπεια να παραλείπονται βασικά τμήματά της, κατά την εκ των υστέρων εφαρμογή της στο σύστημα. Εδώ θα πρέπει να διευκρινιστεί πως η αρχή δεν περιορίζεται στα τεχνικά μέτρα εφαρμογής της ιδιωτικότητας, αλλά αφορά και στα οργανωτικά μέτρα. Δηλαδή κατά την διάρκεια του σχεδιασμού των συστημάτων θα πρέπει να λαμβάνονται υπ'όψιν και τα οργανωτικά μέτρα - δηλαδή του ποιός θα έχει πρόσβαση σε ποιά δεδομένα, και υπό ποιές προϋποθέσεις. Επειδή τα οργανωτικά μέτρα είναι ευκολότερα στην υλοποίησή τους εκ των υστέρων, σε σχέση με τα τεχνικά μέτρα (δηλαδή είναι ευκολότερο να οριστούν περιορισμοί πρόσβασης στους ανθρώπους παρά να αλλάξει το σύστημα για να περιορίζει την επεξεργασία), φαίνεται πως η η ιδιωτικότητα προτιμάται να υλοποιείται με τέτοιου τύπου μέτρα, κάτι που την καθιστά αρκετά ελλιπή. Η συγκεκριμένη αρχή, λοιπόν, έρχεται να τονίσει πως τα μέτρα (οργανωτικά και τεχνικά) θα πρέπει να αποφασίζονται κατά την φάση του σχεδιασμού. Έτσι επιτυγχάνεται το να λαμβάνονται τα κατάλληλα μέτρα προτού το σύστημα τεθεί σε κανονική λειτουργία, με αποτέλεσμα να εφαρμόζεται η ιδιωτικότητα από την στιγμή που το σύστημα θα συλλέξει τα πρώτα δεδομένα.

Πλήρης λειτουργικότητα (θετικό ισοζύγιο, όχι μηδενικό)

Η συγκεκριμένη αρχή (*Full functionality - Positive-Sum, not Zero-Sum*) έρχεται να τονίσει πως πρέπει να αντιμετωπίζεται η ιδιωτικότητα απο τους ιδιοκτήτες, αρχιτέκτονες και χρήστες των συστημάτων επεξεργασίας δεδομένων προσωπικού χαρακτήρα. Δηλαδή είναι αρχή η οποία απευθύνεται στην αλλαγή νοοτροπίας των ενδιαφερομένων. Ο λόγος που τίθεται είναι γιατί υπάρχει η τάση να αντιμετωπίζεται η ιδιωτικότητα περισσότερο ως πρόβλημα παρά ως θετική ιδιότητα του συστήματος. Αυτό συναντάται κυρίως στους τομείς της ασφάλειας και της εξόρυξης δεδομένων, όπου η πρόσβαση άνευ περιορισμών στα συλλεχθέντα δεδομένα θεωρείται νευραλγικής σημασίας για την επίτευξη του στόχου της επεξεργασίας. Αυτή η νοοτροπία όμως θεωρείται ξεπερασμένη και από μόνη της δημιουργεί προβλήματα στην ενσωμάτωση της ιδιωτικότητας στα αντίστοιχα συστήματα.

Ολοκληρωμένη Ασφάλεια - Προστασία σε όλο τον κύκλο ζωής

Σύμφωνα με αυτήν την αρχή (*End-to-End Security - Full Lifecycle Protection*), θεωρώντας δεδομένο πως έχει ενσωματωθεί η ιδιωτικότητα στον σχεδιασμό του συστήματος επεξεργασίας, οι πληροφορίες θα πρέπει να προστατεύονται από την φάση της συλλογής τους έως και την διαγραφή/καταστροφή τους. Δηλαδή, η αρχή τονίζει πως καθίσταται σημαντικό το να προστατεύονται οι πληροφορίες καθ'όλην την διάρκεια της ζωής τους - αυτό συμπεριλαμβάνει και το στάδιο της συλλογής τους. Η συγκεκριμένη αρχή είναι πολύ σημαντική γιατί αποσαφηνίζει τα όρια εντός των οποίων οφείλουμε να σκεφτόμαστε το πως θα εφαρμόσουμε την ιδιωτικότητα των υποκειμένων επεξεργασίας. Αυτό, βεβαίως, είναι σημαντικό γιατί παρατηρείται πως δίνεται βάση στην προστασία των δεδομένων μόνο κατά τις επιχειρησιακές φάσεις (ενεργό επεξεργασία τους) ενώ στις υπόλοιπες φάσεις (συλλογή, αντίγραφα, διαγραφή) δεν δίνεται ιδιαίτερη βάση.

Διαφάνεια

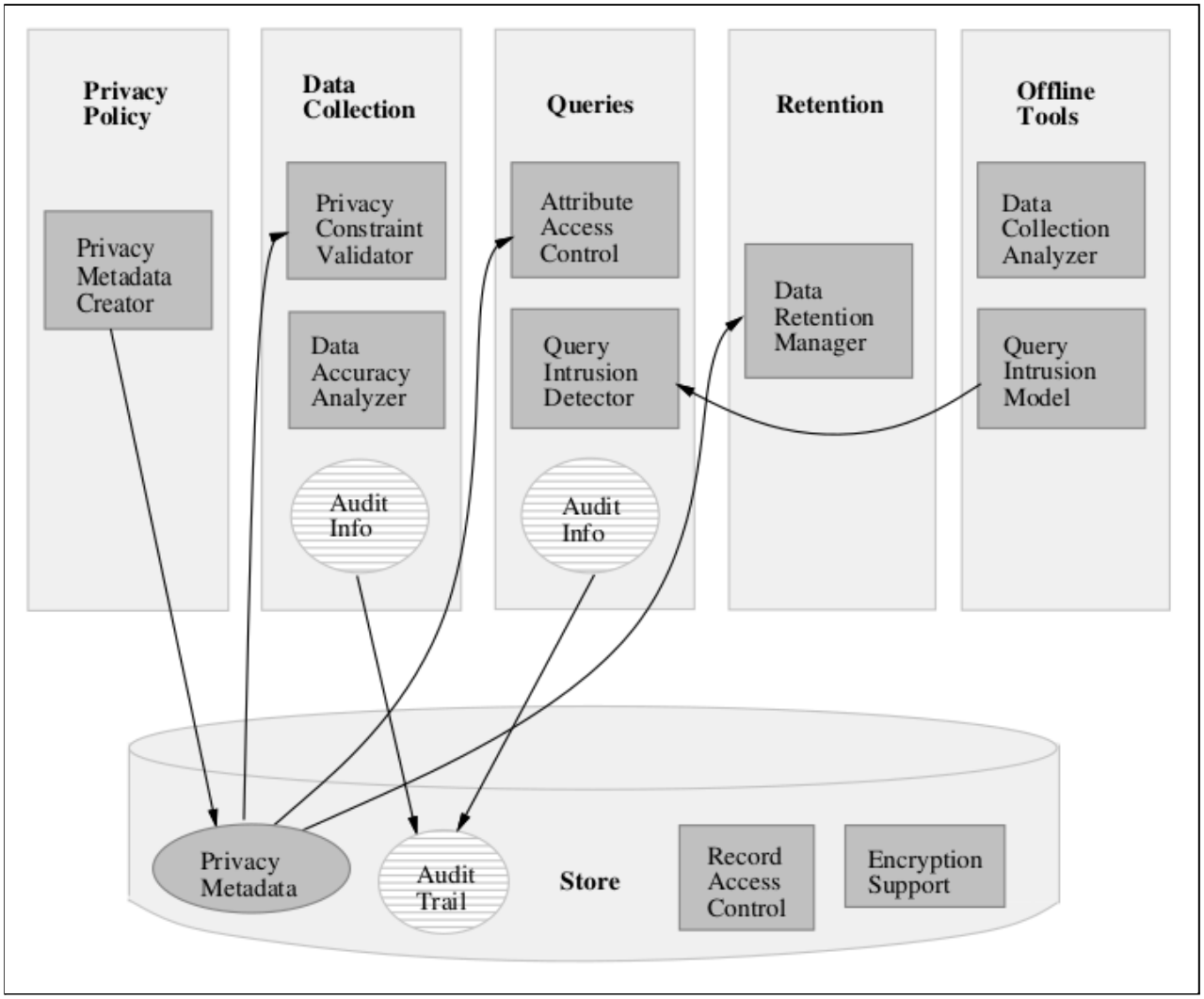
Η αρχή αυτή (*Visibility and Transparency – Keep it Open*) τονίζει την ανάγκη του να μπορεί να ελέγχει και να επιβεβαιώνει ο κάθε ενδιαφερόμενος (πχ χρήστης, υπεύθυνος επεξεργασίας ή νομικός ελεγκτής) τον τρόπο με τον οποίο το κάθε σύστημα εφαρμόζει την ιδιωτικότητα των υποκειμένων επεξεργασίας. Ο έλεγχος θα πρέπει να είναι εύκολος να γίνει και θα πρέπει να αποδεικνύει πως το κάθε σύστημα τηρεί τις αρχές της ιδιωτικότητας.

Σεβασμός στην ιδιωτικότητα των χρηστών

Σύμφωνα με την αρχή αυτή (*Respect for User Privacy – Keep it User-Centric*) οι αρχιτέκτονες των συστημάτων θα πρέπει να δίνουν προτεραιότητα κατά τον σχεδιασμό στην προάσπιση των συμφερόντων των υποκειμένων επεξεργασίας. Αυτό θα πρέπει να επιτυγχάνεται υλοποιώντας κατάλληλες ειδοποιήσεις και δίνοντας στους χρήστες όσο το δυνατόν πιο εύχρηστους τρόπους για να προσαρμόζουν τις ρυθμίσεις της ιδιωτικότητάς τους, σύμφωνα με τις επιθυμίες τους. Σε αυτήν την αρχή βλέπουμε πως τίθεται το ζήτημα του να μην καθίσταται δύσκολο (είτε με ελλιπείς, είτε με δυσνόητες και "κρυμμένες" ρυθμίσεις) στα υποκείμενα επεξεργασίας το να μπορούν να προσαρμόζουν τις ρυθμίσεις της ιδιωτικότητας - αυτό είναι βασικό σημείο στον σχεδιασμό συστημάτων επεξεργασίας δεδομένων.

Ιπποκρατικές βάσεις

Οι Ιπποκρατικές Βάσεις προτάθηκαν για πρώτη φορά το 2002 ([14]) ως μια λύση για την εφαρμογή της ιδιωτικότητας σε επίπεδο βάσης δεδομένων. Το βασικό αρχιτεκτονικό σχεδιάγραμμα των βάσεων αυτών φαίνεται στην **Εικόνα 1**.



Εικόνα 1: Βασικό αρχιτεκτονικό σχεδιάγραμμα μιας Ιπποκρατικής Βάσης Δεδομένων ([14])

Όπως βλέπουμε από την εικόνα, μια Ιπποκρατική βάση αποτελείται από τα εξής μέρη:

- **Πολιτική Ιδιωτικότητας** (*Privacy Policy*): είναι το τμήμα του συστήματος διαχείρισης της βάσης δεδομένων, βάσει του οποίου δημιουργούνται τα **μεταδεδομένα ιδιωτικότητας** (*privacy metadata*). Στα μεταδεδομένα (τα οποία βρίσκονται στην ίδια την βάση δεδομένων) αποθηκεύονται οι σχετικοί με την ιδιωτικότητα περιορισμοί της βάσης, για παράδειγμα:
    - Για πόσο χρόνο θα κρατώνται τα δεδομένα
    - Ποιός θα έχει πρόσβαση σε αυτά

- Για ποιόν λόγο θα μπορούν να χρησιμοποιηθούν τα δεδομένα
- κ.ά

- **Συλλογή δεδομένων** (*Data Collection*): κατά την συλλογή δεδομένων ελέγχονται οι προτιμήσεις των χρηστών, ως προς τα κριτήρια της ιδιωτικότητάς τους. Δηλαδή, ελέγχονται οι ρυθμισεις των χρηστών και συγκρίνονται με την πολιτική ιδιωτικότητας της βάσης - αν είναι συμβατά τότε επιτρέπεται η συλλογή των δεδομένων. Αυτό γίνεται μέσω του **αναλυτή εγκυρότητας δεδομένων** (*Data Accuracy Validator*). Δηλαδή αν τα δεδομένα που συλλέχθηκαν δεν συμφωνούν με τις προτιμήσεις της ιδιωτικότητας των χρηστών τότε η συναλλαγή (στην οποία συλλέχθηκαν τα δεδομένα) απορρίπτεται. Εάν ο αναλυτής επιτρέψει την συλλογή, τότε μόνο επιτρέπεται η αποστολή των δεδομένων από τον χρήστη στην βάση και την αποθήκευσή τους. Αυτή η φάση ονομάζεται **εισαγωγή δεδομένων** (*Data Insertion*). Άλλο ένα βήμα στο οποίο μπορεί να προβεί ο αναλυτής είναι η επεξεργασία των δεδομένων ώστε να είναι ακριβή, βάσει των απαιτήσεων ακρίβειας που έχουν τεθεί στην πολιτική ιδιωτικότητας (αυτή είναι η φάση **προ-επεξεργασίας δεδομένων** - *Data Preprocessing*)

- **Ερωτήματα στην βάση** (*Queries*): σε αυτό το τμήμα υποβάλλονται ερωτήματα στην βάση μαζί με τον σκοπό του κάθε ερωτήματος. Τα ερωτήματα στην βάση χωρίζονται στα παρακάτω στάδια:
  - **Προ της εκτέλεσης του ερωτήματος** (*Before Query Execution*): αυτό το στάδιο ελέγχεται η εγκυρότητα του ερωτήματος (πχ οτι ο χρήστης που έκανε το ερώτημα έχει εξουσιοδοτηθεί να το κάνει) καθώς και αν το ερώτημα προσπελαύνει μόνο δεδομένα που είναι απαραίτητα για τον σκοπό του ερωτήματος
  - **Κατά την εκτέλεση του ερωτήματος** (*During Query Execution*): σε αυτό το στάδιο ελέγχεται ότι το ερώτημα έχει πρόσβαση ("βλέπει") μόνο τα δεδομένα τα οποία εξυπηρετούν τον σκοπό του ερωτήματος, αξιοποιώντας τον **έλεγχο πρόσβασης εγγραφών** (*Record Access Control*)
  - **Μετά την εκτέλεση του ερωτήματος** (*After Query Execution*): σε αυτό το στάδιο, τα αποτελέσματα του ερωτήματος ελέγχονται από τον **ανιχνευτή επιθέσεων** (*Query Intrusion Detector*), ο οποίος διασταυρώνει εάν το ερώτημα ανταποκρίνεται στην συνήθη συμπεριφορά του χρήστη ή αν ο χρήστης έχει αλλάξει συμπεριφορά. Η ανάλυση αυτή της συμπεριφοράς βασίζεται σε ιστορικά στοιχεία που κρατώνται για τον χρήστη, βάσει παλαιοτέρων ερωτημάτων του, και αποθηκεύονται στο **μοντέλο επιθέσεων σε ερωτήματα** (*Query Intrusion Model*)

- **Διατήρηση** (*Retention*): εδώ αξιοποιείται ο **διαχειριστής διατήρησης δεδομένων** (*Data Retention Manager*) ο οποίος διαγράφει τα δεδομένα τα οποία δεν εξυπηρετούν πλέον τον σκοπό για τον οποίο συλλέχθησαν

- **Αναλυτής Συλλογής Δεδομένων** (*Data Collection Analyzer*): ελέγχει αν τα δεδομένα τα οποία έχουν συλλεχθεί, έχουν αξιοποιηθεί από τα διάφορα σχετικά ερωτήματα. Σε περίπτωση που κάποια δεδομένα δεν αξιοποιούνται, ή κάποια ερωτήματα έχουν περισσότερα δικαιώματα από όσα χρειάζονται τότε ο αναλυτής διαγράφει τα σχετικά δεδομένα

Όπως βλέπουμε από την ανάλυση, οι Ιπποκρατικές βάσεις δεδομένων δίνουν αρκετή βάση στην εφαρμογή της ιδιωτικότητας σε επίπεδο σχεδιασμού. Δηλαδή, είναι σχεδιασμένες με τέτοιον

τρόπο ώστε να προστατεύουν με τεχνικά μέσα (καθώς και οργανωτικά, μέσω της πολιτικής ιδιωτικότητας) την πρόσβαση στα δεδομένα προσωπικού χαρακτήρα. Σημαντικό είναι να υπογραμμίσουμε τόσο την αξιοποίηση της κρυπτογραφίας όσο και την πολιτική προσβάσεως - ο συνδυασμός των οποίων αυξάνει το επίπεδο προστασίας των δεδομένων και απο εσωτερικές απειλές ή επιθέσεις. Ενώ όμως αποτελούν σημαντικό βήμα προς την πρακτική υλοποίηση της ιδιωτικότητας δια του σχεδιασμού, είναι ίσως αρκετά δύσκολες στην εφαρμογή τους λόγω της υψηλής πολυπλοκότητας που παρουσιάζει η αρχιτεκτονική τους. Αυτό γίνεται αντιληπτό από το ότι ενώ υπάρχει σχετική βιβλιογραφία και περαιτέρω αναλύσεις (πχ [15], [16], [17]), εν τούτοις φαίνεται να μην κυριαρχούν ως τεχνολογία στην ευρύτερη αγορά.

Απόσταση Ιδιωτικότητας (Privacy Distance - PriDe)

Άλλη μια προσπάθεια εφαρμογής της ιδιωτικότητας στις σχεσιακές βάσεις δεδομένων είναι η μέτρηση της απόστασης ιδιωτικότητας ([18]). Η συγκεκριμένη διαδικασία βασίζεται σε δύο βήματα:

1) Δημιουργία "*συμπεριφοράς αναφοράς*" των χρηστών (*baseline behaviour*): σε αυτό το βήμα, το προτεινόμενο σύστημα καταγράφει τα ερωτήματα που τίθενται στην βάση, αφού τα κανονικοποιήσει πρώτα, και φτιάχνει μια βάση αναμενόμενων ή συνηθισμένων ερωτημάτων που κάνει ο κάθε χρήστης στην βάση. Το συγκεκριμένο βήμα τίθεται σε εφαρμογή για ορισμένο χρονικό διάστημα, ώστε να δημιουργηθεί μια αρκετά μεγάλη και ικανοποιητική βάση με αναμενόμενα ερωτήματα - σε τέτοιον βαθμό ώστε να καλύπτεται μεγάλο κομμάτι της αναμενόμενης συμπεριφοράς του κάθε χρήστη

2) Καταγραφή της "*τρέχουσας συμπεριφοράς*" των χρηστών (*runtime behaviour*): σε αυτό το βήμα, το προτεινόμενο σύστημα καταγράφει τα ερωτήματα που τίθενται στην βάση, αφού τα κανονικοποιήσει πρώτα, και κάθε ερώτημα συγκρίνεται με τα κανονικοποιημένα ερωτήματα του πρώτου βήματος παράγοντας έτσι μια "απόκλιση" (απόσταση ιδιωτικότητας) του τρέχοντος ερωτήματος από τα αποθηκευμένα αναμενόμενα

Τα δύο βήματα υλοποιούνται χρησιμοποιώντας τις καταγραφές των βάσεων (*database logs*)[8]. Από αυτά προκύπτει μια απόκλιση (ή απόσταση ιδιωτικότητας) μεταξύ του τρέχοντος ερωτήματος (*runtime behaviour*) και της συμπεριφοράς αναφοράς (*baseline behaviour*). Η απόκλιση αυτή, η οποία ονομάζεται επίσης και βαθμολογία (*score*), υπολογίζεται χρησιμοποιώντας μια μηχανή βαθμολόγησης (*scoring engine*), η οποία χρησιμοποιεί την μέθοδο σύγκρισης μεταξύ **ν-γραμμάτων** (*n-grams*)[9]. Ανάλογα με το πόσο διαφέρουν τα ν-γράμματα του τρέχοντος ερωτήματος σε σχέση με τα ν-γράμματα της συμπεριφορά αναφοράς, παράγεται η βαθμολογία (απόσταση ιδιωτικότητας). Στην **Εικόνα 2** φαίνεται διαγραμματικά η διαδικασία παραγωγής της απόστασης ιδιωτικότητας απο το προτεινόμενο σύστημα.

[8] Εδώ εννούνται οι καταγραφές συναλλαγών (transaction logs) που χρησιμοποιούνται εσωτερικά από το σύστημα διαχείρισης βάσεων δεδομένων, όχι οι καταγραφές που γίνονται σε επίπεδο εφαρμογής σε κάποιο αρχείο προσβάσιμο από τους εξουσιοδοτημένους χρήστες και που προορίζονται για εξέταση συμβάντων και αποσφαλμάτωση

[9] Όρος που έρχεται από την υπολογιστική γλωσσολογία (computational linguistics) και τις πιθανότητες, το ν-γραμμα (n-gram, επίσης συναντάται και ως Q-gram) είναι μια συνεχής ακολουθία ν αντικειμένων από κάποιο κείμενο ή ομιλία. Τα αντικείμενα αυτά μπορεί να είναι γράμματα, φωνήματα, συλλαβές ή λέξεις, αναλόγως με το ποιά είναι η βαθμίδα αναλύσεως του προγράμματος που χρησιμοποιεί το συγκεκριμένο ν-γραμμα

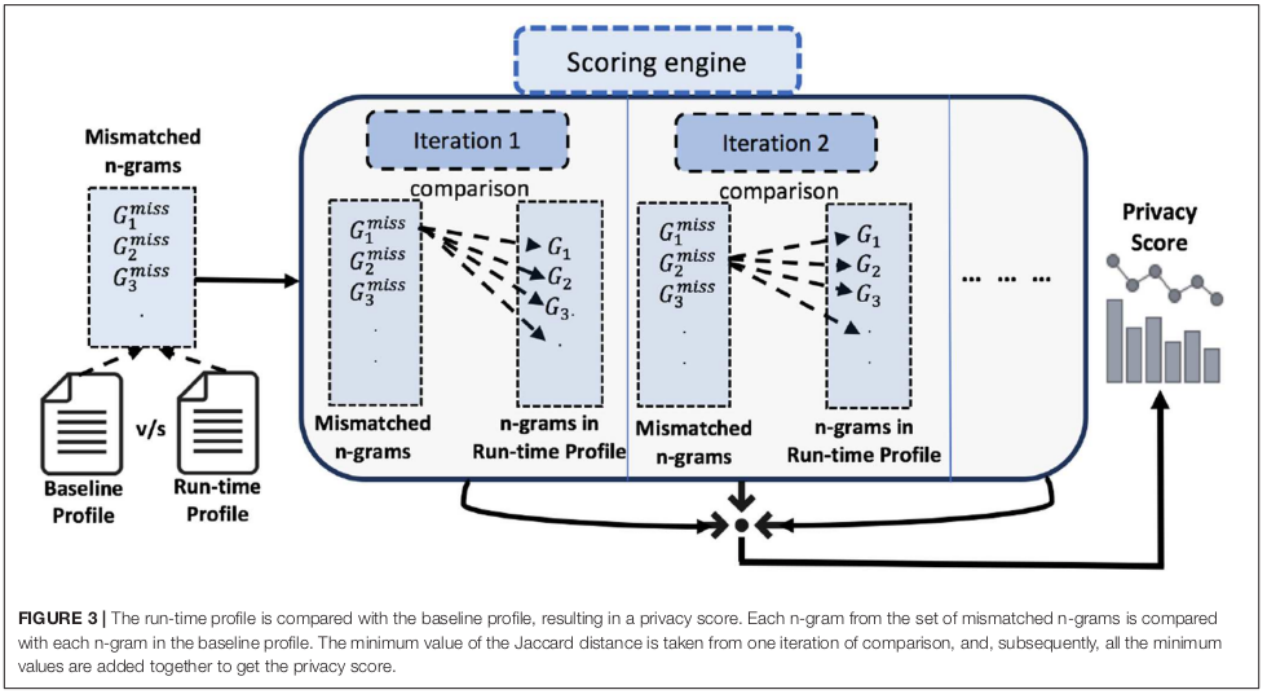


FIGURE 3 | The run-time profile is compared with the baseline profile, resulting in a privacy score. Each n-gram from the set of mismatched n-grams is compared with each n-gram in the baseline profile. The minimum value of the Jaccard distance is taken from one iteration of comparison, and, subsequently, all the minimum values are added together to get the privacy score.

Εικόνα 2: Διάγραμμα παραγωγής της βαθμολογίας ιδιωτικότητας ([18])

Η βαθμολογία παίζει πολύ σημαντικό ρόλο, καθώς στην ουσία δείχνει τον βαθμό κατά τον οποίο η ιδιωτικότητα παραβιάζεται με κάθε ερώτημα. Δηλαδή, με κάθε ερώτημα που γίνεται στην βάση, γίνονται γνωστές πληροφορίες σχετικά με τα υποκείμενα επεξεργασίας και, συνεπώς, η ιδιωτικότητα φθίνει μετά από κάθε ερώτημα. Ο βαθμός του κάθε ερωτήματος αντιπροσωπεύει το πόσο έφθινε η ιδιωτικότητα μετά από το ερώτημα αυτό.

Η βαθμολογία, λοιπόν, μπορεί να είναι ξεχωριστή για κάθε ερώτημα ή μπορεί να είναι αθροιστική για όλα τα ερωτήματα που έχουν τεθεί στην βάση για κάποιο υποκείμενο επεξεργασίας. Στην δεύτερη περίπτωση μπορούμε να υπολογίσουμε το πόσο έχει φθίνει η ιδιωτικότητα στον χρόνο και να το συγκρίνουμε με κάποιο **κατώφλι ανοχής απώλειας ιδιωτικότητας** (*acceptable threshold for privacy loss*). Η βαθμολογία που προκύπτει (είτε ξεχωριστά, είτε αθροιστικά) χρησιμοποιείται ως είσοδος σε κάποιο **σύστημα ανίχνευσης εισβολών** (*intrusion detection system*) το οποίο, με την σειρά του, μπορεί να προβεί σε ενέργειες για τις οποίες έχει ρυθμιστεί.

Σε αυτό το σημείο πρέπει να διευκρινιστεί, πως το προτεινόμενο σύστημα *δεν είναι αυτοτελές ως προς την εφαρμογή της ιδιωτικότητας* των υποκειμένων επεξεργασίας. Αντιθέτως λειτουργεί ως μέρος ή βοήθημα σε κάποιο ευρύτερο σύστημα, καθώς *ο στόχος του PriDe είναι να ανιχνεύει ανωμαλίες στην συμπεριφορά των χρηστών* και όχι να δρα επ'αυτού.

Διαφορική Ιδιωτικότητα (Differential Privacy)

Η διαφορική ιδιωτικότητα ([19], [20]) είναι τρόπος εφαρμογής της ιδιωτικότητας των υποκειμένων επεξεργασίας, κυρίως όσον αφορά τις στατιστικές βάσεις δεδομένων[10]. Η

[10] Στατιστική βάση δεδομένων ονομάζεται η βάση δεδομένων που χρησιμοποιείται για να γίνουν στατιστικές αναλύσεις επι των δεδομένων. Δηλαδή δεν παρουσιάζουν ενδιαφέρον οι συγκεκριμένες εγγραφές στην βάση αλλά η στατιστική που προκύπτει εξετάζοντας πλήθος εγγραφών

προσέγγιση αυτή αφορά στην ελεγχόμενη εισαγωγή θορύβου σε απαντήσεις ερωτημάτων σε βάσεις, με τέτοιον τρόπο ώστε να μην μπορεί να εξάγει συμπεράσματα ο ερωτών σχετικά με την ύπαρξη συγκεκριμένων δεδομένων στην βάση.

Πιο συγκεκριμενα, έστω πως υπάρχουν δύο βάσεις δεδομένων, η Α και η Β, οι οποίες έχουν τα ίδια δεδομένα (σε πληθος και περιεχόμενα) εκτός από μία εγγραφή. Όταν κάποιος κάνει το ίδιο ερώτημα στις δύο βάσεις, θα πρέπει να παίρνει σχεδόν τις ίδιες απαντήσεις. Με αυτόν τον τρόπο, δεν θα μπορεί να εξάγει συμπεράσματα ως προς το είδος των δεδομένων που έχει η κάθε βάση. Αυτό επιτυγχάνεται αν σε κάθε απάντηση εισάγεται κάποιο ποσό θορύβου, το οποίο θα εξισώνει τις πιθανοτικές κατανομές των απαντήσεων. Το ποσό του θορύβου αυτό θα πρέπει να είναι τυχαιοποιημένο, δηλαδή να μην είναι σταθερό για να μην μπορεί να εξαχθεί οποιοδήποτε συμπέρασμα, στην περίπτωση που κάποιος κάνει πολλές ερωτήσεις στην ίδια βάση.

Το ποσό του θορύβου που εισάγεται περιορίζεται και προσδιορίζεται από μια **παράμετρο ε** (*ε-διαφορική ιδιωτικότητα / ε-differential privacy*) η οποία ονομάζεται **παράμετρος απώλειας ιδιωτικότητας**. Η παράμετρος προκύπτει απο την **πιθανοτική κατανομή Λαπλάς** (*Laplace probability distribution*) που προσδιορίζει την απόκλιση στις απαντήσεις της ίδιας ερώτησης σε δύο βάσεις που διαφέρουν μόνο κατα μία εγγραφή. Με άλλα λόγια, αν γίνει η ίδια ερώτηση στην βάση Α, και λάβουμε την απάντηση Χ, και στην βάση Β και λάβουμε την απάντηση Υ, τότε η απόκλιση στις απαντήσεις Χ και Υ αποδίδεται απο μια πιθανοτική κατανομή τύπου Λαπλάς. Αυτή η κατανομή προσδιορίζει και το ποσό του θορύβου που θα εισάγουμε σε κάθε απάντηση ώστε να αποσβεστεί η όποια απόκλιση μεταξύ των απαντήσεων.

Όσο πιο μικρή θέτουμε την παράμετρο ε, τοσο πιο μικρή απώλεια ιδιωτικότητας περιμένουμε να έχουμε από την απάντηση ερωτημάτων, άρα πιο μεγάλο ποσό θορύβου θα εισαχθεί στις απαντήσεις που οδηγεί σε πιο ανακριβείς απαντήσεις. Αντιθέτως, μεγάλη παράμετρος ε, σημαίνει ότι αποδεχόμαστε μεγάλη απώλεια ιδιωτικότητας στις απαντήσεις, άρα βάζουμε λιγότερο θόρυβο στις απαντήσεις και, συνεπώς, εκείνες είναι πιο ακριβείς.

Απο τα παραπάνω συμπεραίνουμε πως η *επιλογή του μεγέθους της παραμέτρου ε αποτελεί συμβιβασμό* ανάμεσα στην ακρίβεια που επιθυμούμε στις απαντήσεις των ερωτημάτων και στην απώλεια της ιδιωτικότητας που επέρχεται εξαιτίας των ερωτημάτων στην βάση. Επίσης, βλέπουμε πως και αυτή *η μέθοδος ενεργεί στις απαντήσεις των ερωτημάτων και όχι στα ίδια τα δεδομένα που περιέχουν τις προσωπικές πληροφορίες των υποκειμένων επεξεργασίας*.

6 *Σχεσιακές Βάσεις Δεδομένων (ΣΒΔ)*

Εισαγωγή

Σε αυτό το κεφάλαιο θα κάνουμε μια σύντομη αναφορά στις σχεσιακές βάσεις δεδομένων (ΣΒΔ), στα σημεία που παρουσιάζουν ενδιαφέρον για την παρούσα διπλωματική. Αυτά είναι η κανονικοποίηση μιας βάσης δεδομένων και το πως αυτή επιτυγχάνεται με την εύρεση υποψηφίων κλειδιών στις σχεσιακές μεταβλητές.

Ορισμοί

Προτού αναφέρουμε την διαδικασία της κανονικοποίησης μιας βάσης, κρίνεται σκόπιμο να αναφερθούν κάποιοι ορισμοί που χρησιμοποιούνται στην παρούσα εργασία:

απλό κλειδί (simple key): το υποψήφιο κλειδί που αποτελείται από ένα μόνο χαρακτηριστικό

γραμμή ή **εγγραφή** (row ή record): είναι μια καταχώρηση δεδομένων σε έναν πίνακα που αντιστοιχεί στον φυσικό κόσμο σε μία γραμμή του πίνακα

εναλλακτικό κλειδί (alternate key): το υποψήφιο κλειδί που δεν χρησιμοποιείται ως πρωτεύον του πίνακα

λειτουργική εξάρτηση (functional dependency): ο προσδιορισμός κατά μοναδικό τρόπο ενός χαρακτηριστικού (ή περισσοτέρων) από ένα άλλο χαρακτηριστικό (ή περισσοτέρων). Κάθε εγγραφή ενός πίνακα εξαρτάται λειτουργικά απο τα υποψήφια κλειδιά της εγγραφής

ξένο κλειδί (foreign key): μεταξύ δύο πινάκων Α και Β, στην περίπτωση που ο πίνακας Α αναφέρεται στις εγγραφές του πίνακα Β, ο συσχετισμός αυτός γίνεται όταν ο πίνακας Α έχει ένα ή περισσότερα χαρακτηριστικά τα οποία παίρνουν τιμές από τα αντίστοιχα χαρακτηριστικά του

πίνακα Β, τα οποία αποτελούν το πρωτεύον κλειδί του πίνακα Β. Τα χαρακτηριστικά συσχετισμού του πίνακα Α ονομάζονται ξένο κλειδί

πίνακας (table): είναι η υλοποίηση μιας σχέσης (relation) από ένα σύστημα διαχείρισης βάσεων δεδομένων, η οποία αναφέρεται στην βιβλιογραφία και ως σχεσιακή μεταβλητή (relational variable - "relvar", [5])

πρωτεύον κλειδί (primary key): το υποψήφιο εκείνο κλειδί το οποίο επιλέχθηκε για να χρησιμοποιείται ως αναγνωριστικό μοναδικότητας των εγγραφών του πίνακα

πρωτεύον χαρακτηριστικό (prime attribute): χαρακτηριστικό που αποτελεί μέρος ενός υποψηφίου κλειδιού

σύνθετο κλειδί (composite key): το υποψήφιο κλειδί που αποτελείται από δύο ή περισσότερα χαρακτηριστικά

υπερκλειδί (superkey): οποιοσδήποτε συνδυασμός χαρακτηριστικών ενός πίνακα που μπορεί να προσδιορίσει με μοναδικό τρόπο κάθε εγγραφή του πίνακα

υποκατάστατο κλειδί (surrogate key): ένα, συνήθως, χαρακτηριστικό (στήλη) ενός πίνακα, το οποίο προσδιορίζει μοναδικά κάθε εγγραφή του πίνακα, δεν σχετίζεται με τα υποψήφια κλειδιά του πίνακα και το οποίο, κατά σύμβαση, παίρνει τιμές ακεραίων αριθμών μοναδικές ανά εγγραφή (συνήθως, δε, αυξανόμενες). Χρησιμοποιείται κυρίως για λόγους ευκολίας στον σχεδιασμό μιας βάσης, ειδικά στην περίπτωση που το πρωτεύον κλειδί είναι σύνθετο

υποψήφιο κλειδί ή **κλειδί** (candidate key ή key): το ελάχιστο υποσύνολο χαρακτηριστικών ενός υπερκλειδιού το οποίο μπορεί να προσδιορίσει μοναδικά κάθε εγγραφή του πίνακα - προκύπτει αν από ένα υπερκλειδί αφαιρεθούν όλα τα χαρακτηριστικά εκείνα που δεν είναι απαραίτητα για να προσδιοριστεί μοναδικά η κάθε εγγραφή του πίνακα

φυσικό κλειδί (natural key): ένα κλειδί του οποίου τα χαρακτηριστικά υπάρχουν στην πραγματικότητα, στον φυσικό κόσμο, εκτός της βάσης δεδομένων (πχ ΑΦΜ, αριθμός ταυτότητας κτλ)

χαρακτηριστικό ή **στήλη** (attribute ή column): είναι η διάταξη δεδομένων στον πίνακα που αντιστοιχεί στον φυσικό κόσμο σε μία στήλη του

Οι παραπάνω ορισμοί μας είναι απαραίτητοι για να κατανοήσουμε το πως επιτυγχάνεται η κανονικοποίηση μιας βάσης, αλλά και το ποια δεδομένα στοχεύει η παρούσα διπλωματική. Προτού όμως εξηγήσουμε τα θεμελιώδη βήματα της κανονικοποίησης, θα εξετάσουμε εν συντομία την διαφορά μεταξύ μιας βάσης δεδομένων και του συστήματος διαχείρισής της.

Βάση δεδομένων και ΣΔΒΔ

Μία βασική έννοια που πρέπει να αποσαφηνιστεί σχετικά με τις βάσεις δεδομένων, είναι η διαφορά μεταξύ της ίδιας της βάσης δεδομένων και του συστήματος διαχείρισής της.

Ορισμός: βάση δεδομένων (ΒΔ) ορίζεται η συλλογή μη-πλεοναζόντων δεδομένων ([4])

Δηλαδή, *η βάση δεδομένων είναι τα ίδια τα δεδομένα*. Η αναφορά σε μη-πλεονάζοντα δεδομένα γίνεται ώστε στην έννοια της βάσης δεδομένων να συμπεριλαμβάνεται και ο τρόπος που ομαδοποιούνται και συσχετίζονται τα δεδομένα, ώστε να μην υπάρχουν πληροφορίες οι οποίες επαναλαμβάνονται χωρίς να υπάρχει τέτοια ανάγκη. Επίσης, όταν αναφερόμαστε στις σχέσεις μεταξύ των δεδομένων δεν εννοούμε το πως τα τοποθετούμε στο αποθηκευτικό μέσο για να έχουμε εύκολη ή ταχεία πρόσβαση σε αυτά. Αντιθέτως, εννοείται η σημασιολογική σχέση που έχουν τα δεδομένα μεταξύ τους, δηλαδή πως συσχετίζονται τα δεδομένα σε επίπεδο σημασιολογικό. Για παράδειγμα το όνομα μιας πόλης συσχετίζεται με το όνομα μιας περιοχης, της περιοχής ενός γεωγραφικού διαμερίσματος και του διαμερίσματος με το όνομα μιας χώρας.

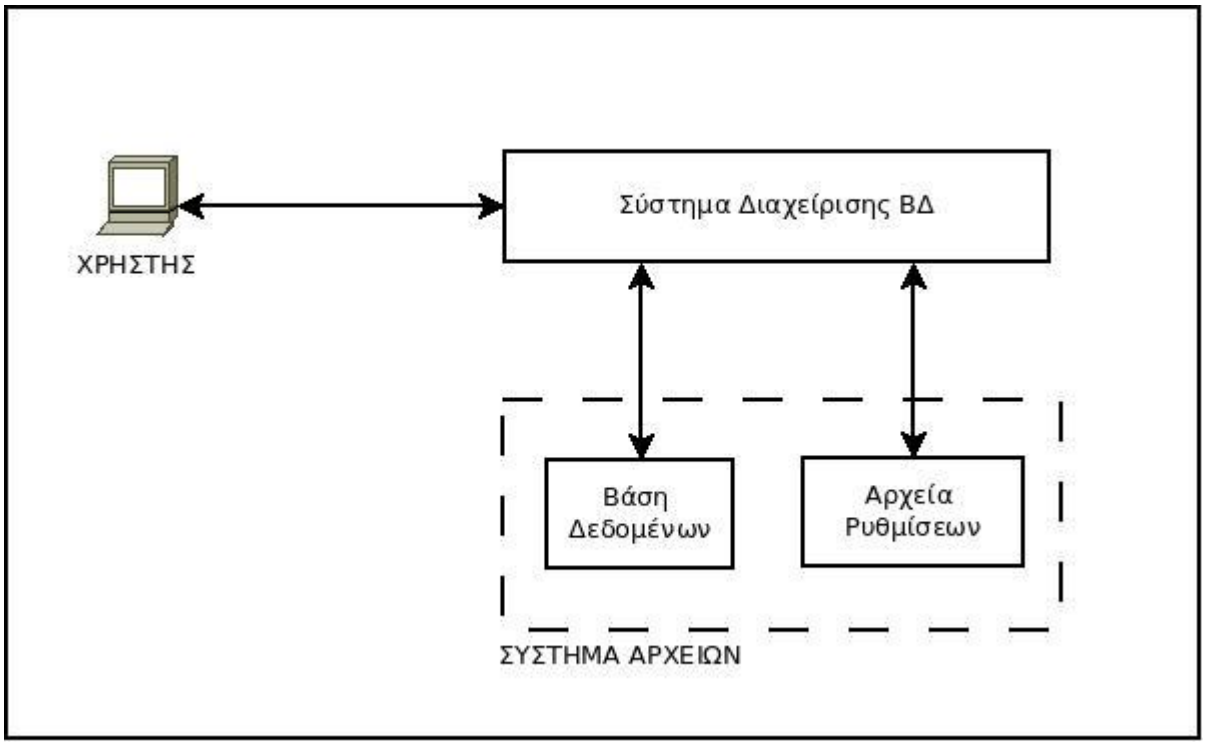
Αυτή *η συσχέτιση αποτελεί κομμάτι της βάσης δεδομένων, όμως ο τρόπος με τον οποίο θα αποθηκευτούν σε έναν σκληρό δίσκο όχι*. Συνήθως στο αποθηκευτικό μέσο αποθηκεύονται με τρόπο που εξυπηρετεί την εύκολη και ταχεία ανάκτηση των δεδομένων, ώστε να μπορεί να χρησιμοποιείται η βάση χωρίς μεγάλες και πολλές καθυστερήσεις. Το γεγονός όμως ότι τα δεδομένα αποθηκεύονται με συγκεκριμένο τρόπο ώστε να διευκολύνεται η ταχεία επεξεργασία τους, σημαίνει πως δεν είναι εύχρηστα στους χειριστές της βάσης και, συνεπώς, χρειάζονται ειδικά προγράμματα διαχείρισής της.

Ορισμός: σύστημα διαχείρισης βάσεων δεδομένων (ΣΔΒΔ) είναι το πρόγραμμα (ή σύνολο προγραμμάτων) το οποίο βοηθά στο να *μεταφράζει* την μορφή των δεδομένων της βάσης από εκείνην στην οποία βρίσκονται όντας στο αποθηκευτικό μέσο, σε μία άλλη πιο εύχρηστη και διαχειρίσιμη από τις οντότητες που τα επεξεργάζονται ([4])

Ο ορισμός του συστήματος διαχείρισης μιας βάσης δεδομένων συμπεριλαμβάνει και πολλές άλλες ιδιότητες, όπως οτι το σύστημα μπορεί να εφαρμόζει έλεγχο πρόσβασης, διαφορετικές όψεις της βάσης αναλόγως την οντότητα που τα επεξεργάζονται, να προσφέρει την δυνατότητα αντιγράφων ασφαλείας και πολλά άλλα. Για την κατανόηση της παρούσας διπλωματικής δεν μας είναι απαραίτητο να κατανοήσουμε όλο το φάσμα των δυνατοτήτων ενός ΣΔΒΔ, αλλά οτι *ο βασικός τους σκοπός είναι να διευκολύνει την πρόσβαση στην βάση*.

Εδώ είναι σημαντικό να αναφέρουμε πως <u>το ΣΔΒΔ δεν είναι απαραίτητο για την πρόσβαση, διαχείριση και επεξεργασία των δεδομένων της βάσης</u>, όμως είναι πρακτικά αναγκαίο εργαλείο καθώς είναι φτιαγμένο με σκοπό την κατανόηση της μορφής των δεδομένων στο αποθηκευτικό μέσο, καθώς και την μετατροπή των δεδομένων απο την μια μορφή στην άλλην.

Στην **Εικόνα 3** βλέπουμε τα δομικά στοιχεία ενός ΣΔΒΔ, δηλαδή το σύστημα διαχείρισης, την βάση δεδομένων, τα αρχεία ρυθμίσεων του συστήματος και τον χρήστη της βάσης. Όπως βλέπουμε, ο χρήστης αλληλεπιδρά με την βάση μέσω του συστήματος διαχείρισης, η λειτουργία του οποίου ορίζεται από τα αρχεία ρυθμίσεων του ΣΔΒΔ (*configuration files*). Στην πιο συνηθισμένη υλοποίηση των ΣΔΒΔ, η βάση και τα αρχεία ρυθμίσεων αποθηκεύονται στο σύστημα αρχείων (*filesystem*) του λειτουργικού συστήματος που φιλοξενεί το ΣΔΒΔ.



Εικόνα 3: Διάγραμμα Αλληλεπίδρασης Χρήστη με το ΣΔΒΔ

Αφού κατανοήσαμε την διαφορά μεταξύ μιας βάσης δεδομένων και του συστήματος διαχείρισής της, θα κάνουμε μια σύντομη αναφορά στο πως σχεδιάζουμε μια σχεσιακή βάση δεδομένων, βάσει των κανόνων που δίνονται από την θεωρία της σχεσιακής άλγεβρας. Αν και η αναφορά θα είναι σύντομη, κρίνεται σημαντική καθώς πάνω σε αυτήν θα βασιστεί το υπόλοιπο της διπλωματικής, για να θεμελιώσει την προτεινόμενη διαδικασία σχεδιασμού.

Φόρμες κανονικοποίησης

Η σχεδίαση μιας σχεσιακής βάσης δεδομένων χρησιμοποιεί κάποιους κανόνες ή *φόρμες*, οι οποίες βοηθούν τον σχεδιαστή να ομαδοποιήσει και συσχετίσει μεταξύ τους τα δεδομένα με τέτοιον τρόπο ώστε να επιτυγχάνεται η βέλτιστη ευχρηστία και ακεραιότητα των δεδομένων, εκμηδενίζοντας τα πλεονάζοντα τέτοια. Αυτή η διαδικασία ονομάζεται *κανονικοποίηση* ([6])

Στην βιβλιογραφία αναλύονται πέντε ή περισσότερες φόρμες κανονικοποίησης, εμάς όμως θα μας απασχολήσουν οι τρεις πρώτες, γιατί αυτές εφαρμόζονται στην πλειοψηφία των βάσεων δεδομένων. Οι φόρμες αυτές δίνουν κάποιες οδηγίες στους σχεδιαστές των σχεσιακών βάσεων δεδομένων, ως προς το πως να ομαδοποιήσουν τα δεδομένα και να τα συσχετίσουν με τον βέλτιστο τρόπο. Οι τρεις φόρμες (*normalization forms - NF*) είναι οι εξής ([4], [5]):

1. Πρώτη Φόρμα Κανονικοποίησης (**1NF**): δεν θα πρέπει να υπάρχουν επαναλαμβανόμενες ομάδες δεδομένων

2. Δεύτερη Φόρμα Κανονικοποίησης (**2NF**): όλα τα χαρακτηριστικά (attributes) που δεν αποτελούν μέρους του κλειδιού θα πρέπει να εξαρτώνται από όλο το κλειδί (σε περίπτωση που το κλειδί είναι σύνθετο)
3. Τρίτη Φόρμα Κανονικοποίησης (**3NF**): όλα τα χαρακτηριστικά, που δεν αποτελούν μέρος του κλειδιού, θα πρέπει να εξαρτώνται αποκλειστικά από το κλειδί κι όχι από κάποιο άλλο χαρακτηριστικό

Αν και δεν πρόκειται να αναλύσουμε περαιτέρω τις φόρμες κανονικοποίησης (ο αναγνώστης μπορεί να ανατρέξει στην πλούσια βιβλιογραφία επι του θέματος, π.χ. [4], [5]), αναφέρθηκαν γιατί έτσι καθίσταται σαφές πως για να κανονικοποιηθεί μια σχεσιακή βάση δεδομένων πρέπει με κάποιον τρόπο να βρούμε τα υποψήφια κλειδιά της κάθε σχέσης (πίνακα). Αυτά, με την σειρά τους, βρίσκονται αναλύοντας τις λειτουργικές εξαρτήσεις μεταξύ των χαρακτηριστικών του πίνακα. Η συγκεκριμένη διαδικασία είναι κρίσιμη για την παρούσα εργασία, καθώς *η εύρεση των υποψηφίων κλειδιών ενός πίνακα αποτελεί βασικό λίθο στην αναγνώριση των χαρακτηριστικών, που θα συμμετάσχουν στην εφαρμογή της ιδιωτικότητας των υποκειμένων επεξεργασίας*. Να σημειωθεί πως θα χρειαστεί να βρούμε όλα τα υποψήφια κλειδιά, γιατί το γεγονός πως επιλέγουμε κάποιο υποψήφιο κλειδί ως πρωτεύον[11] δεν ακυρώνει την ιδιότητα των υπολοίπων κλειδιών να είναι υποψήφια και, συνεπώς, να προσδιορίζουν μοναδικά την κάθε εγγραφή. Αυτό θα χρησιμοποιηθεί αργότερα για να προσδιορίσουμε το ποιά δεδομένα μπορούν να θεωρηθούν ταυτοποιητικά και ποιά αναγνωριστικά[12].

Απλά και σύνθετα κλειδιά

Θα πρέπει εδώ να γίνει μια ανάλυση σχετικά με το πότε χρησιμοποιούμε απλά και πότε σύνθετα κλειδιά. Ο ρόλος του κλειδιού είναι να προσδιορίζει μοναδικά κάθε εγγραφή και για να γίνει αυτό, θα πρέπει *οι τιμές του κλειδιού να είναι σε πλήθος τόσες ώστε να μπορούν να καλύπτουν τον αριθμό των πιθανών εγγραφών στον εκάστοτε πίνακα*. Στην περίπτωση που οι τιμές ενός πεδίου δεν επαρκούν για να καλύψουν το πλήθος των πιθανών εγγραφών στον πίνακα της βάσης (άρα δεν μπορούμε να έχουμε *απλό κλειδί*), τότε οι τιμές του πεδίου αυτού μπορεί να χρησιμοποιηθούν από περισσότερες από μία εγγραφές (πχ το όνομα "*Σπυρίδων*" μπορεί να χρησιμοποιηθεί από πολλές εγγραφές). Χρησιμοποιούμε λοιπόν δεύτερο πεδίο (φτιάχνουμε *σύνθετο κλειδί*), ώστε σε συνδυασμό με τις τιμές του πρώτου πεδίου να προκυψει πλήθος τιμών που μπορεί να καλύψει σε πλήθος όλες τις πιθανές εγγραφές του πίνακα (πχ το όνομα κι επώνυμο "*Σπυρίδων Νίνος*"). Αν και το δεύτερο πεδίο δεν αρκεί ώστε να καλύψουμε όλες τις πιθανές εγγραφές του πίνακα, τότε προσθέτουμε όσα άλλα πεδία απαιτείται. Το κάθε πεδίο όμως, προσφέρει σε διαφορετικό βαθμό στην μοναδικότητα του κλειδιού, ανάλογα με το πόσες ξεχωριστές τιμές μπορεί να πάρει (*πληθικότητα*).

Για να το καταλάβουμε αυτό, υποθέτουμε ότι το πεδίο *όνομα* μπορεί να λάβει μόνο δύο τιμές. Αν χρησιμοποιήσουμε αυτό το πεδίο ως πρωτεύον κλειδί, τότε είναι προφανές πως ο πίνακας μπορεί να πάρει μόνο δύο εγγραφές οι οποίες θα μπορούν να προσδιοριστούν μοναδικά από το πεδίο *όνομα*. Συνεπώς, η χρήση ενός απλού κλειδιού (το πεδίο όνομα) είναι περιοριστική για την βάση μας.

[11] Να σημειωθεί πως η επιλογή ενός υποψηφίου κλειδιού ως πρωτεύον δεν επιβάλλεται απο την σχεσιακή άλγεβρα. Γίνεται, όμως, λαμβάνοντας υπ'όψιν παράγοντες που δεν έχουν να κάνουν με την άλγεβρα - πχ με κριτήριο τον αποθηκευτικό χώρο που κάποιο κλειδί απαιτεί ή την επίδραση στην απόδοση του ΣΔΒΔ

[12] Βλ. Κεφάλαιο 8

Για να μπορέσουμε να αυξήσουμε τις εγγραφές που μπορεί να δεχτεί η βάση, μπορούμε να χρησιμοποιήσουμε δεύτερο πεδίο σε συνδυασμό με το πρώτο. Έστω ότι το πεδίο που χρησιμοποιείται είναι το *επώνυμο* και ότι μπορεί να πάρει εκατό (100) διαφορετικές τιμές. Συνεπώς, τώρα το κλειδί (που είναι σύνθετο) μπορεί να προσδιορίσει μοναδικά εγγραφές που σε πλήθος αντιστοιχούν στο γινόμενο του πλήθους των τιμών του πρώτου πεδίου με το πλήθος του δεύτερου πεδίου. Πιο συγκεκριμένα, το κλειδί μπορεί τώρα να προσδιορίσει μοναδικά 2 (*όνομα*) * 100 (*επώνυμο*) εγγραφές, δηλαδή 200 εγγραφές.

Συνεπεία αυτού είναι πως, για να σχηματίσουμε κλειδιά, προτιμάμε να επιλέγουμε πεδία τα οποία μπορούν να πάρουν πάρα πολλές διαφορετικές τιμές, ώστε να μην χρειάζονται πολλά τέτοια για να προσδιοριστούν μοναδικά όλες τις πιθανές εγγραφές του πίνακα.

Μια τελευταία, αλλά σημαντική παρατήρηση προκύπτει απο το εξής σκεπτικό: στο ίδιο παράδειγμα με τα δύο πεδία (*όνομα* και *επώνυμο*), έστω ότι έχουμε 200 εγγραφές[13]. Αν αφαιρέσουμε το πεδίο *όνομα* από το κλειδί, τότε το πεδίο *επώνυμο* θα προσδιορίζει δύο εγγραφές ανά τιμή. Αν, αντιθέτως, αφαιρέσουμε το πεδίο *επώνυμο* απο το κλειδί, τότε το πεδίο *όνομα* θα προσδιορίζει 100 εγγραφές ανά τιμή. Βλέπουμε λοιπόν πως αφαιρώντας πεδία από ένα υποψήφιο κλειδί, το κλειδί σταματά να προσδιορίζει μοναδικά κάθε εγγραφή. Αντιθέτως, προσδιορίζει μοναδικά *ομάδες εγγραφών*, το πλήθος των οποίων (ομάδων) ισοδυναμεί με το πλήθος των τιμών που παίρνει το πεδίο που αφαιρέθηκε (ομάδες των 2 όταν αφαιρείται το *όνομα*, ομάδες των 100 όταν αφαιρείται το *επώνυμο*). Αυτή η παρατήρηση είναι σημαντική και θα χρησιμοποιηθεί στο Κεφάλαιο 8 ως κριτήριο προστασίας τμημάτων κλειδιών.

[13] Υποθέτουμε ισοκατανομή των εγγραφών, δηλαδή μία εγγραφή για κάθε συνδυασμό (ζεύγος τιμών) των δύο πεδίων

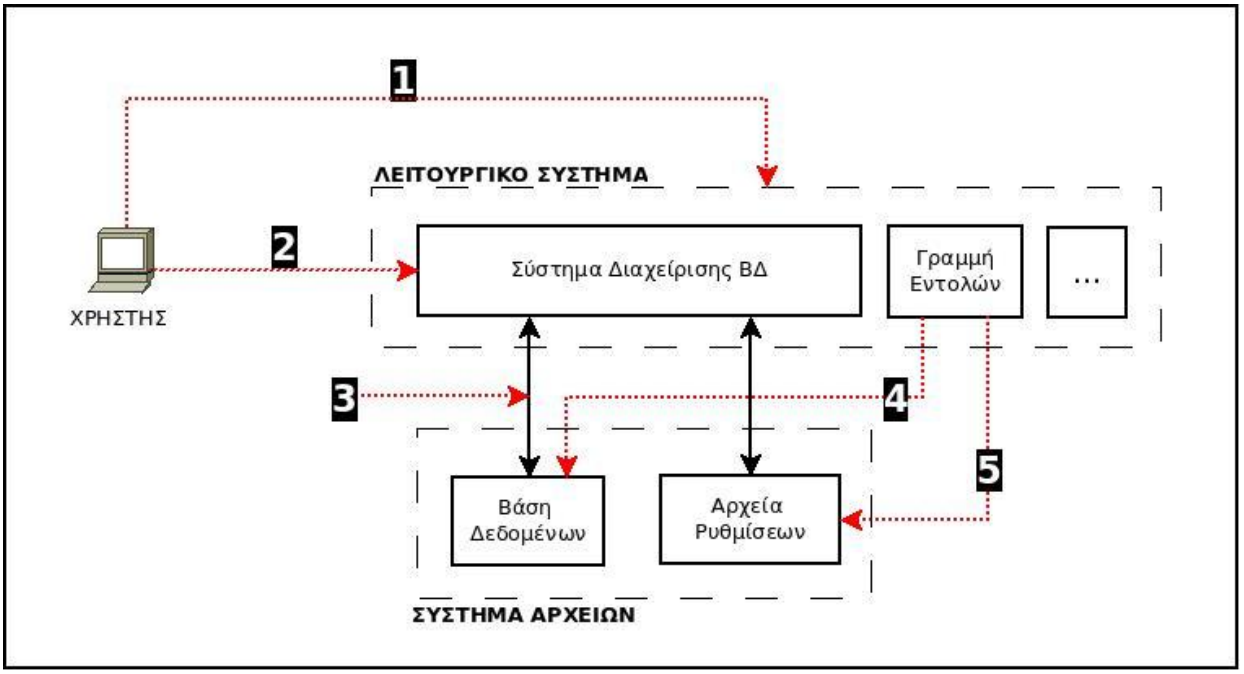
7 *Μοντέλο Απειλών ΣΔΒΔ*

Εισαγωγή

Σε αυτό το κεφάλαιο θα μελετήσουμε το μοντέλο απειλών μιας σχεσιακής βάσης δεδομένων ("ΣΒΔ"). Το μοντέλο είναι απαραίτητο για να μπορέσουμε να κατανοήσουμε τους τρόπους με τους οποίους κάποιος επιτιθέμενος μπορεί να αποκτήσει πρόσβαση στα δεδομένα μιας βάσης δεδομένων και, συνεπώς, να κατανοήσουμε ποιά αντίμετρα απαιτούνται και την συνεισφορά της διπλωματικής σε αυτά.

Μοντέλο Απειλών

Στην **Εικόνα 4** βλέπουμε το μοντέλο απειλών που χρησιμοποιείται από την παρούσα εργασία για να θεμελιώσει την διαδικασία της προστασίας δεδομένων μίας βάσης ώστε να επιτευχθεί η ιδιωτικότητα δια του σχεδιασμού.



Εικόνα 4: Μοντέλο Απειλών σε ΣΔΒΔ

Στην ίδια εικόνα βλέπουμε τα δομικά στοιχεία από τα οποία αποτελείται το περιβάλλον αλληλεπίδρασης μεταξύ χρηστών και της βάσης, και τα οποία είναι τα εξής:

- **χρήστης**: οντότητα, άνθρωπος ή πρόγραμμα, η οποία συνδέεται στο ΣΔΒΔ για να προσπελάσει τα δεδομένα της βάσης
- **λειτουργικό σύστημα**: το λειτουργικό σύστημα εντός του οποίου εκτελείται το ΣΔΒΔ και το οποίο διαχειρίζεται το σύστημα αρχείων (*filesystem*) στο οποίο αποθηκεύονται τα δεδομένα της βάσης
- **γραμμή εντολών**: είναι το πρόγραμμα το οποίο δέχεται εντολές τις οποίες πληκτρολογεί μέσω ενός τερματικού κάποιος χρήστης και τις εκτελεί (πχ οποιοδήποτε κέλυφος, όπως το bash, csh, ksh κτλ)
- **βάση δεδομένων**: τα προσπελάσιμα δεδομένα, όπως έχει οριστεί στο προηγούμενο κεφάλαιο
- **αρχεία ρυθμίσεων**: είναι τα αρχεία εκείνα που διαβάζει το ΣΔΒΔ, συνήθως κατά την εκκίνησή του, και τα οποία ρυθμίζουν την λειτουργία του. Σε αυτά τα αρχεία συμπεριλαμβάνονται εκείνα τα οποία ρυθμίζουν και την ασφάλεια του ΣΔΒΔ, δηλαδή τους τρόπους πρόσβασης, τυχόν κωδικούς πρόσβασης χρηστών κτλ

Διευκρινίζεται πως με μαύρα βέλη (διπλής κεφαλής) απεικονίζονται οι αλληλεπιδράσεις μεταξύ του ΣΔΒΔ με την βάση και τα αρχεία ρυθμίσεων. Με κόκκινα βέλη απεικονίζονται τα σημεία ή τις αλληλεπιδράσεις εκείνες μέσω των οποίων ένας επιτιθέμενος μπορεί να παραβιάσει τα μέτρα προστασίας του ΣΔΒΔ και να αποκτήσει διευρυμένη πρόσβαση στην βάση. Τα σημεία αυτά αναλύονται παρακάτω.

Επίθεση στο λειτουργικό σύστημα (1)

Το σημείο (1) απεικονίζει την επίθεση ενός χρήστη στο λειτουργικό σύστημα, στο οποίο εκτελείται η διαδικασία (*process*) του ΣΔΒΔ. Η επίθεση αυτή μπορεί να αποτελείται από εκμετάλλευση κάποιας ευπάθειας ή την χρήση διαπιστευτηρίων χρηστών[14] του λειτουργικού συστήματος. Το αποτέλεσμα της επίθεσης θα είναι να δοθεί πρόσβαση υψηλού ή χαμηλού επιπέδου εξουσιοδότησης στον επιτιθέμενο. Στις περιπτώσεις που ο επιτιθέμενος λάβει εξουσιοδότηση υψηλού επιπέδου, πχ χρήστη που μπορεί να διαβάσει το σύστημα αρχείων στο οποίο αποθηκεύεται η βάση ή διαχειριστή του ΣΔΒΔ, τότε προβαίνει σε ανεπιθύμητη επεξεργασία της βάσης. Αν, απο την άλλη ο επιτιθέμενος, αποκτήσει εξουσιοδότηση χαμηλού επιπέδου (πχ απλού χρήστη) τότε πρέπει να χρησιμοποιήσει τεχνικές διεύρυνσης εξουσιοδότησης ή να *περιστραφεί* (pivot) μέσω ενός άλλου, καταλλήλου, χρήστη ώστε να αποκτήσει πρόσβαση στα δεδομένα της βάσης.

Επίθεση στο ΣΔΒΔ (2)

Το σημείο (2) απεικονίζει την επίθεση ενός χρήστη απ'ευθείας στο ΣΔΒΔ. Εδώ θεωρείται πως ο χρήστης θα χρησιμοποιήσει είτε κάποια ευπάθεια του ΣΔΒΔ, ώστε να αλληλεπιδράσει με αυτό ως ο διαχειριστής του, ή με την χρήση διαπιστευτηρίων καταλλήλων χρηστών που θα του δώσουν την δυνατότητα να προβούν σε επεξεργασία της βάσης. Μόλις ο επιτιθέμενος αποκτήσει πρόσβαση στο ΣΔΒΔ, θα μπορεί να εκτελέσει ερωτήματα στην βάση.

Κατάχρηση των ερωτημάτων στην βάση (3)

Το σημείο (3) απεικονίζει την αλληλεπίδραση του ΣΔΒΔ και της βάσης. Στο σημείο αυτό θεωρούμε πως ο επιτιθέμενος θα μπορεί να εκτελέσει ερωτήματα σε όλο το εύρος των δεδομένων, τα οποία ερωτήματα θα ειναι απεριόριστου πλήθους και μορφής. Η επίθεση αυτή προϋποθέτει την απόκτηση κατάλληλης εξουσιοδότησης από την πλευρά του επιτιθέμενου, πχ επίπεδο διαχειριστή της βάσης ή του ΣΔΒΔ. Κατάλληλη εξουσιοδότηση επίσης θεωρείται η ελλιπής προστασία των δεδομένων από τους χρήστες εκτός του διαχειριστή, δηλαδή το να μπορεί ο οποιοσδήποτε χρήστης να εκτελέσει ερωτήματα στην βάση χωρίς περιορισμούς (τόσο στην μορφή των ερωτημάτων όσο και στο εύρος των υπο επεξεργασία δεδομένων).

Πρόσβαση δεδομένων εκτός ΣΔΒΔ (4)

Το σημείο (4) απεικονίζει την απ'ευθείας πρόσβαση των δεδομένων απο τον επιτιθέμενο, χρησιμοποιώντας μια γραμμή εντολών (*command prompt*). Αυτό προϋποθέτει πως ο επιτιθέμενος έχει εξουσιοδότηση για να προσπελάσει το σύστημα αρχείων (*filesystem*) επι του οποίου βρίσκεται η βάση δεδομένων. Αυτό μπορεί να συμβεί αν ο επιτιθέμενος έχει δικαιώματα διαχειριστή της βάσης ή του λειτουργικού συστήματος (για τον οποίον συνήθως δεν υπάρχουν περιορισμοί). Η συγκεκριμένη πρόσβαση θεωρείται μια απο τις πιο δύσκολες για να αποφευχθούν, καθώς στα σύγχρονα λειτουργικά συστήματα δεν υπάρχουν πολλές επιλογές για

[14] Σε κάθε περίπτωση στην οποία αναφέρουμε την χρήση διαπιστευτηρίων χρηστών, θα εννοείται πως τα διαπιστευτήρια αυτά (πχ όνομα χρήστη και κωδικός πρόσβασης ή κλειδί ασύμμετρης κρυπτογραφίας) θα έχουν αποκτηθεί είτε με υποκλοπή, είτε με εξαντλητική αναζήτηση (brute force) διαπιστευτηρίων επι του ταυτοποιητικού μηχανισμού του συστήματος (πχ login process, ssh κτλ)

περιορισμό της πρόσβασης του διαχειριστή σε κάποιο σημείο του συστήματος αρχείων του λειτουργικού συστήματος. Επίσης, η επεξεργασία γίνεται απ'ευθείας επι των δεδομένων της βάσης, χωρίς την διαμεσολάβηση του ΣΔΒΔ, οπότε δεν μπορούν να επιβληθούν μέτρα προστασίας εκ μέρους του απέναντι σε μη εξουσιοδοτημένη επεξεργασία.

Μεταβολή αρχείων ρυθμίσεων ΣΔΒΔ (5)

Το σημείο (5) απεικονίζει την πρόσβαση που μπορεί να αποκτήσει κάποιος επιτιθέμενος στα αρχεία ρυθμίσεων του ΣΔΒΔ μέσω μιας γραμμής εντολών. Αυτό επίσης προϋποθέτει πως ο επιτιθέμενος έχει εξουσιοδότηση να προσπελάσει το σύστημα αρχείων στο οποίο είναι αποθηκευμένα τα αρχεία ρυθμίσεων - άρα ο επιτιθέμενος ενεργεί ώς ο διαχειριστής της βάσης ή ως ο διαχειριστής του λειτουργικού συστήματος. Η συγκεκριμένη επίθεση δεν αποτελεί από μόνη της απειλή για την βάση, αλλά επειδή μέσω αυτής μπορούν να μεταβληθούν οι παράμετροι ασφαλείας (πχ διαπιστευτήρια χρηστών) του ΣΔΒΔ, μπορεί να χρησιμοποιηθεί σε συνδυασμό με άλλη επίθεση για να διευρύνει την εξουσιοδότηση του επιτιθέμενου κι έτσι να αυξήσει την πιθανότητα επιτυχίας της τελικής επίθεσης.

Από το μοντέλο απειλών που απεικονίζεται στην **Εικόνα 3** εξάγουμε την επιφάνεια επιθέσεων, η οποία αποτελείται από 4 διανύσματα επιθέσεων:

(α) (1) + (4): αποτελείται από συνδυασμό των επιθέσεων (1) και (4), δηλαδή ο επιτιθέμενος πρώτα αποκτά πρόσβαση στο λειτουργικό σύστημα κι έπειτα, χρησιμοποιώντας κάποια γραμμή εντολών, προσπελαύνει τα δεδομένα της βάσης χωρίς την χρήση του ΣΔΒΔ. Εδώ εννοείται πως αν η αρχική πρόσβαση στο λειτουργικό σύστημα δεν γίνεται με κατάλληλα εξουσιοδοτημένο χρήστη, τότε ο επιτιθέμενος θα πρέπει να χρησιμοποιήσει τεχνικές διεύρυνσης εξουσιοδότησης (*privilege escalation*) ή να *περιστραφεί* (*pivot*) χρησιμοποιώντας άλλα σημεία πρόσβασης, για να μπορέσει να αποκτήσει πρόσβαση στην βάση στο σύστημα αρχείων

(β) (1) + (3): αποτελείται από συνδυασμό των επιθέσεων (1) και (3), όπως και το προηγούμενο διάνυσμα. Η διαφορά είναι ότι εδώ υποθέτουμε πως ο επιτιθέμενος, μετά την πρόσβαση στο λειτουργικό σύστημα, έχει την δυνατότητα να συνδεθεί στο ΣΔΒΔ με διευρυμένη εξουσιοδότηση, ώστε να μπορέσει να τελέσει ερωτήματα χωρίς περιορισμούς. Για παράδειγμα, είναι συνήθης πρακτική οτι ο διαχειριστής του ΣΔΒΔ μπορεί να συνδεθεί σε οποιαδήποτε βάση χωρίς να απαιτείται η ταυτοποίησή του, όταν συνδέεται στο ΣΔΒΔ απο το λειτουργικό σύστημα που εκτελεί το ΣΔΒΔ (συνδέεται "τοπικά" δηλαδή). Έτσι, χρησιμοποιεί τον λογαριασμό με τα δικαιώματα του οποίου εκτελείται το ΣΔΒΔ (πχ. όταν ο χρήστης *postgres* συνδέεται τοπικά με το ΣΔΒΔ *postgresql*). Σε αυτήν την περίπτωση, ο επιτιθέμενος το μόνο που χρειάζεται να κάνει είναι να βρει τρόπο να συνδεθεί στο λειτουργικό σύστημα ως τον συγκεκριμένο χρήστη κι έπειτα να συνδεθεί στην βάση ως ο διαχειριστής του ΣΔΒΔ

(γ) (2) + (3): αποτελείται από συνδυασμό των επιθέσεων (2) και (3). Εδώ, όπως και στο προηγούμενο διάνυσμα, ο επιτιθέμενος κάνει κατάχρηση των ερωτημάτων στην βάση. Στην συγκεκριμένη περίπτωση όμως, ο επιτιθέμενος έχει αποκτήσει πρόσβαση στο ΣΔΒΔ απ'ευθείας ως χρήστης με διευρυμένη εξουσιοδότηση

(δ) (1) + (5) + (2) + (3): για το μοντέλο που παρουσιάζεται στην παρούσα εργασία, αυτό το διάνυσμα επιθέσεως είναι το πιο πολύπλοκο. Αποτελείται από δύο βασικά διανύσματα, το (1) + (5) και το (2) + (3). Το (2) + (3) το μελετήσαμε αμέσως παραπάνω. Το (1) + (5) δείχνει ότι ο επιτιθέμενος αποκτά πρόσβαση στο λειτουργικό σύστημα με κατάλληλη εξουσιοδότηση, ώστε να προσπελάσει τα αρχεία ρυθμίσεων του ΣΔΒΔ. Με αυτήν την πρόσβαση, αλλάζει τις ρυθμίσεις ασφαλείας του ΣΔΒΔ, ώστε μετά να μπορεί να συνδεθεί στο ΣΔΒΔ ως ο διαχειριστής της (ή άλλος χρήστης με την κατάλληλη εξουσιοδότηση) και να κάνει χρήση ερωτημάτων στην βάση χωρίς περιορισμούς.

Μέτρα προστασίας από τις απειλές

Όπως είδαμε και πιο πάνω, στο μοντέλο απειλών εντάσσονται και τα μέτρα προστασίας κατά των απειλών. Οπότε χρησιμοποιώντας τα διανύσματα επιθέσεων, τα οποία βρήκαμε στην προηγούμενη ενότητα, μπορούμε να βρούμε και τα κατάλληλα μέτρα προστασίας. Παρατηρώντας τις συνιστώσες των διανυσμάτων, βλέπουμε ότι όλα τα διανύσματα καταλήγουν στις επιθέσεις (3) και (4), οι οποίες ουσιαστικά αποτελούν διαφορετικές απόψεις της ίδιας επίθεσης - δηλαδή της πρόσβασης στα δεδομένα χωρίς περιορισμούς. Συνεπώς, οποιοδήποτε μέτρο το οποίο περιορίζει ή προστατεύει από τις επιθέσεις (3) και (4) θα συνδράμει ουσιαστικά στον περιορισμό της ακτίνας δράσης οποιασδήποτε άλλης επιθέσεως που αναλύεται στο μοντέλο απειλών της παρούσας ενότητας. Οι υπόλοιπες επιθέσεις είναι εκτός του σκοπού της παρούσας εργασίας και ο αναγνώστης προτρέπεται να ανατρέξει στην πλούσια βιβλιογραφία επί του αντικειμένου.

8 *Ιδιωτικότητα δια του σχεδιασμού στις ΣΒΔ*

Εισαγωγή

Σε αυτό το κεφάλαιο θα συνδυάσουμε τις γνώσεις που πήραμε από τα προαναφερθέντα κεφάλαια ώστε να εφαρμόσουμε μια πρακτική λύση στα προβλήματα που προκύπτουν κατά την εφαρμογή της ιδιωτικότητας. Βασικό μέλημα θα είναι η λύση αυτή να μπορεί να εφαρμοστεί στο στάδιο του σχεδιασμού των βάσεων δεδομένων, ώστε να αντιμετωπιστεί η αιτία του προβλήματος που είναι η εξ αρχής διάθεση των προσωπικών δεδομένων φυσικών ανθρώπων χωρίς περιορισμούς.

Προβλήματα με τις μέχρι τώρα προσεγγίσεις

Όπως είδαμε και στα προηγούμενα κεφάλαια, οι περισσότερες προσεγγίσεις προσπαθούν να εφαρμόσουν την ιδιωτικότητα στις απαντήσεις των ερωτημάτων. Αυτό σημαίνει ότι θεωρούν δεδομένο πως οι χρήστες θα χρησιμοποιούν το σύστημα διαχείρισης των βάσεων δεδομένων. Στην πραγματικότητα, όπως είδαμε στο προηγούμενο κεφάλαιο (Κεφάλαιο 7: Μοντέλο Απειλών ΣΒΔ), υπάρχει η δυνατότητα κάποιος εξουσιοδοτημένος χρήστης να αποκτήσει πρόσβαση στην βάση και να κάνει ερωτήματα επί αυτής, χωρίς να περιοριστεί από το σύστημα διαχείρισής της. Συνεπώς, η παραδοχή ότι ένας επιτιθέμενος θα χρησιμοποιεί το σύστημα διαχείρισης της βάσης που έχουμε ορίσει, είναι εγγενής αδυναμία απέναντι στις παραβιάσεις της ιδιωτικότητας[15].

Όσον αφορά στην γενική προσέγγιση που ακολουθούν οι υπάρχουσες μελέτες, παρατηρείται ότι κάποιες απαιτούν κρυπτογράφηση των δεδομένων για προστασία από κλοπή, άλλες μεταδεδομένα στις εγγραφές που εφαρμόζουν την πολιτική ιδιωτικότητας της βάσης και άλλες τίποτα από τα δύο, δεδομένου ότι περιορίζουν τα αποτελέσματα των ερωτήσεων στην βάση. Η χρήση της κρυπτογράφησης και των μεταδεδομένων είναι, βεβαίως, πολύ καλύτερη από το να διατίθενται τα δεδομένα αυτούσια, αλλά παρουσιάζει υψηλή πολυπλοκότητα και κάποιες φορές είναι δύσκολη να χρησιμοποιεί στον βαθμό που απαιτείται για την καθολική προστασία των δεδομένων..

[15] Σημαντική εξαίρεση αποτελούν οι Ιπποκρατικές βάσεις, οι οποίες εφαρμόζουν πολυεπίπεδη προστασία

Γι'αυτόν τον λόγο κρίνεται σκόπιμο να μελετηθεί άλλη προσέγγιση, η οποία θα κινείται στο πνεύμα της πολυεπίπεδης προστασίας των Ιπποκρατικών βάσεων, αλλά θα δίνει μεγαλύτερη έμφαση στην δομή και προστασία των δεδομένων.

Κατηγορίες και κύκλος ζωής δεδομένων

Προτού αναλύσουμε τις ανάγκες που πρέπει να καλύπτει η μέθοδος σχεδιασμού, θα πρέπει να κατανοήσουμε τα είδη των δεδομένων καθώς και τον κύκλο ζωής τους (data lifecycle). Τα δεδομένα χωρίζονται σε δύο ευρείες κατηγορίες[16] ([9], [21]):

1. **Βασικά ή κύρια δεδομένα** (*master data*): δεδομένα που είναι σημαντικά για την λειτουργία μιας επιχείρησης - περιγράφουν ανθρώπους, περιοχές και αντικείμενα τα οποία συμμετέχουν στο επιχειρησιακό κομμάτι μιας εταιρείας

2. **Δεδομένα συναλλαγών** (*transactional data*): τα δεδομένα που αντιπροσωπεύουν τις αλληλεπιδράσεις μεταξύ των κυρίων δεδομένων

Για να κατανοήσουμε τις δύο αυτές κατηγορίες, ας θεωρήσουμε πως μια εταιρεία έχει πελάτες και τους πουλάει προϊόντα. Τα στοιχεία των πελατών (πχ ονόματα, διευθύνσεις κτλ) και των προϊόντων (πχ ονομασία, κωδικοί κτλ) αποτελούν τα κύρια δεδομένα της εταιρείας. Οι παραγγελίες των πελατών (πχ ημερομηνίες παραγγελιών, ποσότητες προϊόντων, τιμολόγηση κτλ) καθώς και οι αποστολές των προϊόντων αποτελούν τα δεδομένα συναλλαγών της εταιρείας.

Τα κύρια δεδομένα δεν τροποποιούνται συχνά, δηλαδή μετά την αρχική καταχώρησή τους στην βάση δεδομένων, θα μεταβληθεί το περιεχόμενό τους μόνο στην περίπτωση που υπάρχει αντίστοιχη αλλαγή στον φυσικό κόσμο (πχ αλλαγή διεύθυνσης πελάτη ή διακοπή προμήθειας προϊόντος). Εν αντιθέσει, τα δεδομένα συναλλαγών από την φύση τους είναι παροδικά και δεν έχουν μακροχρόνια αξία για την εταιρεία, πέραν των περιπτώσεων που επιβάλλονται από την νομοθεσία (πχ τιμολόγια, τα οποία πρέπει να τηρούνται σε αρχείο για ελάχιστο χρονικό διάστημα).

Εξίσου σημαντικό να γίνει κατανοητή είναι η εξάρτηση των δύο τύπων δεδομένων. Όπως φαίνεται και από την χρήση του κάθε τύπου, τα κύρια δεδομένα είναι ανεξάρτητα στην δημιουργία τους. Επίσης, δεν επηρεάζονται από άλλα δεδομένα, τα οποία πιθανόν να έχουμε στην βάση. Αντιθέτως, τα δεδομένα συναλλαγών εξαρτώνται άμεσα από τα κύρια δεδομένα, καθώς από την φύση τους αντιπροσωπεύουν τις αλληλεπιδράσεις μεταξύ των κυρίων δεδομένων. Συνεπώς, δεν μπορούν να υπάρχουν χωρίς να αναφέρονται στα κύρια δεδομένα.

Από τα παραπάνω, μπορούμε να εξάγουμε τον κύκλο ζωής (lifecycle) των δεδομένων[17]([7]):

1. **Συλλογή ή δημιουργία** (*collection ή creation*): το στάδιο κατά το οποίο τα δεδομένα εισέρχονται στο σύστημα. Παράδειγμα, το στάδιο κατά το οποίο ένας χρήστης του συστήματος επεξεργασίας υποβάλει τα στοιχεία του στο σύστημα μέσω ενός περιηγητή

[16] Για τους σκοπούς της παρούσας διπλωματικής, δεν θα ασχοληθούμε και με άλλες κατηγορίες δεδομένων, όπως πχ τα δεδομένα αναφοράς, τα μεταδεδομένα κτλ

[17] Η διαχείριση των δεδομένων (πχ ακρίβεια περιεχομένου, συνέπεια μεταξύ διαφορετικών συστημάτων κτλ) δεν εμπίπτει στον σκοπό της παρούσας διπλωματικής. Αναφέρεται, δε, σε διαφορετικό κλάδο της διαχείρισης των δεδομένων από την προστασία και την ιδιωτικότητα (πχ Master Data Management)

διαδικτύου (*web browser*). Στο στάδιο αυτό τα δεδομένα θεωρούνται δεδομένα σε κίνηση (*data in transit*)

2. **Αποθήκευση** (*storage*): στο στάδιο αυτό τα δεδομένα, που δημιουργήθηκαν στο προηγούμενο στάδιο, αποθηκεύονται σε κάποια βάση δεδομένων. Στο στάδιο αυτό τα δεδομένα θεωρούνται δεδομένα σε στάση (*data at rest*)

3. **Χρήση ή αξιοποίηση** (*usage ή utilization*): στο στάδιο αυτό τα δεδομένα αξιοποιούνται από το σύστημα επεξεργασίας για να εκπληρώσει τον σκοπό συλλογής και επεξεργασίας τους. Τα δεδομένα μεταφέρονται από την βάση δεδομένων στην κύρια μνήμη του συστήματος επεξεργασίας και επιδέχονται προσωρινή μετατροπή (πχ άθροιση τιμών στην μνήμη του συστήματος, η οποία δεν αποθηκεύεται στην βάση) ή μόνιμη (πχ ενημέρωση του περιεχομένου των δεδομένων στην βάση). Τα δεδομένα που βρίσκονται σε αυτό το στάδιο θεωρούνται δεδομένα σε χρήση (*data in use*)

4. **Αντίγραφα ασφαλείας** (*backing up ή archiving*): στο στάδιο αυτό τα δεδομένα αποθηκεύονται σε ειδικά αποθηκευτικά μέσα για μεγάλο χρονικό διάστημα, με σκοπό την επαναφορά τους σε περίπτωση που το σύστημα επεξεργασίας αστοχήσει και χάσει όλα τα δεδομένα από την βάση δεδομένων του. Τα δεδομένα σε αυτό το στάδιο θεωρούνται δεδομένα σε στάση (*data at rest*)

5. **Διαγραφή ή καταστροφή** (*deletion ή destruction*): στο στάδιο αυτό τα δεδομένα καταστρέφονται επειδή δεν θεωρούνται απαραίτητα, έπαψαν να εξυπηρετούν τον σκοπό συλλογής τους ή απαιτείται εν γένει απο την νομοθεσία και τις πολιτικές της εταιρείας. Η καταστροφή πρέπει να είναι μόνιμη, δηλαδή τα δεδομένα θα πρέπει να διαγράφονται από το αποθηκευτικό μέσο (βάση ή αντίγραφα ασφαλείας) με ειδικό τρόπο, ή το μέσο αποθήκευσης να καταστρέφεται. Έτσι, μετά από την διαδικασία της διαγραφής δεν είναι δυνατή η επαναφορά ή ανάκτηση των δεδομένων

Παρατηρώντας τον κύκλο ζωής των δεδομένων, μπορούμε να κάνουμε τις εξής παρατηρήσεις, ως προς τα μέτρα προστασίας τα οποία συνήθως λαμβάνονται ή πρέπει να λαμβάνονται:

- Κατά την συλλογή και την χρήση των δεδομένων χρησιμοποιείται συνήθως η κρυπτογράφηση. Χρησιμοποιείται για την προστασία των δεδομένων κατά την μεταφορά τους μεταξύ του αποθηκευτικού μέσου και του συστήματος επεξεργασίας, καθώς και του συστήματος επεξεργασίας και των τερματικών αλληλεπίδρασης με τους χρήστες (πχ περιηγητές διαδικτύου). Συνήθως χρησιμοποιείται για την προστασία των καναλιών επικοινωνίας, είτε σε επίπεδο δικτύου (πχ εικονικά δίκτυα τύπου IpSec), είτε σε επίπεδο διαδικτυακής μεταφοράς (πχ πρωτόκολλο TLS - Transport Layer Security)

- Κατά την καταστροφή δεδομένων χρησιμοποιούνται ειδικές τεχνικές διαγραφής τους (πχ ασφαλής διαγραφή - secure wipe), ή καταστρέφεται ολοσχερώς το αποθηκευτικό μέσο, το οποίο χρησιμοποιήθηκε για να αποθηκεύσει τα δεδομένα (πχ καταστροφή ταινιών αντιγράφων ασφαλείας - archiving tapes)

- Το μεγαλύτερο πρόβλημα στην εφαρμογή της ιδιωτικότητας παρουσιάζεται στα στάδια της αποθήκευσης (σημείο 2 πιο πάνω) και της δημιουργίας αντιγράφων ασφαλείας (σημείο 4 πιο πάνω). Αυτό γιατί αξιοποιούνται τεχνικές κρυπτογράφησης δεδομένων, αλλά αυτό

τείνει να γίνεται σε επίπεδο αποθηκευτικού μέσου. Δηλαδή κρυπτογραφείται ολόκληρος ο αποθηκευτικός χώρος, έτσι ώστε να μην απασχολείται το σύστημα επεξεργασίας με την διαδικασία κρυπτογράφησης, αλλά μεταφέρεται η ευθύνη στο λειτουργικό σύστημα. Αυτό δημιουργεί κάποια βασικά προβλήματα:

1. χαμηλή απόδοση του συστήματος, καθώς η κρυπτογράφηση και αποκρυπτογράφηση απαιτούν αρκετή επεξεργαστική ισχύ για να γίνουν σε πραγματικό χρόνο και σε τόσα δεδομένα. Το πρόβλημα εντείνεται όταν κρυπτογραφούνται δεδομένα τα οποία δεν χρήζουν προστασίας

2. καθολική αστοχία όταν κάποιος αποκτήσει νόμιμη πρόσβαση στην αποκρυπτογραφημένη μορφή των δεδομένων, πχ ο διαχειριστής του συστήματος της βάσης δεδομένων. Σε αυτήν την περίπτωση μπορεί να τα προσπελάσει το σύνολο των δεδομένων ανενόχλητος, πράγμα που παραβιάζει την βασική αρχή της "αναγκαίας γνώσης" και, συνεπεία αυτού, την ιδιωτικότητα των υποκειμένων επεξεργασίας

3. τα κλειδιά αποκρυπτογράφησης του αποθηκευτικού χώρου βρίσκονται στην κύρια μνήμη του συστήματος επεξεργασίας, οπότε χρήστης με ικανή εξουσιοδότηση (πχ διαχειριστής συστημάτων) μπορεί να τα ανακτήσει και να παραβιάσει την ασφάλεια του κρυπτογραφημένου αποθηκευτικού μέσου[18]

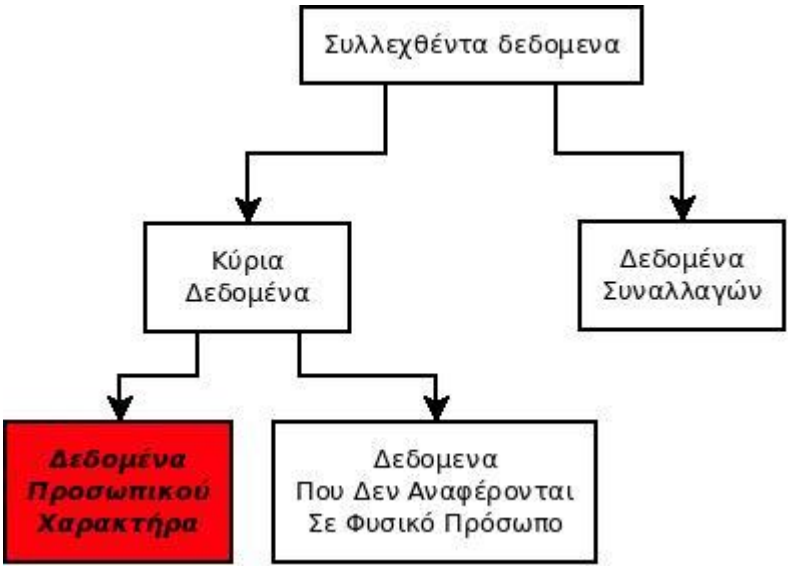
- Τα δεδομένα συναλλαγών εξαρτώνται από τα κύρια δεδομένα καθώς τα πρώτα δεν μπορούν να υπάρξουν χωρίς τα δεύτερα. Η εξάρτηση αυτή αποτυπώνεται στην πράξη με απλή αναφορά από τα δεδομένα συναλλαγών στα κύρια, κι όχι με αντιγραφή του περιεχομένου των κύριων συναλλαγών. Για παράδειγμα, ένας πελάτης αγοράζει ένα προϊόν. Τα κύρια δεδομένα στο παράδειγμα είναι τα στοιχεία του πελάτη και τα στοιχεία του προϊόντος. Τα δεδομένα συναλλαγής είναι η αγορά του προϊόντος από τον πελάτη. Τα δεδομένα συναλλαγής, λοιπόν, δεν αποθηκεύουν τα στοιχεία του πελάτη και του προϊόντος αλλά από μια αναφορά προς τα στοιχεία του πελάτη και του προϊόντος (κυρίως για λόγους εξοικονόμησης χώρου και ακρίβειας). Δηλαδή, δεν θα αποθηκευτεί η συναλλαγή "*Ο Χ αγόρασε έναν υπολογιστή τύπου Ψ*", αλλά αν υποθέσουμε ότι ο πελάτης Χ διαθέτει το μοναδικό αναγνωριστικό[19] *5* και ο υπολογιστής το μοναδικό αναγνωριστικό *14*, τότε η συναλλαγή που θα αποθηκευτεί θα είναι η "*Ο 5 αγόρασε ένα 14*". Αυτή είναι πολύ σημαντική παρατήρηση, καθώς ξεκαθαρίζει πως <u>τα δεδομένα συναλλαγών δεν χρήζουν προστασίας</u> καθώς δεν περιέχουν (ή *δεν πρέπει να περιέχουν*) προσωπικά δεδομένα

- Τα κύρια δεδομένα χωρίζονται σε δύο υποκατηγορίες δεδομένων: τα προσωπικά και τα υπόλοιπα. Η ιδιωτικότητα έχει να κάνει με την προστασία των προσωπικών δεδομένων ενός υποκειμένου και δεν αφορά σε δεδομένα που δεν αντιστοιχούν σε φυσικά πρόσωπα. Για παράδειγμα, τα στοιχεία των εταιρειών και των προϊόντων δεν αποτελούν προσωπικά δεδομένα καθώς δεν αναφέρονται σε φυσικά πρόσωπα. Συνεπώς, <u>απο τα κύρια δεδομένα μόνο τα προσωπικής φύσεως δεδομένα χρήζουν προστασίας</u>

[18] Για λόγους ευκολίας, δεν αναλύουμε την περίπτωση των κρυπτογραφημένων αντιγράφων ασφαλείας, γιατί στην περίπτωση που χρησιμοποιηθούν για να επαναφέρουν κάποιο σύστημα επεξεργασίας το οποίο έχει αστοχήσει, τότε εμπίπτουν στην κατηγορία (2) του κύκλου ζωής των δεδομένων

[19] Στις σχεσιακές βάσεις δεδομένων το αναγνωριστικό αυτό είναι το πρωτεύον κλειδί (primary key) του πίνακα

Απο τις δύο τελευταίες παρατηρήσεις, μπορούμε πλέον να προσδιορίσουμε τον τύπο και το εύρος των δεδομένων τα οποία χρήζουν περαιτέρω ή συγκεκριμένης προστασίας, και είναι τα προσωπικά δεδομένα που εντάσσονται στα κύρια δεδομένα μιας βάσης. Διαγραμματικά αυτό φαίνεται στην **Εικόνα 5**:



Εικόνα 5: Κατηγορίες δεδομένων

Για να μπορέσουμε όμως να προστατέψουμε ένα μόνο τμήμα των δεδομένων μιας βάσης, θα πρέπει να πληρούνται οι παρακάτω συνθήκες:

1. Τα δεδομένα συναλλαγών περιέχουν μόνο αναφορές στα κύρια δεδομένα και όχι οποιοδήποτε τμήμα περιεχομένου των κυρίων δεδομένων (αυτό στις σχεσιακές βάσεις δεδομένων επιτυγχάνεται από την κανονικοποίηση)

2. Τα κύρια δεδομένα θα πρέπει να μην αντιπροσωπεύουν εξαρτήσεις μεταξύ των οντοτήτων (πχ προσώπων ή πραγμάτων) και να είναι ανεξάρτητα μεταξύ τους[20]

3. Οι εξαρτήσεις μεταξύ των οντοτήτων των κυρίων δεδομένων θα πρέπει να αποτυπώνεται αποκλειστικά στα δεδομένα συναλλαγών

Τηρουμένων των παραπάνω συνθηκών, τα κύρια δεδομένα δεν έχουν αναμεταξύ τους εξαρτήσεις και συνεπώς καθίσταται εύκολη η τμηματική τους προστασία. Το τελευταίο είναι ιδιαίτερα ωφέλιμο καθώς, όπως είδαμε και πιο πάνω μας ενδιαφέρει να προστατέψουμε μόνο τα δεδομένα προσωπικού χαρακτήρα για να πετύχουμε την αρχή της ιδιωτικότητας δια του σχεδιασμού. Τα υπόλοιπα δεδομένα δεν παίζουν καποιον άμεσο ρόλο στην ιδιωτικότητα των υποκειμένων επεξεργασίας.

[20] Να υπενθυμίσουμε πως αναφερόμαστε σε σχεσιακές βάσεις δεδομένων και όχι σε άλλου τύπου βάσεις, όπως πχ οι βάσεις δεδομένων γράφων (*graph databases*). Σε εκείνες βασικό ρόλο παίζουν οι σχέσεις των οντοτήτων (ακμές του γράφου) και χρήζουν διαφορετικής αντιμετώπισης σε σχέση με τις ΣΒΔ

Προστασία κατά την εγγραφή και την ανάγνωση

Όπως είδαμε και στο προηγούμενο εδάφιο, στον κύκλο ζωής των δεδομένων υπάρχουν δύο στάδια τα οποία χρήζουν προστασίας, το (2) και το (4) (αποθήκευση και αντίγραφα ασφαλείας). Το μεν στάδιο (2) αφορά στην εξ αρχής αποθήκευση των δεδομένων, δηλαδή την αποθήκευση που λαμβάνει χώρα αμέσως μετά την συλλογή των δεδομένων, ενώ το στάδιο (4) λαμβάνει χώρα σε άλλο χρονικό σημείο, που προκύπτει κυρίως από τα σχέδια διασφάλισης επιχειρηματικότητας και ανάκαμψης από καταστροφή. Συνεπώς, *τα δεδομένα πρέπει να προστατεύονται κατά την πρώτη τους αποθήκευση στην βάση δεδομένων*. Είναι δυνατόν να προστατευτούν και αργότερα, αλλά δεν θεωρείται πρακτικό να αποθηκεύονται χωρίς προστασία και μετά να αντικαθίστανται απο προστατευμένα δεδομένα, λόγω των προβλημάτων που δημιουργούνται από αυτήν την προσέγγιση (πχ ασυνέπεια δεδομένων, δυνατότητα υποκλοπής δεδομένων στο ενδιάμεσο κτλ).

Σε αντίθεση με την παραπάνω θεώρηση, οι περισσότερες προσεγγίσεις[21] που ακολουθούνται από αντίστοιχες εργασίες αφορούν στην εφαρμογή της ιδιωτικότητας κατά την ανάγνωσή τους από το σύστημα διαχείρισης της βάσης, όχι κατά την εγγραφή των δεδομένων. Αυτό έχει ως αποτέλεσμα να εφαρμόζεται εν τέλει η ιδιωτικότητα, αλλά να εξαρτάται από την χρήση του συστήματος διαχείρισης (το οποίο επιβάλλει και την εφαρμογή της ιδιωτικότητας). Αυτή η *εξάρτηση αποτελεί εγγενή αδυναμία των προσεγγίσεων*, καθώς σύμφωνα με το μοντέλο απειλών που έχουμε δει, ένας επιτιθέμενος μπορεί να προσπελάσει την βάση δεδομένων χωρίς να χρησιμοποιήσει το ΣΔΒΔ.

Οι δύο προσεγγίσεις που ακολουθούνται λοιπόν, είναι συνοψίζονται ως εξής:

- **Προστασία κατά την εγγραφή** (*protection on write*): σημαίνει ότι τα δεδομένα προστατεύονται (πχ κρυπτογραφούνται) προτού αποθηκευτούν και αποθηκεύονται προστατευμένα

- **Προστασία κατά την ανάγνωση** (*protection on read*): σημαίνει ότι τα δεδομένα αποθηκεύονται στην αρχική, προσπελάσιμη μορφή τους και προστατεύονται όταν γίνεται ανάγνωσή τους για να εξυπηρετηθούν ερωτήματα στην βάση δεδομένων

Οι προσεγγίσεις αυτές ακολουθούν το παράδειγμα του διαχωρισμού των βάσεων δεδομένων σε συστήματα που απαιτούν τα προς αποθήκευση δεδομένα να τηρούν αυστηρά κάποια δομή πριν αποθηκευτούν (**δομή κατά την εγγραφή** - *schema on write*)[22] και σε συστήματα που αποθηκεύουν τα δεδομένα στην αρχική τους μορφή, ενώ την απαραίτητη δομή την λαμβάνουν κατά την ανάγνωσή τους απο τον αποθηκευτικό χώρο (**δομή κατά την ανάγνωση** - *schema on read*)[23].

Όπως γίνεται αντιληπτό, κι έχουμε αναφέρει ήδη στο Κεφάλαιο 7, εάν τα δεδομένα αποθηκευτούν απροστάτευτα στην βάση δεδομένων, τότε είναι δυνατόν ένας επιτιθέμενος να τα προσπελάσει χωρίς περιορισμούς. Αυτό σημαίνει πως για να εγγυηθούμε την προστασία των δεδομένων θα πρέπει εξ ορισμού να χρησιμοποιείται η αρχή της προστασίας κατά την εγγραφή.

[21] Με αξιοσημείωτη εξαίρεση τις Ιπποκρατικές βάσεις

[22] Τέτοια είναι συνήθως οι σχεσιακές βάσεις δεδομένων (*relational databases*), που λειτουργούν είτε ως **επιχειρησιακές βάσεις** (*operational databases / **O**n**L**ine **T**ransactional **P**rocessing*) είτε ως **αναλυτικές βάσεις** (*data warehouses / **O**n**L**ine **A**nalytical **P**rocessing*)

[23] Τέτοιες είναι συνήθως οι μη σχεσιακές βάσεις δεδομένων, οι επονομαζόμενες βάσεις *NoSQL*, που λειτουργούν υπό την μορφή "**λιμνών δεδομένων**" (*data lakes*)

Ένας συνήθης τρόπος για να επιτευχθεί αυτό, είναι τα δεδομένα να κρυπτογραφούνται πριν την πρώτη αποθήκευση στην βάση και να αποκρυπτογραφούνται κατ'απαίτηση και κατ'ανάγκη. Στην παρούσα διπλωματική, όταν αναφέρεται ο όρος "προστασία" θα εννοείται η κρυπτογράφησή τους.

Στρατηγικές προστασίας δεδομένων

Σημαντικό σημείο της προσέγγισης που παρουσιάζεται στην παρούσα διπλωματική, είναι πως δεν χρειάζεται να προστατευτούν όλα τα δεδομένα της βάσης, αλλά μόνο τα απαραίτητα (όπως φαίνεται στην **Εικόνα 5**). Συνεπώς, τα δεδομένα συναλλαγών και τα μη-προσωπικά κύρια δεδομένα μπορούν να αποθηκευτούν χωρίς προστασία. Αυτά τα δεδομένα, λοιπόν, μπορούν *να αντιμετωπίζονται ως δημοσιευμένα δεδομένα*, τηρουμένων πάντα των νομοθετικών ή άλλων περιορισμών (πχ για λόγους εμπορικής εμπιστευτικότητας, κάποια δεδομένα οφείλονται να μην δημοσιεύονται). Απο την πλευρά όμως της ιδιωτικότητας, τα δεδομένα αυτά δεν παίζουν κάποιον ρόλο, συνεπώς δεν θα μας απασχολήσουν στην διπλωματική.

Τα προσωπικού χαρακτήρα κύρια δεδομένα πρέπει να προστατευτούν. Όμως, και σε αυτήν την κατηγορία δεδομένων υπάρχει διαβάθμιση όσον αφορά στο επίπεδο προστασίας που οφείλουμε να παρέχουμε. Αυτή η διαβάθμιση εξαρτάται κυρίως από τον ρόλο που παίζει κάθε πληροφορία των προσωπικών δεδομένων στην ταυτοποίηση του υποκειμένου επεξεργασίας. Για να μπορέσουμε να αναγνωρίσουμε τον ρόλο, θα πρέπει να εξετάσουμε την σχέση των δεδομένων με τα υποψήφια κλειδιά.

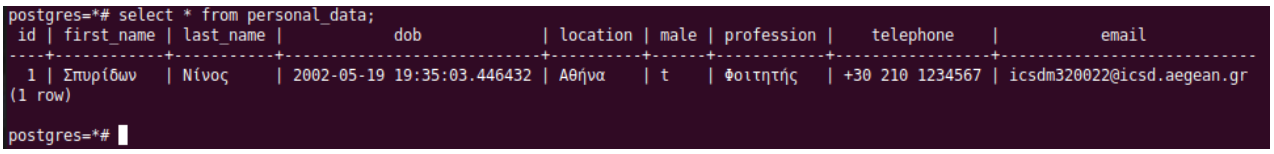
Λειτουργική εξάρτηση και υποψήφια κλειδιά

Στο Κεφάλαιο 6 ("Σχεσιακές Βάσεις Δεδομένων") εξηγήσαμε την λειτουργική εξάρτηση και πως αυτή συνδέει τα υποψήφια κλειδιά με τα υπόλοιπα δεδομένα της κάθε εγγραφής ενός πίνακα. Σε αυτήν την ενότητα θα εξηγήσουμε μια σημαντική ιδιότητα της λειτουργικής εξάρτησης των χαρακτηριστικών ενός πίνακα, η οποία είναι πως η εξάρτηση είναι συνάρτηση 1-1. Πιο συγκεκριμένα:

Ιδιότητα 1-1 λειτουργικής εξάρτησης

Το περιεχόμενο ενός υποψηφίου κλειδιού προσδιορίζει μοναδικά τα υπόλοιπα περιεχόμενα της εγγραφής. Αντίστοιχα, τα περιεχόμενα της εγγραφής που δεν ανήκουν στο υποψήφιο κλειδί προσδιορίζουν μοναδικά το περιεχόμενο του υποψηφίου κλειδιού.

Αυτή η ιδιότητα είναι σημαντικότατη, διότι εξηγεί την αιτια εξαιτίας της οποίας λειτουργούν οι **επιθέσεις επαναταυτοποίησης** (*re-identification attacks*) προσώπων, στην περίπτωση που κάποιος επιτιθέμενος έχει πρόσβαση στα ανωνυμοποιημένα δεδομένα και όχι στα ταυτοποιητικά (δηλαδή το σχετικό υποψήφιο κλειδί). Για να το κατανοήσουμε καλύτερα, ας δούμε ένα παράδειγμα: έστω ότι ένας πίνακας ("*personal_data*") έχει τα εξής χαρακτηριστικά: όνομα (*first_name*), επώνυμο (*last_name*), ημερομηνία γέννησης (*dob*), τοποθεσία (*location*), επιλογή άρρεν/θήλυ (*male*), επάγγελμα (*profession*), τηλέφωνο (*telephone*) και διεύθυνση ηλ. ταχυδρομείου (*email*) (βλ. **Εικόνα 6**).



Εικόνα 6: Πίνακας βάσης "*personal_data*"

(Απο τον πίνακα, ας αγνοήσουμε το πεδίο "*id*", το οποίο παίζει τον ρόλο του υποκατάστατου κλειδιού)

Ας υποθέσουμε πως επιλέγουμε το πρωτεύον κλειδί να αποτελείται από τα πεδία του πίνακα *first_name* και *last_name* γιατί προσδιορίζουν μοναδικά όλα τα υπόλοιπα πεδία του πίνακα. Συνεπώς, γνωρίζοντας το όνομα και το επώνυμο για την συγκεκριμένη εγγραφή ("*Σπυρίδων Νίνος*") γνωρίζουμε και τα υπόλοιπα περιεχόμενα της συγκεκριμένης εγγραφής, δεδομένου πως τα δύο αυτά πεδία προσδιορίζουν μοναδικά τα υπόλοιπα πεδία. Να υπενθυμίσουμε πως το οτι προσδιορίζουν μοναδικά σημαίνει πως το περιεχόμενο των πεδίων του κλειδιού θα πρέπει να προσδιορίζει μοναδικά όλη την εγγραφή, δηλαδή δεν γίνεται ένα κλειδί να αναφέρεται σε δύο ή περισσότερες εγγραφές, ή δύο ή περισσότερα κλειδιά να αναφέρονται στην ίδια εγγραφή. Άρα, για να είναι το ονοματεπώνυμο (στον πίνακα) πρωτεύον κλειδί, σημαίνει πως το κάθε ονοματεπώνυμο θα πρέπει να συναντάται μία μόνο φορά και να αναφέρεται στα συγκεκριμένα υπόλοιπα περιεχόμενα της εγγραφής (δείτε το Κεφάλαιο 6 για περισσότερες πληροφορίες). Επι του συγκεκριμένου παραδείγματος λοιπόν, όταν γνωρίζουμε τα στοιχεία:

- Σπυρίδων (*όνομα*)
- Νίνος (*επώνυμο*)

θα πρέπει απαραίτητα να γνωρίζουμε και τα υπόλοιπα στοιχεία:

- 2002-05-19 (*ημερομηνία γέννησης*)
- Αθήνα (*τοποθεσία*)
- Άρρεν (*φύλο*)
- Φοιτητής (*επάγγελμα*)
- +30 210 1234567 (*τηλέφωνο*)
- icsdm320022@icsd.aegean.gr (*email*)

Αντίστοιχα όμως, γνωρίζοντας τα στοιχεία "*άρρεν, γεννηθείς την 2002-05-19, κατοικεί στην Αθήνα, είναι φοιτητής, το τηλέφωνό του είναι +30 210 1234567 και η διεύθυνση ηλ. ταχυδρομείου του είναι icsdm320022@icsd.aegean.gr*" προκύπτει ότι το ονοματεπώνυμο θα πρέπει να είναι το "*Σπυρίδων Νίνος*".

Εδώ θα πρέπει να υπενθυμίσουμε πως ένας πίνακας μπορεί να έχει περισσότερα του ενός υποψήφια κλειδιά (*εναλλακτικά κλειδιά*), ανεξαρτήτως αν εμείς έχουμε επιλέξει ένα από αυτά για να το χρησιμοποιήσουμε ως πρωτεύον. Στο συγκεκριμένο παράδειγμα, δηλαδή, υποψήφια (*εναλλακτικά*) κλειδιά είναι και το τηλέφωνο αλλά και η διεύθυνση ηλ. ταχυδρομείου, υπο την προϋπόθεση ότι δεν είναι κοινόχρηστα. Συνεπώς, η συγκεκριμένη εγγραφή στον πίνακα μπορεί να προσδιοριστεί μοναδικά και απο το κάθε ένα από τα δύο αυτά πεδία. Επεκτείνοντας το σκεπτικό της 1-1 ιδιότητας της λειτουργικής εξάρτησης μεταξύ των υποψηφίων κλειδιών και των υπολοίπων δεδομένων, συμπεραίνουμε πως η ύπαρξη υποψηφίου κλειδιού σε μια εγγραφή ισοδυναμεί με την δυνατότητα ταυτοποίησης της εγγραφής με κάποιο φυσικό πρόσωπο

(δεδομένου ότι αναφερόμαστε σε προσωπικά δεδομένα). Γι'αυτόν τον λόγο θα πρέπει να λαμβάνουμε υπ'όψιν μας κατά τον σχεδιασμό της βάσης οτι θα πρέπει να προστατέψουμε όλα τα υποψήφια κλειδιά.

Προστασία απλών κλειδιών

Στην περίπτωση των απλών υποψηφίων κλειδιών, θα πρέπει να προστατευτεί εξ ολοκλήρου το κλειδί.

Προστασία σύνθετων κλειδιών

Στην περίπτωση των σύνθετων κλειδιών, μας δίνεται η ευκαιρία να προστατέψουμε τμήμα του κλειδιού, δεδομένου πως το κλειδί για να λειτουργήσει ως κλειδί χρειάζεται όλα τα χαρακτηριστικά (πεδία) συστατικά του. Συνεπώς, αν αφαιρέσουμε κάποιο ή κάποια από τα πεδία του, τότε σταματάει να λειτουργεί ως κλειδί με την αυστηρή έννοια του όρου. Δηλαδή σταματάει να προσδιορίζει με μοναδικό τρόπο μια εγγραφή και προσδιορίζει πλέον μια ομάδα εγγραφών.

Ο λόγος που θα θέλαμε να μην προστατέψουμε ολόκληρο το κλειδί αλλά τμήμα του είναι καθαρά η απόδοση του συστήματος. Επειδή η κρυπτογράφηση απαιτεί υπολογιστικούς πόρους, είναι επιθυμητό να περιορίσουμε την χρήση της κρυπτογράφησης στο ελάχιστο δυνατό. Η μερική προστασία ενός κλειδιού προκαλεί ένα πρόβλημα όμως: με ποιά κριτήρια θα επιλέξουμε τα πεδία που θα προστατευτούν και ποιά θα αφήσουμε απροστάτευτα;

Για να λυθεί το πρόβλημα θα πρέπει να επιστρατεύσουμε τα κριτήρια πληθικότητας[24] των πεδίων:

Κριτήριο φυσικής πληθικότητας (*natural cardinality*): *πόσες τιμές μπορεί να πάρει το πεδίο στον φυσικό κόσμο*

Κριτήριο πληθικότητας στην βάση (*database cardinality*): *πόσες τιμές παίρνει το πεδίο στην βάση*

Το κριτήριο φυσικής πληθικότητας αποτελεί το μέγιστο όριο του κριτηρίου πληθικότητας στην βάση, γιατί στην βάση δεν μπορούν να μπουν σε κάποιο πεδίο περισσότερες τιμές από όσες υπάρχουν στην φύση.

Τα κριτήρια χρησιμοποιούνται ως εξής: όσο πιο μικρά είναι, τόσο πιο μικρή η συμβολή του πεδίου στην μοναδικότητα του κλειδιού. Όσο πιο μεγάλα είναι, τόσο πιο μεγάλη η συμβολή του πεδίου στην μοναδικότητα του κλειδιού (δες Κεφάλαιο 6 για την επεξήγηση). Συνεπώς, εμάς μας ενδιαφέρει *να προστατέψουμε τα πεδία με υψηλή πληθικότητα*. Το πόσα πεδία θα προστατέψουμε και πόσα θα αφήσουμε απροστάτευτα, βασίζεται στο επίπεδο ανωνυμοποίησης που απαιτούμε από την βάση (ακολουθούμε δηλαδή ακριβώς την ίδια προσέγγιση που ακολουθείται στην διαδικασία της κ-ανωνυμοποίησης[25]).

[24] Εδώ η πληθικότητα αναφέρεται στην έννοια που συναντάται στις βάσεις δεδομένων και όχι στην σχεσιακή άλγεβρα, δηλαδή στο πλήθος των διακριτών τιμών που μπορεί να λάβει ένα πεδίο

[25] k-anonymity

Διαχείριση κρυπτογράφησης

Αν και η διαχείριση των κλειδιών κρυπτογράφησης των δεδομένων δεν εμπίπτει στο αντικείμενο της διπλωματικής, εν τούτοις θα γίνει μια σύντομη αναφορά σε μια τεχνική η οποία μπορεί να συντελέσει καθοριστικά στην εφαρμογή της ιδιωτικότητας.

Στον τομέα της διαχείρισης των κλειδιών κρυπτογράφησης υπάρχουν οι έννοιες του **αλατιού** (*salt*) και του **πιπεριού** (*pepper*) ([22])[26]. Το αλάτι είναι μια τυχαία τιμή, η οποία χρησιμοποιείται για να παράξει (συνήθως) έναν κωδικό πρόσβασης σε ένα σύστημα και η αποθηκεύεται μαζί με τον κωδικό. Με αυτόν τον τρόπο αποτρέπονται επιθέσεις εύρεσης κωδικών με πίνακες αντιστοιχίας (*rainbow tables*). Το πιπέρι είναι επίσης μια τυχαία τιμή, η οποία χρησιμοποιείται για την παραγωγή κωδικών, αλλά σε αντίθεση με το αλάτι δεν αποθηκεύεται μαζί με τον παραχθέν κωδικό. Συνήθως το πιπέρι αποθηκεύεται σε άλλο σύστημα, στο οποίο πρόσβαση έχει μόνο ο νόμιμα εξουσιοδοτημένος κάτοχος του κωδικού πρόσβασης.

Αντίστοιχα με την παραπάνω τεχνική, θα μπορούσαμε να ορίσουμε δύο κλειδιά για τις δύο κατηγορίες προστασίας των δεδομένων (βλ. επόμενη ενότητα για ανάλυση όρων):

- **Κλειδί συστήματος** (*system key*): είναι το κλειδί που προστατεύει τα αναγνωριστικά δεδομένα και στο οποίο θα πρέπει έχει πρόσβαση το σύστημα επεξεργασίας των δεδομένων. *Υπάρχει ένα κλειδί για όλα τα αναγνωριστικά δεδομένα*

- **Κλειδι υποκειμένου** (*subject key*): είναι το κλειδί που προστατεύει τα ταυτοποιητικά δεδομένα ενός υποκειμένου και στο οποίο δεν θα πρέπει να έχει πρόσβαση το σύστημα επεξεργασίας δεδομένων, αλλά μόνο το υποκείμενο. *Υπάρχει από ένα κλειδί για κάθε υποκείμενο επεξεργασίας*

Χρησιμοποιώντας τα δύο αυτά κλειδιά, μεγιστοποιούμε την ασφάλεια των δεδομένων και παράλληλα την ιδιωτικότητα των υποκειμένων επεξεργασίας.

Κατηγορίες ταυτοποίησης

Για να μπορέσουμε να πετύχουμε τον στόχο της ιδιωτικότητας δια του σχεδιασμού, θα πρέπει η βάση δεδομένων να είναι σχεδιασμένη με τέτοιον τρόπο ώστε τα δεδομένα να αποθηκεύονται με ασφαλή τρόπο κατά την πρώτη συλλογή τους. Αυτό όμως προϋποθέτει πως θα μπορούμε να ταξινομήσουμε τα δεδομένα σε κατηγορίες ταυτοποίησης, οι οποίες θα υποδεικνύουν τον βαθμό προστασίας που πρέπει να λάβουμε για την κάθε κατηγορία. Οι κατηγορίες, λοιπόν, είναι τρείς (3):

- **Ταυτοποιητικές πληροφορίες** (*personal or directly identifying information*): σε αυτήν την κατηγορία ανήκουν όλες οι πληροφορίες προσωπικού χαρακτήρα, οι οποίες μπορούν *να ταυτοποιήσουν* ένα φυσικό πρόσωπο ως το υποκείμενο επεξεργασίας το οποίο αφορούν οι πληροφορίες ενδιαφέροντος σκοπού. Οι πληροφορίες αυτής της κατηγορίας πρέπει να προστατευτούν εξ ολοκλήρου, με τρόπο που θα τα καθιστά μη-προσπελάσιμα από το σύστημα επεξεργασίας χωρίς την έγκριση του υποκειμένου επεξεργασίας

[26] Στην συγκεκριμένη βιβλιογραφική αναφορά, αντί για τον όρο "πιπέρι" χρησιμοποιείται ο όρος "μυστικό αλάτι"

- **Αναγνωριστικές πληροφορίες** (*quasi-identifying information*): σε αυτήν την κατηγορία ανήκουν όλες οι πληροφορίες προσωπικού χαρακτήρα, οι οποίες από μόνες τους δεν μπορούν να ταυτοποιήσουν κάποιο φυσικό πρόσωπο ως το υποκείμενο επεξεργασίας. Είναι δηλαδή *ανωνυμοποιημένες πληροφορίες*. Βεβαίως, σε συνδυασμό με άλλες πληροφορίες (από άλλες πηγές)[27] θα μπορούν να ταυτοποιήσουν το φυσικό πρόσωπο ως το υποκείμενο. Οι πληροφορίες αυτής της κατηγορίας πρέπει να προστατευτούν εξ ολοκλήρου με τρόπο τέτοιον ώστε να είναι προσπελάσιμα από το σύστημα επεξεργασίας, αλλά να μην είναι δυνατή η προσπέλασή τους χωρίς την χρήση αυτού. Η απαίτηση να προστατεύονται οι πληροφορίες αυτής της κατηγορίας υπάρχει με σκοπό να μειωθεί η επιφάνεια επίθεσης (*attack surface*) των επιθέσεων συνδυασμού δεδομένων (*linkage attacks*), δεδομένου ότι στο σύστημα επεξεργασίας μπορούν να υλοποιηθούν αντίμετρα κατά αυτών. Επίσης, αν κάποιος προσπαθήσει να προσπελάσει την βάση χωρίς την χρήση του συστήματος διαχείρισης της βάσης (όπως συζητήθηκε στο αντίστοιχο μοντέλο απειλών) τότε δεν θα έχει πρόσβαση στα δεδομένα, καθώς θα προστατεύονται από τέτοιου είδους ενέργειες (πχ με την χρήση κρυπτογράφησης)

- **Ενδιαφέροντος σκοπού** (*related to the purpose of collection*): σε αυτήν την κατηγορία ανήκουν οι πληροφορίες οι οποίες *πληρούν τον σκοπό της συλλογής και επεξεργασίας τους*. Για παράδειγμα, εάν ο σκοπός της συλλογής και επεξεργασίας είναι οι αγοραστικές συνήθειες των νοικοκυριών στα αστικά κέντρα, τότε σε αυτήν την κατηγορία εντάσσονται τα αντικείμενα που αγοράζουν οι πελάτες, τα είδη των καταστημάτων, οι ώρες και ημέρες αγορών κ.τ.λ. Ένα άλλο παράδειγμα είναι ο ιατρικός φάκελος ασθενών του εθνικού συστήματος υγείας. Σε αυτήν την περίπτωση, οι πληροφορίες που ανήκουν στην παρούσα κατηγορία είναι οι ασθένειες και οι αγωγές που ακολουθούν οι ασθενείς. Αυτές οι πληροφορίες δεν θα πρέπει να δίνουν την δυνατότητα να προσδιοριστούν τα φυσικά πρόσωπα τα οποία αφορούν. Τα δεδομένα σε αυτήν την κατηγορία δεν είναι απαραίτητο να προστατευτούν[28] με κάποιον τρόπο

Τα δεδομένα που διαθέτει μια βάση, λοιπόν, ανήκουν σε κάποια[29] από τις τρεις προαναφερθείσες κατηγορίες. Βάσει της ανάλυσης που κάναμε στα προηγούμενα κεφάλαια, για έναν πίνακα που περιέχει δεδομένα προσωπικού χαρακτήρα, τα δεδομένα ταξινομούνται ως εξής στις κατηγορίες:

- **Ταυτοποιητικές πληροφορίες** (*identifiers*): σε αυτήν την κατηγορία ανήκουν όλα τα απλά υποψήφια κλειδιά των οντοτήτων που ανήκουν στα κύρια δεδομένα προσωπικού χαρακτήρα, καθώς και τα πεδία των σύνθετων υποψηφίων κλειδιών των κυρίων δεδομένων (προσωπικού χαρακτήρα και μή) τα οποία θέλουμε να προστατεύσουμε

- **Αναγνωριστικές πληροφορίες** (*quasi-identifiers*): σε αυτήν την κατηγορία ανήκουν όλα τα πεδία των σύνθετων υποψηφίων κλειδιών των κυρίων δεδομένων (προσωπικού χαρακτήρα και μη) που επιλέγουμε να μην προστατέψουμε, τα πεδία των κυρίων δεδομένων που δεν αποτελούν χαρακτηριστικά κλειδιού, καθώς και τα πεδία των πινάκων που ανήκουν στις προσωπικές πληροφορίες αλλά δεν αποτελούν χαρακτηριστικά κάποιου κλειδιού

[27] Σε επιθέσεις συνδυασμού δεδομένων (linkage attacks)

[28] Υπενθυμίζουμε ότι με τον όρο "προστασία" αναφερόμαστε στην εμπιστευτικότητα και όχι σε προβλήματα ακεραιότητας ή διαθεσιμότητας

[29] Τα δεδομένα δεν μπορούν να ανήκουν ταυτόχρονα σε περισσότερες από μία κατηγορίες

- **Ενδιαφέροντος σκοπού** (*purpose related*): σε αυτήν την κατηγορία ανήκουν τα δεδομένα συναλλαγών

Αντιμετώπιση καταχρηστικών ερωτημάτων

Συνδυάζοντας τις κατηγορίες δεδομένων, τα μέτρα προστασίας τους και τα είδη δεδομένων που ταξινομούνται σε αυτές, προκύπτει η απαραίτητη προσέγγιση για την εφαρμογή της ιδιωτικότητας στην βάση:

1. Τα απλά υποψήφια κλειδιά και κάποια πεδία σύνθετων υποψηφίων κλειδιών, από τα κύρια δεδομένα προσωπικού χαρακτήρα πρέπει να προστατεύονται εξ ολοκλήρου, με τρόπο που να χρειάζεται έγκριση από το υποκείμενο επεξεργασίας ώστε να προσπελαστούν

2. Τα υπόλοιπα πεδία σύνθετων υποψηφίων κλειδιών (που δεν ανήκουν στην προηγούμενη κατηγορία) και τα υπόλοιπα κύρια δεδομένα προσωπικού χαρακτήρα πρέπει να προστατεύονται εξ ολοκλήρου, όμως με τρόπο που το σύστημα επεξεργασίας να μπορεί να τα προσπελάσει χωρίς την έγκριση του υποκειμένου επεξεργασίας

3. Τα κύρια δεδομένα μη-προσωπικού χαρακτήρα και τα δεδομένα συναλλαγών δεν είναι απαραίτητο να προστατευτούν με κάποιον τρόπο

Η ταξινόμηση των δεδομένων κατ'αυτόν τον τρόπο δίνει το πλεονέκτημα ότι μπορεί να αντιμετωπιστεί η κατάχρηση ερωτημάτων στην βάση. Για να κατανοήσουμε το γιατί, ας θεωρήσουμε ότι ένας επιτιθέμενος αποκτά πρόσβαση στο σύστημα διαχείρισης της βάσης και την ίδια την βάση.

Όσον αφορά στα δεδομένα ενδιαφέροντος σκοπού, από την στιγμή που δεν μπορούν να αξιοποιηθούν για την ταυτοποίηση κάποιου υποκειμένου, τότε δεν παρουσιάζουν κάποιον κίνδυνο στην ιδιωτικότητα, ανεξάρτητα απο την μορφή ή τον αριθμό των ερωτημάτων που μπορεί να κάνει ο επιτιθέμενος.

Για τα αναγνωριστικά δεδομένα, το πρόβλημα της προσπέλασης των δεδομένων εξαρτάται από τις συνθήκες. Αν είναι απλός χρήστης που χρησιμοποιεί το σύστημα διαχείρισης, τότε μπορεί να κάνει όσες ερωτήσεις στην βάση θέλει - αλλά από την στιγμή που τα αναγνωριστικά δεδομένα είναι ανωνυμοποιημένα δεν δημιουργείται κάποιο πρόβλημα (πέραν των επιθέσεων σύνδεσης δεδομένων). Σε περίπτωση που είναι χρήστης που δεν έχει πρόσβαση στο σύστημα διαχείρισης της βάσης τότε, όπως έχει συζητηθεί, δεν θα μπορεί να προσπελάσει τα δεδομένα. Το μόνο πρόβλημα που μπορεί να υπάρξει είναι ότι εάν ο χρήστης έχει δικαιώματα διαχειριστή στο λειτουργικό σύστημα, θα μπορεί να διαβάσει την μνήμη της βάσης και να εξάγει τα κλειδιά προστασίας των δεδομένων. Το οποίο, βεβαίως, ανάγεται σε πρόβλημα επιθέσεων σύνδεσης δεδομένων, καθώς τα δεδομένα είναι ανωνυμοποιημένα.

Τέλος, για τις ταυτοποιητικές πληροφορίες, δεδομένου ότι δεν μπορούν να προσπελαστούν χωρίς την έγκριση των υποκειμένων επεξεργασίας (πχ τα δεδομένα κρυπτογραφούνται και το κλειδί κρυπτογράφησης το έχει μόνο το υποκείμενο επεξεργασίας για τα δεδομένα που το αφορούν), δεν παρουσιάζεται πρόβλημα στην ιδιωτικότητα.

Προστασία των σχέσεων μεταξύ των κατηγοριών

Ένα βήμα το οποίο μπορεί να γίνει και να προσφέρει επιπλέον επίπεδο προστασίας των δεδομένων, είναι η απόκρυψη των σχέσεων μεταξύ των δεδομένων που είναι ταξινομημενα στις διάφορες κατηγορίες. Πιο συγκεκριμένα, εφόσον τα δεδομένα ταξινομούνται στις τρεις κατηγορίες, θα πρέπει να υπάρχουν αναφορές στα δεδομένα μεταξύ των κατηγοριών ώστε να μπορεί κάποιος εξουσιοδοτημένος να συγκροτήσει μια πλήρη εγγραφή. Για παράδειγμα, έστω ότι έχουμε τρεις πίνακες, τους Α, Β και Γ ως εξής:

1. **Α**: περιέχει δεδομένα ταυτοποιητικά
2. **Β**: περιέχει δεδομένα αναγνωριστικά, έχει **ξένο κλειδί** (*foreign key*) για τον Α
3. **Γ**: περιέχει δεδομένα σκοπού συλλογής, έχει **ξένο κλειδί** (*foreign key*) για τον Β

Ένας εξουσιοδοτημένος χρήστης θα πρέπει να μπορεί να συνδυάσει τα δεδομένα από τους τρεις πίνακες ώστε να εξάγει πλήρεις εγγραφές. Για να γίνει αυτό θα πρέπει να χρησιμοποιήσει τις **αναφορές** (*ξένα κλειδιά*) από τους πίνακες Β και Γ.

Σε περίπτωση που οι αναφορές αυτές είναι απροστάτευτες και μπορεί να τις προσπελάσει ο οποιοσδήποτε, τότε το πρόβλημα του επιτιθέμενου είναι απλά το πως θα αναγνώσει τις πληροφορίες από τον πίνακα Α και (ενδεχομένως) απο τον Β. Σε περίπτωση όμως που οι αναφορές αυτές είναι προστατευμένες και δεν μπορεί να τις προσπελάσει ο οποιοσδήποτε (πχ οι αναφορές είναι κρυπτογραφημένες ή τυχαιοποιημένες) τότε προστίθεται το πρόβλημα της αντιστοιχίας των σωστών δεδομένων μεταξύ των τριών πινάκων. Αυτό το πρόβλημα πολλαπλασιάζει την προσπάθεια ενός επιτιθέμενου στο να ανακτήσει τις πληροφορίες της βάσης.

Μεταφορά Πεδίων

Υπάρχουν περιπτώσεις που για λόγους απόδοσης δεν θέλουμε να προστατέψουμε τις αναφορές μεταξύ πινάκων. Αυτό επειδή τα αναλυτικά ερωτήματα απαιτούν συνδυασμό πληροφοριών από αρκετούς πίνακες (*table joins*), οπότε το να προστατεύουμε (κρυπτογραφούμε) τις αναφορές (*ξένα κλειδιά*) μεταξύ πινάκων ουσιαστικά εμποδίζει την λειτουργία των αναλυτικών ερωτημάτων σε μια βάση. Για να μπορέσουμε να ξεπεράσουμε αυτό το εμπόδιο, εφαρμόζουμε την διαδικασία της **μεταφοράς πεδίων** ως εξής:

- επιλέγουμε έναν πίνακα, που περιέχει ήδη πεδία προστατευμένα ή έναν καινούργιο πίνακα (που τον ονομάζουμε "**προστατευμένη επέκταση**")

- μεταφέρουμε τα προς προστασία πεδία, από τους αρχικούς πίνακες που τα περιέχουν στην προστατευμένη επέκταση (εννοείται ότι αφαιρούμε τα μεταφερμένα πεδία από τους πίνακες που βρίσκονταν αρχικά)

- κάνουμε αναφορά στον πίνακα αυτόν από τον πίνακα που περιέχει τα κύρια δεδομένα προσωπικού χαρακτήρα (που τον ονομάζουμε "**προστατευμένη βάση**"), στην περίπτωση που η προστατευμένη επέκταση είναι διαφορετικός πίνακας από την προστατευμένη βάση

Ουσιαστικά λοιπόν, προβαίνουμε σε μια μορφή *επιλεγμένης αποκανονικοποίησης* της βάσης, χρησιμοποιώντας ως καθοριστικό κριτήριο την ιδιωτικότητα. Κατ'αυτόν τον τρόπο επιτυγχάνεται

και ο σκοπός της αποκανονικοποίησης της βάσης, που είναι η απόδοση στην ταχύτητα των ερωτημάτων.

Διαδικασία Κανονικοποίησης με Υλοποίηση Ιδιωτικότητας

Έχοντας αναλύσει τα επιμέρους βήματα και ενέργειες που απαιτούνται για να σχεδιάσουμε μια βάση η οποία θα υλοποιήσει την ιδιωτικότητα των υποκειμένων επεξεργασίας, μπορούμε πλέον να παραθέσουμε ολοκληρωμένα την διαδικασία του σχεδιασμού:

1. Επιλέγουμε τα δεδομένα που θέλουμε να συλλέξουμε για τον σκοπό της επεξεργασίας
2. Διαχωρίζουμε τα κύρια δεδομένα από τα δεδομένα συναλλαγών
3. Διαχωρίζουμε τα κύρια δεδομένα προσωπικού χαρακτήρα (από εδώ και πέρα θα αναφέρονται απλά ως "δεδομένα") από τα κύρια δεδομένα μη-προσωπικού χαρακτήρα
4. Βρίσκουμε όλα τα υποψήφια κλειδιά των δεδομένων (απλά και σύνθετα)
5. Επιλέγουμε ποιά πεδία των σύνθετων υποψηφίων κλειδιών θα προστατέψουμε, βάσει των κριτηρίων πληθικότητας, σύμφωνα με τις απαιτήσεις ανωνυμίας που θα έχουμε για τα αναγνωριστικά δεδομένα
6. Ταξινομούμε τα δεδομένα στις κατηγορίες των ταυτοποιητικών και αναγνωριστικών δεδομένων
7. Κανονικοποιούμε τα δεδομένα σύμφωνα με τους κανόνες της σχεσιακής άλγεβρας
8. Κανονικοποιούμε τα κύρια δεδομένα μη-προσωπικού χαρακτήρα και τα δεδομένα συναλλαγών
9. Αποφασίζουμε αν θα προστατέψουμε τις σχέσεις μεταξύ των κατηγοριών των δεδομένων με τα κλειδιά συστήματος και υποκειμένων, βάσει του βήματος (5), ή αν θα προβούμε στην διαδικασία της *μεταφοράς πεδίων*

Όσον αφορά στο σύστημα επεξεργασίας, μένει μόνο ένα βήμα προς υλοποίηση:

10. Σχεδιάζουμε το σύστημα επεξεργασίας (που θα περιέχει το κλειδί συστήματος) και το σύστημα παροχής πρόσβασης στα κλειδιά υποκειμένων

Όπως βλέπουμε, η προσέγγιση αυτή προσθέτει μόνο 5 βήματα στην κανονική διαδικασία σχεδιασμού μιας σχεσιακής βάσης (βήματα 2, 3, 5, 6 και 9), καθώς τα υπόλοιπα είναι κοινά.

Παραδείγματα Κανονικοποίησης με Υλοποίηση Ιδιωτικότητας

Σε αυτήν την ενότητα θα δώσουμε τρία παράδειγμα σχεδιασμού μιας βάσης:

- στο πρώτο θα υλοποιηθεί μια απλή κανονικοποίηση
- στο δεύτερο θα γίνει κανονικοποίηση με υλοποίηση της ιδιωτικότητας για συστήματα επεξεργασίας συναλλαγών (*OLTP - Online Transaction Processing*)

- στο τρίτο θα γίνει κανονικοποίηση με υλοποίηση της ιδιωτικότητας, με κάποια βελτιστοποίηση για συστήματα αναλυτικής επεξεργασίας (*OLAP - Online Analytical Processing*)

Για την υλοποίηση των παραδειγμάτων, θα θεωρήσουμε πως θέλουμε να σχεδιάσουμε την βάση για ένα σύστημα επεξεργασίας ιστορικού νοσηλειών για το Εθνικό Σύστημα Υγείας της Ελλάδας. Για το συγκεκριμένο σύστημα, αποφασίζουμε ότι θα χρειαστούμε τα παρακάτω δεδομένα για τον κάθε πολίτη:

1. Όνομα (first_name)
2. Επώνυμο (last_name)
3. Πατρώνυμο (fathers_name)
4. Μητρώνυμο (mothers_name)
5. Ημερομηνία γεννήσεως (date of birth - dob)
6. Φύλο (sex)
7. ΑΜΚΑ (social security number - ssn)
8. ΑΦΜ (VAT number - vat_num)
9. Αριθμός σταθερού τηλεφώνου (landline_num)[30]
10. Αριθμός κινητού τηλεφώνου (mobile_phone)
11. Διεύθυνση ηλ. ταχυδρομείου (email)
12. Χώρα κατοικίας[31] (country)
13. Πόλη κατοικίας (city)
14. Ταχυδρομικός Κώδικας περιοχής κατοικίας(postal_code)
15. Οδός κατοικίας (street)
16. Αριθμός οδού κατοικίας (street_no)
17. Οικογενειακή κατάσταση (marital_status)
18. Μορφωτικό επίπεδο (education_level)
19. Επάγγελμα (profession)
20. Νοσοκομείο (hospital)
21. Ημερομηνία εισαγωγής (admission_datetime)
22. Ημερομηνία εξιτηρίου (discharge_datetime)
23. Θεράπων ιατρός (treating_doctor)
24. Διάγνωση (diagnosis)
25. Διαγνωστικές εξετάσεις (medical_examinations)
26. Φάρμακο θεραπείας (treatment_medicine)

Αιτιολόγηση σκοπού επεξεργασίας

Το σκεπτικό πίσω από την συλλογή των συγκεκριμένων δεδομένων είναι το εξής: τα δεδομένα 1-19 τα χρειαζόμαστε για να μπορούμε να ταυτοποιήσουμε το υποκείμενο επεξεργασίας. Τα δεδομένα 11 έως 26 τα χρειαζόμαστε επίσης για στατιστικά και εξόρυξη δεδομένων (πχ εξαγωγή ιστορικού νοσηλειών ομάδων πληθυσμού). Να τονίσουμε ότι τα δεδομένα 11-19 έχουν διπλή χρησιμότητα, μπορούν και *να αναγνωρίσουν το υποκείμενο επεξεργασίας αλλά και να συμμετασχουν σε συγκεκριμένου τύπου στατιστικά*.

[30] Θεωρούμε (όπως και για τον αριθμό κινητού τηλεφώνου και την διεύθυνση ηλ. ταχυδρομείου) πως δεν είναι κοινόχρηστα και ανήκουν αποκλειστικά στο υποκείμενο επεξεργασίας

[31] Το συμπεριλαμβάνουμε για λόγους πληρότητας - δεδομένου ότι όλοι οι πολίτες θα πρέπει να κατοικούν στην Ελλάδα, το πεδίο δεν προσφέρει κάποια ουσιαστική πληροφορία

Παράδειγμα 1: Απλή κανονικοποίηση

Για να σχεδιάσουμε μια σχεσιακή βάση, πρέπει να εφαρμόσουμε τις 3 πρώτες φόρμες κανονικοποίησης[32].

1η φόρμα

Σύμφωνα με αυτήν την φόρμα, δεν θα πρέπει να υπάρχουν επαναλαμβανόμενες ομάδες δεδομένων στα πεδία. Δεδομένου πως μπορεί ένας άνθρωπος να νοσηλευτεί πάνω από 1 φορά στο ίδιο νοσοκομείο, αμέσως προκύπτει πως όλα τα στοιχεία του ανθρώπου (*ομάδα δεδομένων*) θα επαναληφθούν όσες φορές νοσηλευτεί. Συνεπώς, τα δεδομένα πρέπει να χωριστούν σε δύο, τουλάχιστον, πίνακες:

Πίνακας 1: Δεδομένα 1-18. Μια βελτιστοποίηση που πρέπει να εισάγουμε είναι ένα υποκατάστατο κλειδί, ώστε οι αναφορές στον πίνακα να μην αποτελούνται από σύνθετα κλειδιά

Πίνακας 2: Δεδομένα 19-26, με ένα επιπλέον πεδίο ("*ασθενής*"), το οποίο (όπως και το πεδίο 20, "*θεράπων ιατρός*") περιέχει αναφορά (*ξένο κλειδί*) στον πίνακα 1

2η φόρμα

Σύμφωνα με αυτήν την φόρμα, όλα τα πεδία θα πρέπει να εξαρτώνται από το σύνολο του (πρωτεύοντος) κλειδιού (*στην περίπτωση που είναι σύνθετο*). Συνεπώς, θα πρέπει να βρούμε το κλειδί του κάθε πίνακα. Για να γίνει αυτό, θα πρέπει να εξετάσουμε τις λειτουργικές εξαρτήσεις μεταξύ των πεδίων του κάθε πίνακα. Για τους δοθέντες πίνακες, έχουμε βρεί τα εξής:

1. Για τον **πίνακα 1**: επιλέγουμε ως πρωτεύον κλειδί ("ΠΚ") την ομάδα πεδίων: *όνομα, επώνυμο, πατρώνυμο, μητρώνυμο, ημ. γεννήσεως*. Οι εξαρτήσεις που προκύπτουν είναι οι εξής:
    a. ΠΚ →[33]διεύθυνση κατοικίας (η ανάλυση θα πρέπει να εξετάζει το κάθε μέρος της διεύθυνσης ξεχωριστά, αλλά για λόγους συντομίας αναφέρουμε την διεύθυνση ως σύνολο)
    b. ΠΚ → οικογενειακή κατάσταση
    c. ΠΚ → μορφωτικό επίπεδο
    d. ΠΚ → επάγγελμα
2. Για τον **πίνακα 2**: επιλέγουμε ως πρωτεύον κλειδί την ομάδα πεδίων: *ασθενής, νοσοκομείο, ημερομηνία εισαγωγής*. Οι εξαρτήσεις που προκύπτουν είναι οι εξής:
    a. ΠΚ → ημερομηνία εξιτηρίου
    b. ΠΚ → θεράπων ιατρός
    c. ΠΚ → διάγνωση
    d. ΠΚ → διαγνωστικές εξετάσεις
    e. ΠΚ → φάρμακο θεραπείας

[32] Όπως έχουμε αναφέρει, γενικά υπάρχουν περισσότερες φόρμες κανονικοποίησης, αλλά για τον σκοπό της διπλωματικής θα περιοριστούμε στις τρεις πρώτες

[33] Το δεξιάς φοράς βέλος απεικονίζει λειτουργική εξάρτηση ως εξής: το δεξί μέρος εξαρτάται από το αριστερό. Δηλαδή το βέλος αποδίδεται εννοιολογικά ως "*ορίζει*"

3η φόρμα

Σύμφωνα με αυτήν την φόρμα, όλα τα πεδία που δεν είναι μέρος του κλειδιού θα πρέπει να εξαρτώνται αποκλειστικά από αυτό και όχι από άλλο πεδίο. Εξετάζοντας τους δύο πίνακες, βλέπουμε τα εξής:

- Πίνακας 1: τα επιμέρους στοιχεία της διεύθυνσης δεν παρουσιάζουν κάποια εξάρτηση από το πρωτεύον κλειδί. Αντιθέτως, η εξάρτηση είναι ως εξής:
    - Αριθμός οδού → οδός
    - Οδός → Ταχυδρομικός κώδικας[34]
    - Ταχυδρομικός κώδικας → πόλη
    - Πόλη → χώρα

- Πίνακας 2: όλα τα πεδία εξαρτώνται από το πρωτεύον κλειδί

Μια βελτιστοποίηση που θα κάνουμε στο συγκεκριμένο παράδειγμα (που είναι σύνηθες σε πραγματικά συστήματα βάσεων δεδομένων) είναι πως τα πεδία που περιέχουν την οικογενειακή κατάσταση, το μορφωτικό επίπεδο, το επάγγελμα, το νοσοκομείο, τον θεράποντα ιατρό, την διάγνωση, τις διαγνωστικές εξετάσεις και τα φάρμακα θεραπείας θα πρέπει να εξαχθούν σε άλλους πίνακες, για λόγους εξοικονόμησης αποθηκευτικού χώρου της βάσης.

Συνεπώς, οι πίνακες που προκύπτουν απο την κανονικοποίηση των δεδομένων, είναι οι εξής:

- Πίνακας: "**πολίτες**"[35]:
    - υποκατάστατο κλειδί (*αύξων αριθμός*)
    - όνομα
    - επώνυμο
    - πατρώνυμο
    - μητρώνυμο
    - ημ. γεννήσεως
    - φύλο
    - ΑΜΚΑ
    - ΑΦΜ
    - Αριθμός σταθερού τηλεφώνου
    - Αριθμός κινητού τηλεφώνου
    - Διεύθυνση ηλ. ταχυδρομείου
    - Διεύθυνση (*αναφορά στον αντίστοιχο πίνακα*)
    - Οικογενειακή κατάσταση (*αναφορά στον αντίστοιχο πίνακα*)
    - Μορφωτικό επίπεδο (*αναφορά στον αντίστοιχο πίνακα*)
    - Επάγγελμα (*αναφορά στον αντίστοιχο πίνακα*)

- Πίνακας "**διεύθυνση**"[36]

[34] Τα ονόματα των οδών είναι μοναδικά για κάθε ταχυδρομικό κώδικα, επαναλαμβάνονται ανά περιοχές

[35] Η ονοματοδοσία των πινάκων δεν ακολουθεί συγκεκριμένη πρακτική, χρησιμοποιείται πληθυντικός για διακριτές οντότητες και ενικός για έννοιες

[36] Στην πράξη ο πίνακας αυτός θα έπρεπε να σπάσει περαιτέρω, για βελτιστοποίηση του αποθηκευτικού χώρου της βάσης, αλλά για λόγους απλότητας δεν θα το πράξουμε στην παρούσα διπλωματική

  - υποκατάστατο κλειδί (*αύξων αριθμός*)
  - αριθμός οδού
  - οδός
  - ταχυδρομικός κώδικας
  - πόλη
  - χώρα

- Πίνακας "**οικογενειακή κατάσταση**" (με υποκατάστατο κλειδί - αύξοντα αριθμό)

- Πίνακας "**μορφωτικό επίπεδο**" (με υποκατάστατο κλειδί - αύξοντα αριθμό)

- Πίνακας "**επαγγέλματα**" (με υποκατάστατο κλειδί - αύξοντα αριθμό)

- Πίνακας "**νοσοκομεία**" (με υποκατάστατο κλειδί - αύξοντα αριθμό)

- Πίνακας "**διαγνώσεις**" (με υποκατάστατο κλειδί - αύξοντα αριθμό)

- Πίνακας "**διαγνωστικές εξετάσεις**" (με υποκατάστατο κλειδί - αύξοντα αριθμό)

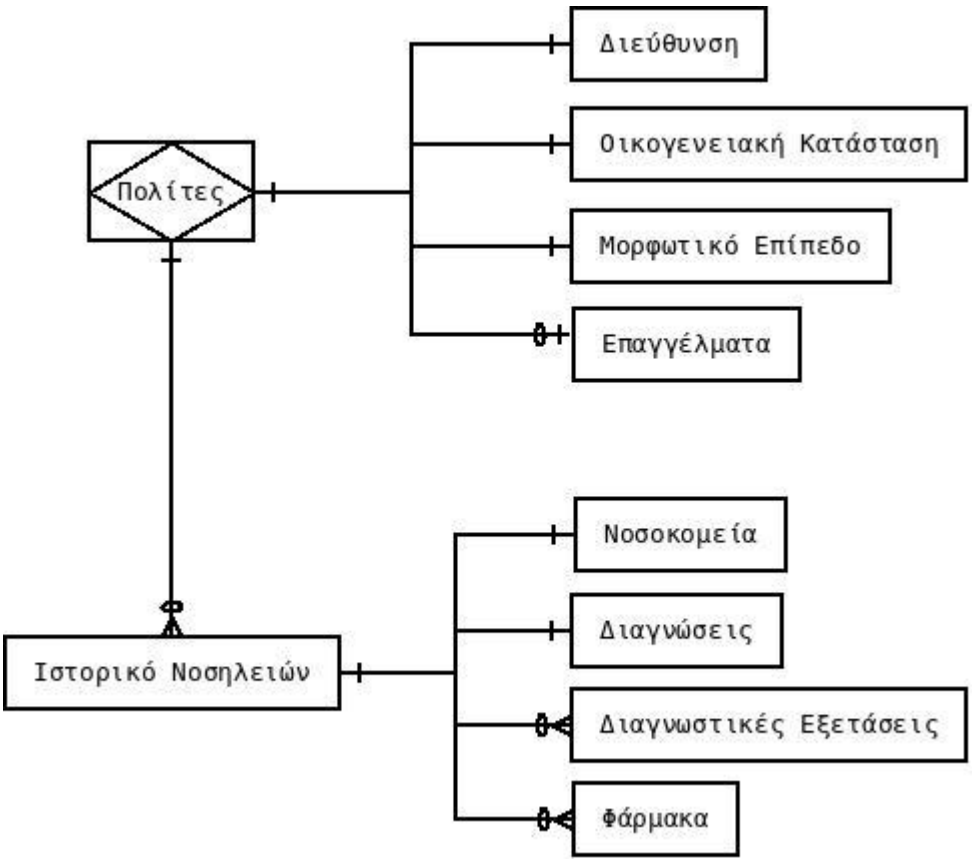
- Πίνακας "**φάρμακα**" (με υποκατάστατο κλειδί - αύξοντα αριθμό)

- Πίνακας "**ιστορικό νοσηλειών**":
  - υποκατάστατο κλειδί (*αύξων αριθμός*)
  - ασθενής (*αναφορά στον αντίστοιχο πίνακα*)
  - νοσοκομείο (*αναφορά στον αντίστοιχο πίνακα*)
  - θεράπων ιατρός (*αναφορά στον αντίστοιχο πίνακα*)
  - ημερομηνία εισαγωγής
  - ημερομηνία εξιτηρίου
  - διάγνωση (*αναφορά στον αντίστοιχο πίνακα*)
  - διαγνωστικές εξετάσεις (*αναφορά στον αντίστοιχο πίνακα*)
  - φάρμακα (*αναφορά στον αντίστοιχο πίνακα*)

<u>Διάγραμμα Οντοτήτων-Σχέσεων</u>[37]

Το διάγραμμα οντοτήτων-σχέσεων που προκύπτει απο την σχεδίαση της βάσης φαίνεται στην **Εικόνα 7**:

[37] Entity-Relationship (ER) diagram



Εικόνα 7: Διάγραμμα οντοτήτων-σχέσεων

Παράδειγμα 2: Κανονικοποίηση με Υλοποίηση Ιδιωτικότητας (OLTP)

Χρησιμοποιούμε τα βήματα της διαδικασίας κανονικοποίησης, που προτάθηκε στην προηγούμενη ενότητα, με υλοποίηση ιδιωτικότητας για να σχεδιάσουμε την βάση:

Βήμα 1: Επιλέγουμε τα δεδομένα που θέλουμε να συλλέξουμε για τον σκοπό της επεξεργασίας

Παίρνουμε τα δεδομένα από την εισαγωγή της ενότητας

Βήμα 2: Διαχωρίζουμε τα κύρια δεδομένα από τα δεδομένα συναλλαγών

Παρατηρώντας τα δεδομένα που θέλουμε να συλλέξουμε για να σχεδιάσουμε την βάση, βλέπουμε τα εξής:

- Τα δεδομένα 1 έως και 20, καθώς και τα 23 έως και 26 είναι *κύρια δεδομένα*
- Τα δεδομένα 21 και 22 είναι *δεδομένα συναλλαγής*

Βήμα 3: Διαχωρίζουμε τα κύρια δεδομένα προσωπικού χαρακτήρα από τα κύρια δεδομένα μη-προσωπικού χαρακτήρα

Αναλύοντας τα κύρια δεδομένα από το Βήμα 2 προκύπτει ότι:

- Τα δεδομένα 1 έως και 19 είναι *κύρια δεδομένα προσωπικού χαρακτήρα*
- Τα δεδομένα 20, 23, 24, 25 και 26 είναι *κύρια δεδομένα μη προσωπικού χαρακτήρα*

Απο την ανάλυση των προηγούμενων κεφαλαίων προκύπτει πως *τα δεδομένα που πρέπει να προστατέψουμε είναι τα δεδομένα 1 έως και 19*.

Βήμα 4: Βρίσκουμε όλα τα υποψήφια κλειδιά των δεδομένων (απλά και σύνθετα)

Τα υποψήφια κλειδιά για τα δεδομένα 1-19 είναι τα εξής:

1. Όνομα, επώνυμο, πατρώνυμο, μητρώνυμο, ημ. γεννήσεως: *σύνθετο κλειδί*
2. ΑΜΚΑ: *απλό κλειδί* (φυσικό)
3. ΑΦΜ: *απλό κλειδί* (φυσικό)
4. Αριθμός σταθερού τηλεφώνου: *απλό κλειδί* (φυσικό)
5. Αριθμός κινητού τηλεφώνου: *απλό κλειδί* (φυσικό)
6. Διεύθυνση ηλ. ταχυδρομείου: *απλό κλειδί* (φυσικό)
7. Χώρα, πόλη, ταχυδρομικός κώδικας, οδός, αριθμός οδου κατοικίας, ημ. γεννήσεως: *σύνθετο κλειδί*
8. Χώρα, πόλη, ταχυδρομικός κώδικας, οδός, αριθμός οδου κατοικίας, επάγγελμα: *σύνθετο κλειδί*

Βήμα 5: Επιλέγουμε ποιά πεδία των σύνθετων υποψηφίων κλειδιών θα προστατέψουμε, βάσει των κριτηρίων πληθικότητας, σύμφωνα με τις απαιτήσεις ανωνυμίας που θα έχουμε για τα αναγνωριστικά δεδομένα

Απο το *Βήμα 4* προέκυψαν 8 υποψήφια κλειδιά. Τα σύνθετα κλειδιά είναι τα **1**, **7** και **8**. Επειδή το κλειδί **1** έχει ήδη επιλεχθεί ως πρωτεύον κλειδί, θεωρούμε ότι θα προστατεύσουμε (κρυπτογραφήσουμε) όλα τα πεδία του. Θα μπορούσαμε να δούμε αν μπορούμε να αφήσουμε το πεδίο "ημ. γεννήσεως" εκτός προστασίας, όμως παρουσιάζει υψηλή πληθικότητα (και φυσική αλλά και στην βάση δεδομένων), οπότε επειδή χρησιμοποιείται και στο κλειδί **7**, θα το προστατεύσουμε.

Για το κλειδί **7**, το πεδίο "*ημ. γεννήσεως*" προστατεύεται λόγω συμμετοχής του στο κλειδί **1**. Απο τα υπόλοιπα πεδία, αυτά που παρουσιάζουν υψηλή πληθικότητα είναι τα πεδία "*οδός*" και "*αριθμός οδού*". Αν αυτά τα πεδία προστατευτούν, τότε οι πληροφορίες που παρέχονται από τον συνδυασμό των πεδίων "*χώρα*", "*πόλη*" και "*ταχυδρομικός κώδικας*" είναι αρκούντως ανωνυμοποιημένες. Συνεπώς, επιλέγουμε να προστατεύσουμε τα πεδία "*οδός*" και "*αριθμός οδού*".

Τέλος, για το κλειδί **8**, μένει να εξετάσουμε το πεδίο "*επάγγελμα*", το οποίο παρουσιάζει υψηλή πληθικότητα. Οπότε θα το προστατεύσουμε.

Βήμα 6: Ταξινομούμε τα δεδομένα στις κατηγορίες των ταυτοποιητικών και αναγνωριστικών δεδομένων

Ταξινομούμε τα δεδομένα στις κατηγορίες ταυτοποίησης ως εξής:

- Τα δεδομένα 1 έως και 11 ανήκουν στις ταυτοποιητικές πληροφορίες
- Τα δεδομένα 12 έως και 19 ανήκουν στις αναγνωριστικές πληροφορίες

Οι λόγοι που ταξινομούμε με αυτόν τον τρόπο τις πληροφορίες παρατίθενται παρακάτω:

- Τα δεδομένα 1 έως και 11 είναι πληροφορίες που ταυτοποιούν το υποκείμενο επεξεργασίας. Επίσης, τα πεδία *ΑΜΚΑ*, *ΑΦΜ*, *σταθερό τηλέφωνο*, *κινητό τηλέφωνο* και *διεύθυνση ηλ. ταχυδρομείου* αναφέρονται κατά μοναδικό τρόπο στο υποκείμενο επεξεργασίας, δεδομένου ότι αποτελούν φυσικά κλειδιά
- Τα δεδομένα 12 έως και 19 μπορούν να συνδεθούν με άλλα δεδομένα και να ταυτοποιήσουν το υποκείμενο επεξεργασίας. Για παράδειγμα, σε συγκεκριμένη διεύθυνση, η πιθανότητα να έχουν δύο άνθρωποι την ίδια ημερομηνία γεννήσεως είναι αμελητέα (αρκεί να σκεφτούμε ότι το 2021 γεννήθηκαν συνολικά[38] 84.767 άτομα, οπότε με ισοκατανομή ανα ημέρα, υπολογίζονται ότι γεννήθηκαν περίπου 232 άνθρωποι την κάθε ημέρα. Η πιθανότητα να μένουν στην ίδια διεύθυνση 2 ή περισσότεροι από τους 232 αυτούς ανθρώπους είναι αμελητέα). Συνεπώς, αν συνδεθούν τα δεδομένα με άλλο ένα στοιχείο (ή περισσότερα), τότε μπορούμε στην πράξη να ταυτοποιήσουμε το υποκείμενο επεξεργασίας

Βήμα 7: Κανονικοποιούμε τα δεδομένα σύμφωνα με τους κανόνες της σχεσιακής άλγεβρας

Η κανονικοποίηση γίνεται όπως στο **παράδειγμα 1**

Βήμα 8: Κανονικοποιούμε τα κύρια δεδομένα μη-προσωπικού χαρακτήρα και τα δεδομένα συναλλαγών

Η κανονικοποίηση γίνεται όπως στο **παράδειγμα 2**

Βήμα 9: Αποφασίζουμε αν θα προστατέψουμε τις σχέσεις μεταξύ των κατηγοριών των δεδομένων με τα κλειδιά συστήματος και υποκειμένων, βάσει του βήματος (5), ή αν θα αποκανονικοποιήσουμε μέρος της βάσης για να συγκεντρώσουμε τα προστατευμένα δεδομένα

Στο συγκεκριμένο παράδειγμα, επειδή μας ενδιαφέρει η βάση να λειτουργεί ως OLTP, μπορούμε να επιλέξουμε να προστατέψουμε τις σχέσεις μεταξύ των κατηγοριών των δεδομένων, αντί να αποκανονικοποιήσουμε την βάση. Για να γίνει αυτό, θα πρέπει να πάμε στην κανονικοποιημένη βάση (βλ. **Εικόνα 7**), να σημειώσουμε τους πίνακες οι οποίοι περιέχουν προστατευμένα πεδία. Οπότε, προστατεύουμε τις αναφορές που γίνονται από τους πίνακες που περιέχουν τα κύρια δεδομένα προσωπικού χαρακτήρα στους σημειωμένους πίνακες.

Πιο συγκεκριμένα, για τους πίνακες του παραδείγματος, βρήκαμε ότι οι πίνακες "**διευθύνσεις**" και "**επαγγέλματα**" περιέχουν προστατευμένα πεδία. Οπότε, στον πίνακα "**πολίτες**" θα προστατεύσουμε τις αναφορές προς τους συγκεκριμένους πίνακες. Αυτό έχει ως αποτέλεσμα να μην μπορούν να συνδεθούν οι πίνακες χωρίς την προμήθεια του αντιστοίχου κλειδιού προστασίας.

[38] Ελληνικη Στατιστική Υπηρεσία, αναφορά για το 2021, https://www.statistics.gr/documents/20181/17529706/GreeceInFigures_2021Q4_GR.pdf/63a2d3b8-23d5-2376-a189-a1446ec6ffaa (τελευταία επίσκεψη Μάιος 2022)

Παράδειγμα 3: Κανονικοποίηση με Υλοποίηση Ιδιωτικότητας (OLAP)

Η σχεδίαση της βάσης με βελτιστοποιήσεις για αναλυτικά ερωτήματα (OLAP), ακολουθεί την ίδια διαδικασία σχεδιασμού με την σχεδίαση για επεξεργασία συναλλαγών (OLTP), με εξαίρεση το **Βήμα 9**.

Η διαφορά έγκειται στο ότι επιλέγουμε να μην προστατεύσουμε τις αναφορές μεταξύ πινάκων, αλλά να εφαρμόσουμε την *μεταφορά πεδίων* (διαδικασία που αναφέρθηκε σε προηγούμενη ενότητα).

Στην πράξη λοιπόν, βασιζόμενοι στο διάγραμμα οντοτήτων-σχέσεων της **Εικόνας 7** και στο **Βήμα 5** (που περιγράφεται στο προηγούμενο παράδειγμα) βρίσκουμε οτι:

- Στον πίνακα "**διευθύνσεις**" έχουμε τα πεδία προς προστασία "*αριθμός οδού*" και "*οδός*"
- Στον πίνακα "**επαγγέλματα**" έχουμε το μοναδικό πεδίο προς προστασία, το "*επάγγελμα*"

Για λόγους απλότητας, επιλέγουμε να μεταφέρουμε τα πεδία "*αριθμός οδού*" και "*οδός*" από τον πίνακα "**διευθύνσεις**" στον πίνακα "**πολίτες**" (που είναι *η προστατευμένη βάση*).

Σε αυτό το σημείο προκύπτει το εξής ζήτημα: αν προστατεύσουμε το πεδίο "*επάγγελμα*" μεταφέροντάς το στην *προστατευμένη βάση*, τότε θα πρέπει να έχουμε δύο στιγμιότυπά του: ένα ως κύριο δεδομένο, το οποίο θα χρησιμοποιείται για την επιλογή του επαγγέλματος του υποκειμένου κατά την εισαγωγή των στοιχείων του στην βάση δεδομένων, και μία στην *προστατευμένη βάση* (πίνακα)[39]. Για να λύσουμε αυτό το πρόβλημα, πρέπει να επανεξετάσουμε τα συμπεράσματα από το **Βήμα 5** για το κλειδί **8** και να δούμε μήπως μπορούμε να πετύχουμε ισοδύναμο βαθμό ιδιωτικότητας για το κλειδί, προστατεύοντας άλλο πεδίο αντί για το συγκεκριμένο.

Πράγματι, μπορούμε να πετύχουμε ισοδύναμη ιδιωτικότητα, αν αντί να προστατέψουμε το πεδίο "*επάγγελμα*" προστατέψουμε το πεδίο "*ταχυδρομικός κώδικας*". Αυτο διαισθητικά φαίνεται από την εξέταση του συνδυασμού "*χώρα, πόλη, επάγγελμα*", ο οποίος δίνει πολύ γενικά αποτελέσματα. Το ίδιο ισχύει και για τον ολικό συνδυασμό των μη-προστατευμένων πεδίων απο την *προστατευμένη βάση* (πίνακα), δηλαδή ο συνδυασμός "*φύλο*", "*χώρα*", "*πόλη*", "*οικογενειακή κατάσταση*", "*μορφωτικο επίπεδο*" και "*επάγγελμα*" δίνουν πολύ γενικά αποτελέσματα.

Συνεπως, αν θεωρήσουμε ότι προτιμούμε να μειωθεί η τοπογραφική ακρίβεια των στατιστικών (αναγωγή στο επίπεδο των πόλεων και όχι στο επίπεδο των περιοχών των πόλεων) αλλά να διατίθεται η κατηγορία τους επαγγέλματος, τότε μπορούμε να αφήσουμε απροστάτευτες τις αναφορές (*ξένα κλειδιά*) μεταξύ των πινάκων και να αλλάξουμε τα περιεχόμενά τους ως εξής:

- Πίνακας: "**πολίτες**":
    - υποκατάστατο κλειδί (*αύξων αριθμός*)
    - όνομα
    - επώνυμο
    - πατρώνυμο
    - μητρώνυμο
    - ημ. γεννήσεως

[39] Το ίδιο συμβαίνει επίσης και για τα πεδία "*οδός*" και "*αριθμός οδού*". Όμως, σχετικά με τις οδούς, επειδή είναι πάρα πολλές ανά χώρα, καθώς και αρκετές από αυτές θα μείνουν αχρησιμοποίητες στην βάση, δεν ενδείκνυνται να χρησιμοποιούνται σε πεδία επιλογών (dropdown menus). Αντίστοιχα και για τον αριθμό οδού.

  - φύλο
  - ΑΜΚΑ
  - ΑΦΜ
  - Αριθμός σταθερού τηλεφώνου
  - Αριθμός κινητού τηλεφώνου
  - Διεύθυνση ηλ. ταχυδρομείου
  - αριθμός οδού
  - οδός
  - ταχυδρομικός κώδικας
  - Διεύθυνση (*αναφορά στον αντίστοιχο πίνακα*)
  - Οικογενειακή κατάσταση (*αναφορά στον αντίστοιχο πίνακα*)
  - Μορφωτικό επίπεδο (*αναφορά στον αντίστοιχο πίνακα*)
  - Επάγγελμα (*αναφορά στον αντίστοιχο πίνακα*)

- Πίνακας "**διεύθυνση**":
  - υποκατάστατο κλειδί (*αύξων αριθμός*)
  - πόλη
  - χώρα

Με αυτόν τον τρόπο, συγκεντρώνουμε όλα τα προστατευμένα πεδία σε έναν πίνακα (στο συγκεκριμένο παράδειγμα), αφήνοντας ανέπαφες τις αναφορές προς άλλους πίνακες ώστε να μην εμποδίζουμε τον συνδυασμό πινάκων (*table joins*) και χωρίς να μειώνουμε το επίπεδο της ιδιωτικότητας των υποκειμένων επεξεργασίας.

Εφαρμογή στα αντίγραφα ασφαλείας

Ο παραπάνω σχεδιασμός δίνει εξ ίσου την δυνατότητα να εφαρμοστεί η ιδιωτικότητα των υποκειμένων επεξεργασίας στο επίπεδο των αντιγράφων ασφαλείας. Απο τον τρόπο με τον οποίο ταξινομούμε τα δεδομένα, βλέπουμε πως υπάρχουν τρία επίπεδα προστασίας των δεδομένων:

- τα δεδομένα συναλλαγών και τα κύρια μη-προσωπικά δεδομένα δεν χρήζουν προστασίας, γιατί είναι πλήρως ανωνυμοποιημένα

- τα κύρια δεδομένα που ανήκουν στα ταυτοποιητικά δεδομένα προστατεύονται από τα κλειδιά υποκειμένων, στα οποία δεν υπάρχει πρόσβαση από το σύστημα επεξεργασίας

- τα κύρια δεδομένα που ανήκουν στα αναγνωριστικά δεδομένα προστατεύονται από το κλειδί συστήματος

Όταν λοιπόν κάνουμε αντίγραφα ασφαλείας, παρατηρούμε τα εξής:

- για τα δεδομένα συναλλαγών και τα κύρια μη-προσωπικά δεδομένα, μπορούμε να κάνουμε αντίγραφα τόσο τύπου καθολικά (*full backups*) όσο και σταδιακά (*incremental backups*), σε μη προστατευμένη μορφή

- για τα ταυτοποιητικά δεδομένα, δεν υπάρχει η δυνατότητα να γίνουν σταδιακά αντίγραφα, καθώς τα αντίγραφα θα αναφέρονται σε κλειδιά προστασίας που είναι ενεργά. Τα αντίγραφα θα πρέπει να λαμβάνονται (και να επαναφέρονται) με καθολικό τρόπο, καθώς για τα συγκεκριμένα δεδομένα δεν υπάρχει πρόσβαση στο περιεχόμενο (λόγω προστασίας). Τα αντίγραφα θα είναι από μόνα τους προστατευμένα

- για τα αναγνωριστικά δεδομένα μπορούν να γίνουν αντίγραφα ασφαλείας, τόσο καθολικά όσο και σταδιακά, καθώς υπάρχει η δυνατότητα να εξεταστεί το περιεχόμενο (δεδομένου ότι το κλειδί συστήματος είναι διαθέσιμο). Βεβαίως, ο πιο αποδοτικός τρόπος θα ήταν να γίνονται καθολικά αντίγραφα, αλλά το κλειδί συστήματος που τα προστατεύει να αποθηκεύεται ξεχωριστά - όπως προτείνουν οι κανόνες διαχείρισης κλειδιών

Συνεπώς, έχουμε ως γενική οδηγία πως τα δεδομένα ενδιαφέροντος σκοπού μπορούμε να τα αποθηκεύουμε με σταδιακά αντίγραφα (*incremental*) ενώ τα υπόλοιπα μόνο ως καθολικά (*full backups*). Αυτό προκύπτει όχι μόνο λόγω του διαφορετικού επιπέδου προστασίας των δύο τύπων δεδομένων, αλλά κυρίως λόγω της διαφοράς του όγκου των δεδομένων. Αναμένουμε δηλαδή ότι ο όγκος των δεδομένων ενδιαφέροντος σκοπού θα είναι πολλαπλάσιος απο τα υπόλοιπα - καθώς για κάθε υποκείμενο επεξεργασίας περιμένουμε να συλλέγουμε πολλές πληροφορίες, αναλόγως με τις δυνατότητες του συστήματος επεξεργασίας.

9 *Συμπεράσματα*

Εισαγωγή

Στην παρούσα διπλωματική εξετάσαμε την σημασία της ιδιωτικότητας στα σύγχρονα πληροφοριακά συστήματα. Είδαμε τι είναι η ιδιωτικότητα και πως θεμελιώνεται νομικά και κοινωνικά, ποιές είναι οι κύριες παραβιάσεις της, πως θεμελιώνεται η έννοια της ιδιωτικότητας δια του σχεδιασμού και εξετάσαμε κάποιες προσπάθειες και μελέτες που έχουν γίνει στον συγκεκριμένο τομέα. Έπειτα, κάναμε μια σύντομη υπενθύμιση της διαδικασίας κανονικοποίησης μιας σχεσιακής βάσης δεδομένων, μελετήσαμε ένα μοντέλο απειλών για μια σχεσιακή βάση δεδομένων και, τέλος, προτείναμε την διαδικασία της κανονικοποίησης με υλοποίηση ιδιωτικότητας, η οποία ως στόχο έχει την επίλυση κάποιων θεμελιωδών προβλημάτων που παρουσιάζονται στα συστήματα επεξεργασίας δεδομένων προσωπικού χαρακτήρα.

Απο την μέχρι τώρα μελέτη προέκυψε πως, η διαδικασία σχεδιασμού της βάσης που προτείνεται, δεν αλλάζει σε μεγάλο βαθμό την κλασική κανονικοποίηση της βάσης, καθώς προσθέτει μόνο 5 απλά βήματα στην διαδικασία. Τα βήματα αυτά αφορούν περισσότερο στον χαρακτηρισμό και την ανάθεση προτεραιότητας στην προστασία των δεδομένων, παρά στο οτι τροποποιούν την διαδικασία της κανονικοποίησης. Για την ακρίβεια, η μόνη αλλαγή που εισάγεται είναι η μερική αποκανονικοποίηση της βάσης για λόγους απόδοσης - κάτι που ούτως ή άλλως χρησιμοποιείται ευρέως σε συστήματα αναλυτικών ερωτημάτων.

Σύγκριση με τα 7 κριτήρια της Ann Cavoukian

Για να εκτιμήσουμε την συνδρομή της συγκεκριμένης διαδικασίας στην προστασία της ιδιωτικότητας, θα προσπαθήσουμε να κατανοήσουμε πως η προτεινόμενη προσέγγιση πληροί τα κριτήρια ιδιωτικότητας δια του σχεδιασμού της Αν Καβούκιαν ([11]):

Προληπτικά όχι αντιδραστικά - αποτρεπτικά, όχι θεραπευτικά: η προστασία εφαρμόζεται στο επίπεδο αποθήκευσης των δεδομένων και όχι στις απαντήσεις των ερωτημάτων, άρα λειτουργεί προληπτικά και αποτρεπτικά

Ιδιωτικότητα ως η προκαθορισμένη ρύθμιση: τα ταυτοποιητικά δεδομένα προστατεύονται από το κλειδί του υποκειμένου και τα αναγνωριστικά απο το κλειδί συστήματος. Για να προσπελαστούν τα δεδομένα χρειάζεται να δοθεί πρόσβαση στα αντίστοιχα κλειδιά - συνεπώς η προκαθορισμένη ρύθμιση είναι η προστασία της ιδιωτικότητας και μόνο με θετική ενέργεια (opting in) μειώνεται το επίπεδό της

Ιδιωτικότητα ενσωματωμένη στον σχεδιασμό: η αρχή αυτή πληρούται προφανώς, γιατί η παρούσα προσέγγιση αναφέρεται στον σχεδιασμό της βάσης κι όχι σε μέτρα που θα πρέπει να εφαρμοστούν εκ των υστέρων

Πλήρης λειτουργικότητα (θετικό ισοζύγιο, όχι μηδενικό): ο τρόπος που προσεγγίζεται ο διαχωρισμός των δεδομένων και η ταξινόμησή τους στις ταυτοποιητικές κατηγορίες, πετυχαίνει την ελαχιστοποίηση του όγκου των προστατευμένων δεδομένων και την μεγιστοποίηση των άνευ προστασίας, προσπελάσιμων δεδομένων. Αυτό παρέχει την καλύτερη δυνατή εφαρμογή της ιδιωτικότητας και ταυτόχρονα την μέγιστη δυνατότητα επεξεργασίας των δεδομένων χωρίς υψηλό υπολογιστικό κόστος

Ολοκληρωμένη Ασφάλεια - Προστασία σε όλο τον κύκλο ζωής: η αρχή της προστασίας κατά την εγγραφή εξασφαλίζει πως θα υπάρχει προστασία σε όλο τον κύκλο ζωής των δεδομένων

Διαφάνεια: ο τρόπος που υπολογίζονται οι πληθικότητες και προσδιορίζεται η ταξινόμηση των δεδομένων στις ταυτοποιητικές κατηγορίες είναι αρκετά τεχνικός κι αφήνει μικρά περιθώρια σφάλματος. Έτσι, εξασφαλίζεται η δυνατότητα να τεκμηριώνεται ο σχεδιασμός της βάσης όσον αφορά την ιδιωτικότητα, πράγμα που συνδράμει στην διαφάνεια του συστήματος

Σεβασμός στην ιδιωτικότητα των χρηστών: το γεγονός ότι τα ταυτοποιητικά δεδομένα κάθε υποκειμένου επεξεργασίας προστατεύονται από το κλειδί του υποκειμένου, στο οποίο πρόσβαση θα πρέπει να έχει αποκλειστικά το κάθε υποκείμενο, σημαίνει πως δεν υπάρχει ευκολία στην παραβίαση της ιδιωτικότητας και πως γίνεται απολύτως σεβαστή η ιδιωτικότητα των υποκειμένων

Βλέπουμε πως η προτεινόμενη προσέγγιση πληροί τα κριτήρια της ιδιωτικότητας δια του σχεδιασμού της Αν Καβούκιαν με τρόπο σαφή και ευνόητο. Κάτι που της δίνει εγκυρότητα και θέτει τις βάσεις για περαιτέρω μελέτη και ανάπτυξή της.

Βιβλιογραφία

[1] "**Cisco Annual Security Report (2014)**", https://www.cisco.com/c/dam/assets/global/UK/pdfs/executive_security/sc-01_casr2014_cte_liq_en.pdf (τελευταία επίσκεψη Νοέ. 2021)

[2] G. Apruzzese, F. Pierazzi, M. Colajanni, M. Marchetti, "**Detection and Threat Prioritization of Pivoting Attacks in Large Networks**", *IEEE Transactions on Emerging Topics in Computing*, Oct. 2017, DOI: 10.1109/TETC.2017.2764885

[3] M. Husák, G. Apruzzese, S. J. Yang and G. Werner, "**Towards an Efficient Detection of Pivoting Activity**", *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2021, pp. 980-985

[4] D. R. Howe, "**Data Analysis for Database Design**", *Butterworth-Heinemann, 3η έκδοση*, 2001 (ISBN 978-0-750-65086-1)

[5] C. J. Date, "**SQL and relational theory**", *O'Reilly (2η έκδοση),* 2012 (ISBN 978-1-491-94117-1)

[6] E. F. Codd, "**A relational model of data for large shared data banks**", *Communication of the ACM - Information Retrieval, vol 13,* June 1970

[7] J. Ladley, "**Data governance**", *Morgan Kaufmann (Elsevier)*, 2012 (ISBN 978-0-12-415829-0)

[8] P. de Guise, "**Data protection: ensuring data availability**", *CRC Press*, 2017 (ISBN 978-1-4822-4415-1)

[9] D. Cervo, M. Allen, "**Master data management in practice**", *Wiley*, 2011 (ISBN 978-0-470-91055-9)

[10] E. Bertino, "**Data protection from insider threats**", *Morgan & Claypool*, 2012 (ISBN 978-1-608-45768-7)

[11] A. Cavoukian, "**Privacy by Design: the 7 foundational principles**", *Office of the Information and Privacy Commissioner of Ontario*, 2009, https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf, (ημερομηνία τελευταίας επίσκεψης: 29/12/2021)

[12] Ευρωπαϊκό Κοινοβούλιο, "**Οδηγία 95/46/ΕΚ του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου**", (Προστασία φυσικών προσώπων έναντι της επεξεργασίας δεδομένων προσωπικού χαρακτήρα και για την ελεύθερη κυκλοφορία των δεδομένων αυτών), 1995, https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:31995L0046&from=EN, (ημερομηνία τελευταίας επίσκεψης: 29/12/2021)

[13] Ευρωπαϊκό Κοινοβούλιο, "**Κανονισμός (ΕΕ) 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου**", (Γενικός Κανονισμός Προστασίας Δεδομένων - GDPR), 2016,

https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:32016R0679&from=EN, (ημερομηνία τελευταίας επίσκεψης: 29/12/2021)

[14] R. Agrawal, J. Kiernan, R. Srikant, Y. Xu, "**Hippocratic Databases**", *Proceedings of the 28th VLDB Conference, Hong Kong*, 2002

[15] M. Gertz, S. Jajodia (ed.), "**Handbook of Database Security, Applications and Trends**", *Springer*, 2008 (ISBN 978-0-387-48532-4)

[16] N. A. Ghani, Z. M. Sidek, "**Hippocratic Database: A Privacy-Aware Database**", *International Journal of Computer and Information Engineering*, 2008, Vol. 2, No. 6

[17] K. LeFevre, R. Agrawal, V. Ercegovac, R. Ramakrishnan, Y. Xu, D. DeWitt, "**Limiting Disclosure in Hippocratic Databases**", *Proceedings of the 30th VLDB Conference*, *Toronto, Canada*, 2004

[18] M. I. Khan, S. N. Foley, B. O'Sullivan, "**Quantitatively Measuring Privacy in Interactive Query Settings Within RDBMS Framework**", *Frontiers in Big Data*, 2020, Vol 3, DOI: 10.3389/fdata.2020.00011

[19] C. Dwork, F. McSherry, K. Nissim, A. Smith, "**Calibrating noise to sensitivity in private data analysis**", *In: Halevi, S., Rabin, T. (eds) Theory of Cryptography, vol 3876. Springer, Berlin, Heidelberg*, 2006, DOI: 10.1007/11681878_14

[20] C. Dwork, "**Differential Privacy**", *In: M. Bugliesi, B. Preneel, V. Sassone, I. Wegener (eds) ICALP 2006, LNCS, vol. 4052, pp. 1-12, Spring, Heidelberg*, 2006

[21] D. McGilvray, "**Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information**", *Morgan Kaufmann*, 2008 (ISBN: 978-0-123-74369-5)

[22] NIST, "**Digital Identity Guidelines: authentication and lifecycle management**", *Special Publication 800-63B*, 2017

[23] M. H. Murphy, "**Technological solutions to privacy questions: what is the role of law?**", *Information & Communications Technology Law, Vol. 25, No. 1*, 2016

[24] Θ' Αναθεωρητική Βουλή των Ελλήνων, "**Σύνταγμα**", 2019, (https://www.hellenicparliament.gr/Vouli-ton-Ellinon/To-Politevma/Syntagma, τελευταία πρόσβαση 06/2021)

[25] S. Trepte, L. Reinecke (editors), "**Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web**", *Springer*, 2011

[26] A. Tamo-Larrieux, "**Designing for Privacy and its Legal Framework: Data Protection by Design and Default for the Internet of Things**", *Law, Governance and Technology Series – Issues in Privacy and Data Protection, Vol. 40*, *Springer*, 2018

[27] D. George, K. Reutimann, A. Tamo-Larrieux, "**GDPR bypass by design? Transient processing of data under the GDPR**", *International Data Privacy Law, Vol. 9, No 4*, 2019

[28] W. Hartzog, "**Privacy's Blueprint: the battle to control the design of new technologies**", *Harvard University Press*, 2018

[29] Ευρωπαϊκό Σώμα για την Προστασία Προσωπικών Δεδομένων (European Data Protection Board), "**Data Protection by Design and by Default**", *Οδηγίες 4/2019 επι του Άρθρου 25 του ΓΚΠΔ (Guidelines 4/2019 on Article 25)*, 2019

[30] D. McAuley, A. Koene, J. Chen, "**Comments on the European Data Protection Board's Guidelines 4/2019 on Article 25 Data Protection by Design and by Default**", *Horizon Digital Economy Research Institute*, 16 Jan. 2020

[31] S. Gutwirth, R. Leenes, P. de Hert (editors), "**Reforming European Data Protection Law**",

Law, Governance and Technology Series – Issues in Privacy and Data Protection, Vol. 20, Springer, 2015

[32] D. Klitou, "**Privacy-Invading Technologies and Privacy by Design**", *Information Technology and Law Series, Vol. 25, Springer*, 2014

[33] D. Wright, P. de Hert (editors), "**Enforcing Privacy: Regulatory, Legal and Technological Approaches**", *Law, Governance and Technology Series, Vol. 25, Springer*, 2016

[34] Technical Committee ISO/PC 317, "**Consumer protection: privacy by design for consumer goods and services**", 2018 (https://www.iso.org/committee/6935430.html – τελευταία πρόσβαση 06/2021)

[35] European Court of Human Rights, Council of Europe, "**European Convention on Human Rights**", (όπως τροποποιήθηκε τελευταία με το Πρωτόκολλο 15), 2021 (https://www.echr.coe.int/Documents/Convention_ENG.pdf - τελευταία πρόσβαση 05/2022)