**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**

**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ**
**ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ**
**ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

# ΠΡΟΣΤΑΣΙΑ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΣΕ ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ

Γκάτσου Η. Ευγενία

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Μάιος 2013

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**

**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ**
**ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ**
**ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

# ΠΡΟΣΤΑΣΙΑ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΣΕ ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ

Γκάτσου Η. Ευγενία

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Μάιος 2013

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ 1η/03.10.2011 συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Πελέκης Νικόλαος, Λέκτορας (Επιβλέπων)
- Θεοδωρίδης Ιωάννης, Αναπλ. Καθηγητής
- Κοφίδης Ελευθέριος, Επίκ. Καθηγητής

Η έγκριση της Διπλωματική Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.

**U N I V E R S I T Y O F P I R A E U S**

**DEPARTMENT OF STATISTICS**

**AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN**

**APPLIED STATISTICS**

# PRIVACY PROTECTION IN MOBILITY DATA

By

Gkatsou I. Evgenia

MSc Dissertation

submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfilment of the requirements for the degree of Master of Science in Applied Statistics

Piraeus, Greece
May 2013

*Στους γονείς μου*
*Ηλία και Ελισάβετ*
*και στην αδερφή μου*
*Χρυσάνθη*

# Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω τον Λέκτορα κ. Πελέκη Νικόλαο που με εμπιστεύτηκε για να ασχοληθώ με το πολύ ενδιαφέρον αντικείμενο της εξόρυξης δεδομένων και πιο συγκεκριμένα της προστασίας ιδιωτικότητας. Καθώς και για τη συνεργασία που είχαμε όλο αυτό τον καιρό.

Επίσης, θα ήθελα να ευχαριστήσω την κ. Κοπανάκη Δέσποινα για τη βοήθειά της και την καθοδήγησή της στην εκπόνηση αυτής της διπλωματικής εργασίας.

Τέλος, ένα μεγάλο ευχαριστώ στους συμφοιτητές και φίλους για τη βοήθεια και την παρουσία τους όλα αυτά τα χρόνια, καθώς και την οικογένειά μου για την κατανόηση και την υποστήριξή.

# Περίληψη

Η διπλωματική εργασία ασχολείται με το αντικείμενο της προστασίας ιδιωτικότητας σε δεδομένα κίνησης. Εξετάζει μεθόδους που έχουν σαν στόχο τη προφύλαξη των ευαίσθητων πληροφοριών των χρηστών, προσφέροντας ένα είδος ανωνυμίας. Σαν παράδειγμα της αναγκαιότητας των μεθόδων αυτών μπορούμε να αναφέρουμε τους χρήστες των συστημάτων πλοήγησης (*GPS),* οι οποίοι προκειμένου να λάβουν τις πληροφορίες που χρειάζονται δημοσιοποιούν προσωπικά τους δεδομένα. Η παρούσα εργασία εστιάζει σε δύο κύριες θεματικές ενότητες, τη βιβλιογραφική ανασκόπηση και την ιδιωτικότητα βάσει σχεδιασμού *(Privacy by Design).*

Το πρώτο μέρος, περιλαμβάνει την εισαγωγή της διπλωματικής, στην οποία αναλύονται κάποιες πολύ βασικές έννοιες που είναι χρήσιμες για την εξέλιξη της εργασίας. Επίσης, γίνεται και μια παρουσίαση για το περιεχόμενο των επόμενων κεφαλαίων.

Στο δεύτερο μέρος, γίνεται εκτενής αναφορά στις τεχνικές που παρέχουν προστασία δεδομένων μέσω της βιβλιογραφικής ανασκόπησης. Το μοντέλο της *k*-ανωνυμίας *(k-anonymity)* είναι το πιο δημοφιλές και με βάση αυτό έχουν ακολουθήσει πολλές ακόμα προσεγγίσεις. Το μοντέλο αυτό, για την δημοσιοποίηση των δεδομένων απαιτεί το χωρισμό σε ισοδύναμες τάξεις μεγέθους τουλάχιστον *k*, όπου σε κάθε κλάση οι καταγραφές είναι δυσδιάκριτες ως προς τα ευαίσθητα χαρακτηριστικά. Στην προσπάθεια να ξεπεραστούν οι περιορισμοί της *k*-ανωνυμίας και να προσφερθεί μια πιο ισχυρή έννοια της ιδιωτικότητας, ακολούθησαν τα μοντέλα της *l*-πολυμορφίας *(l-diversity)* και της *t*-εγγύτητας *(t-closeness).* Γίνεται επίσης παρουσίαση βασικών τεχνικών προστασίας ιδιωτικότητας σε χωρικά και χωροχρονικά δεδομένα.

Στο τρίτο μέρος, αναφέρεται η γενική αρχή της ιδιωτικότητας βάσει σχεδιασμού *(Privacy By Design).* Αναλύονται οι τρεις υποθέσεις και οι επτά θεμελιώδεις αρχές που έχουν σαν στόχο την επίτευξη ισορροπίας μεταξύ της προστασίας και της χρησιμότητας των δεδομένων. Επιπλέον γίνεται έλεγχος για την τήρησή τους σε κάποιες από τις πιο σημαντικές τεχνικές ανωνυμίας.

Στο τέταρτο και τελευταίο μέρος, παρουσιάζονται τα συμπεράσματα από όλη την διπλωματική εργασία καθώς και οι προτάσεις για μελλοντικές εργασίες.

# Abstract

This thesis is focused on the subject of privacy protection in mobility data, examining methods that aim to protect sensitive user information by offering ways of anonymity. As an example of the necessity of these methods, we can mention GPS users, who in order to obtain the information they need, disclose their personal data. The present work focuses on two main themes, the literature review and Privacy by Design.

In the first part, the introduction of the thesis is presented. We analyze some basic concepts that are helpful for the progress of the work. There is also a presentation of the contents of the subsequent chapters.

The second part, focuses on techniques that provide data protection. The model of *k*-anonymity is the most popular technique, based on which many other approaches were "born". This model, in order to disclose data, requires the separation into equivalent classes sized at least *k*, where each class records are indistinguishable to sensitive attributes. In an attempt to overcome the limitations of *k*-anonymity and to provide a stronger sense of privacy, the models of *l*-diversity and *t*-closeness were introduced. In this part, basic privacy protection techniques in spatial and spatiotemporal data, are also presented.

The third part, is referred to the general principle of Privacy by Design. We analyze the three assumptions and the seven fundamental principles needed to be followed by all techniques, in order to achieve an optimal trade-off between privacy and data utility. In addition, we check whether the most significant privacy protection techniques comply with the rules.

In the fourth and final part, we present the conclusions of the entire thesis and our suggestions for future work.

# Περιεχόμενα

**ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ ..... XV**

**ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ ..... XVII**

**ΑΛΦΑΒΗΤΙΚΟΣ ΚΑΤΑΛΟΓΟΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ ..... XIX**

**ΚΕΦΑΛΑΙΟ 1 ..... 1**

**Εισαγωγή ..... 1**
1.1 Δεδομένα κίνησης ..... 1
1.2 Τροχιές δεδομένων ..... 1
1.3 Τροχιές δεδομένων με σημασιολογική πληροφορία ..... 2
1.4 Προστασία δεδομένων-Κίνδυνοι ..... 3
1.5 Το πρόβλημα της ανωνυμίας ..... 4
1.6 Κακόβουλοι χρήστες των Βάσεων Δεδομένων ..... 4
1.7 Τεχνικές ανωνυμίας ..... 4
1.8 Σκοπός της διπλωματικής εργασίας ..... 5

**ΚΕΦΑΛΑΙΟ 2 ..... 7**

**Βιβλιογραφική Ανασκόπηση ..... 7**
2.1 Δημοσιοποίηση Δεδομένων ..... 7
2.1.1 Η *k*-ανωνυμία ..... 8
2.1.2 Η *l*-πολυμορφία *(l-diversity)* ..... 10
2.1.3 Η *t*-εγγύτητα *(t-closeness)* ..... 11
2.2 Ανωνυμία σε Υπηρεσίες Εντοπισμού Θέσης *(Anonymity in Location-Based Services)* ..... 12
2.2.1 Ταξινόμηση των προσεγγίσεων *k*-ανωνυμίας ..... 14
2.2.1.1 Οι τεχνικές της απόκρυψης *(Cloaking Techniques)* ..... 15
2.2.1.2 Χωρική *k*-ανωνυμία *(Location k-anonymity)* ..... 16
2.2.1.3 Ιστορική *k*-ανωνυμία *(Historical k-anonymity)* ..... 21
2.2.1.4 *k*-ανωνυμία σε τροχιές κινούμενων αντικειμένων *(Trajectory k-anonymity)* ..... 21
2.3 Προστασία ιδιωτικότητας σε βάσεις δεδομένων τροχιών ..... 22
2.3.1 Προστασία Ιδιωτικότητας κατά τη Δημοσιοποίηση Δεδομένων Κίνησης *(Privacy-preserving mobility data publishing)* ..... 23
2.3.1.1 Προστασία Ιδιωτικότητας Τοποθεσίας μέσω Σύγχυσης Διαδρομής *(Protecting Location Privacy Through Path Confusion)* ..... 23
2.3.1.2 Προσεγγίσεις με βάση την ομαδοποίηση *(Clustering-based approaches)* ..... 24
2.3.1.3 Προσεγγίσεις με βάση τη γενίκευση *(Generalization-based approaches)* ..... 29
2.3.1.4 Προστασία δεδομένων κίνησης με απόκρυψη μονοπατιού *(Path Cloaking)* ..... 31
2.3.2 Προστασία Ιδιωτικότητας σε Επερωτήσεις Δεδομένων Κίνησης *(Privacy-Aware Mobility Data Querying)* ..... 31
2.3.2.1 Μέθοδοι ελέγχου ασφαλείας για στατιστικές βάσεις δεδομένων ..... 31
2.3.2.1.1 Τύποι στατιστικών δεδομένων και συστημάτων υπολογιστών ..... 32
2.3.2.2 Τεχνική βασισμένη στη μηχανή αναζήτησης *Hermes ++* ..... 33
2.4 Ανωνυμία σε σημασιολογικά ευαίσθητες τοποθεσίες ..... 35
2.4.1 *LBS* για σημασιολογικά ευαίσθητες τοποθεσίες ..... 35

2.4.1.1 Η τεχνική της απόκρυψης θέσεων με σημασιολογική πληροφορία *(Cloaking of semantic locations)* ... 35
2.4.1.2 Η επέκταση της τεχνικής απόκρυψης θέσεων με σημασιολογική πληροφορία *(The extension of cloaking of semantic locations)* ... 36
2.4.2 Προστασία ιδιωτικότητας σε τροχιές με σημασιολογική πληροφορία ... 37
2.4.2.1 Η τεχνική *C*-ασφάλεια *(C-safety)* ... 38

**ΚΕΦΑΛΑΙΟ 3 ... 41**

**Σχεδιασμένη Προστασία της Ιδιωτικότητας *(Privacy by Design)* ... 41**
3.1 Η ιδέα ... 41
3.2 Οι 7 θεμελιώδεις αρχές ... 42
3.3 Έλεγχος τήρησης των τριών υποθέσεων και των επτά θεμελιωδών αρχών της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας ... 43
3.3.1 Προστασία Ιδιωτικότητας σε Υπηρεσίες Εντοπισμού θέσης *(Privacy in Location Based Services)* . 43
3.3.2 Προστασία Ιδιωτικότητας κατά τη Δημοσιοποίηση Δεδομένων Κίνησης *(Privacy-preserving Mobility Data Publishing)* ... 46
3.3.3 Προστασία Ιδιωτικότητας σε Επερωτήσεις Δεδομένων Κίνησης *(Privacy-aware data querying)* .... 51
3.3.4 Έλεγχος τήρησης των τριών υποθέσεων και των επτά θεμελιωδών αρχών της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας για δεδομένα με σημασιολογική πληροφορία ..... 53
3.4 Πίνακες συνοπτικών αποτελεσμάτων ... 57

**ΚΕΦΑΛΑΙΟ 4 ... 61**

**Συμπεράσματα-Συζήτηση ... 61**
4.1 Συμπεράσματα ... 61
4.2 Παρεμβάσεις ... 62
4.3 Μελλοντικές επεκτάσεις ... 63

**ΒΙΒΛΙΟΓΡΑΦΙΑ ... 65**

# Κατάλογος Πινάκων

**2.1** Ιατρικά δεδομένα.....9
**2.2** 4-ανωνυμία στα ιατρικά δεδομένα.....9
**2.3** 3-πολυμορφία στα ιατρικά δεδομένα.....11
**3.1** Συνοπτικά αποτελέσματα ελέγχου τήρησης των τριών υποθέσεων της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας για δεδομένα κίνησης.....58
**3.2** Συνοπτικά αποτελέσματα ελέγχου τήρησης των 7 θεμελιωδών αρχών της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας για δεδομένα κίνησης.....58
**3.3** Συνοπτικά αποτελέσματα ελέγχου τήρησης των τριών υποθέσεων της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας για δεδομένα κίνησης με σημασιολογική πληροφορία.....59
**3.4** Συνοπτικά αποτελέσματα ελέγχου τήρησης των 7 θεμελιωδών αρχών της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας για δεδομένα κίνησης με σημασιολογική πληροφορία.....59

# Κατάλογος Σχημάτων

**1.1** Απεικόνιση ενός δείγματος τροχιάς και μιας τροχιάς με σημασιολογική πληροφορία *(Monreale et al., 2011)*.....2
**2.1** Το παράδειγμα της *Sweeney*-Συνδυασμός δύο διαφορετικών συνόλων δεδομένων.....7
**2.2** Το συγκεντρωτικό μοντέλο προστασίας ιδιωτικότητας σε *LBS*.....13
**2.3** Ένα παράδειγμα χωρικής *k*-ανωνυμίας.....14
**2.4** Απόκρυψη βασιζόμενη στα δεδομένα.....15
**2.5** Απόκρυψη βασιζόμενη στο χώρο.....16
**2.6** Ένα παράδειγμα της τεχνικής απόκρυψης κλίκας.....17
**2.7** Ένα παράδειγμα της τεχνικής της απόκρυψης του κοντινότερου γείτονα.....18
**2.8** Ένα παράδειγμα της τεχνικής του *Casper*.....19
**2.9** Ένα παράδειγμα της τεχνικής της απόκρυψης του *Hilbert* .....20
**2.10** *k*-ανωνυμία βασισμένη σε μετακινήσεις του παρελθόντος.....21
**2.11** Παράδειγμα διασταύρωσης διαδρομών μεταξύ δύο τροχιών.....24
**2.12** Περιοχή αβεβαιότητας: η τροχιά, ο όγκος τροχιάς και η πιθανή καμπύλη κίνησης.....25
**2.13** Σύνολο ανωνυμίας που σχηματίζεται από δύο συνεντοπισμένες τροχιές, τους αντίστοιχους όγκους αβεβαιότητας και τον κεντρικό κυλινδρικό όγκο της ακτίνας δ/2 που περιέχει και τις δύο τροχιές.....26
**2.14** Ομαδοποίηση τροχιών με χρήση της $1^{ης}$ στρατηγικής για την τεχνική της άπληστης ομαδοποίησης.....28
**2.15** Ομαδοποίηση τροχιών με χρήση της $3^{ης}$ στρατηγικής για την τεχνική της άπληστης ομαδοποίησης.....29
**2.16** Διαδικασία Ανωνυμίας της τεχνικής *AWO*.....29
**2.17** (α) περιορισμός ερωτημάτων, (β) διατάραξη δεδομένων, (γ) διατάραξη εξόδου.....32
**2.18** Ο αλγόριθμος παραγωγής ψεύτικης τροχιάς.....34
**2.19** (α) Πλέγμα τοποθεσίες με σημασιολογική πληροφορία στον πραγματικό κόσμο, (β) Πλέγμα με τις πιθανότητες τοποθεσιών με σημασιολογική πληροφορία, (γ) Πλέγμα με τις ευαίσθητες πληροφορίες.....36
**2.20** (α) επικαλυπτόμενη απόκρυψη *(overlapping cloaking)*, (β) ασυνεχής απόκρυψη *(disjoint cloaking)*.....37
**2.21** Παράδειγμα περιοχών ταξινόμησης.....38

# Αλφαβητικός Κατάλογος Συντομογραφιών

| | |
|---|---|
| ASR | Anonymity Spatial Region |
| AWO | Always Walk with Others |
| CRS | Cloaked Regions |
| EDR | Effective Detection Radius |
| GC | Greedy Clustering |
| GIC | Group Insurance Commission |
| GPS | Global Positioning System |
| LBS | Location Based Services |
| MBR | Minimum Bounding Rectangle |
| NN-Cloak | Nearest Neighbor Cloak |
| NWA | Never Walk Alone |
| PbD | Privacy by Design |
| POI | Place Of Interest |
| SCS | Shared Computer System |
| SDB | Statistical Data Base |
| W4M | Wait For Me |

# ΚΕΦΑΛΑΙΟ 1

## Εισαγωγή

### 1.1 Δεδομένα κίνησης

Στη σύγχρονη εποχή, η παραγωγή των δεδομένων γίνεται με πολύ γρήγορους ρυθμούς. Υπάρχουν δεδομένα που αφορούν επιστήμες όπως: ιατρικά, βιολογικά, γεωγραφικά, τραπεζικά, τηλεπικοινωνιακά και τόσα ακόμα που οι πηγές τους στους περισσότερους ανθρώπους είναι άγνωστες. Η αποθήκευσή τους ποικίλει, εφόσον υπάρχουν οικονομικά και γρήγορα μέσα αποθήκευσης, όπως επίσης και μεγάλες βάσεις που φυλάσσουν δεδομένα.

Πιο συγκεκριμένα, τα δεδομένα κίνησης μπορεί να είναι σημεία, γραμμές, περιοχές ή ποσότητες, που κινούνται στο πέρασμα του χρόνου. Ένας αυστηρός ορισμός της κίνησης είναι η αλλαγή της φυσικής θέσης. Η φυσική αυτή κυκλοφορία συνεπάγεται ένα αντικείμενο και ένα σύστημα στο οποίο γίνεται αναφορά των θέσεών του. Πιο συχνά, το σύστημα είναι ένας γεωγραφικός χώρος και μιλάμε για αντικείμενα που κινούνται στο χώρο.

### 1.2 Τροχιές δεδομένων

Μια τροχιά συνιστά την περιγραφή της κίνησης των κινούμενων αντικειμένων. Ορίζεται σαν μια συνάρτηση χρόνου και γεωγραφικού χώρου, είναι δηλαδή η καταγραφή των θέσεων κάποιου αντικειμένου σε συγκεκριμένες χρονικές στιγμές. Πιο συγκεκριμένα, έχουμε τον παρακάτω ορισμό:

❖ Ορισμός 1.1

Τροχιά δεδομένων. Μια τροχιά Τ είναι η γραφική παράσταση μιας συνεχούς απεικόνισης από το $I \subseteq R$ στο $R^2$ σε επίπεδο δύο διαστάσεων.

$$I \subseteq R \rightarrow R^2 : t \mapsto a(t) = (a_x(t), a_y(t))$$

Οπότε: $T = \{(a_x(t), a_y(t), t) | t \in I\} \subset R^2 \times R$

Εν συνεχεία, θα θέλαμε να τονίσουμε ότι κάθε τροχιά που συνδέεται με ένα δείγμα δεδομένων πρέπει να περιλαμβάνει τα στοιχεία του δείγματος. Για όλα τα σημεία $(x_i, y_i, t_i)$, στο δείγμα έχουμε $(x_i, y_i, t_i) = (a_x(t_i), a_y(t_i), t_i)$.

### 1.3 Τροχιές δεδομένων με σημασιολογική πληροφορία

Τροχιές με σημασιολογική πληροφορία *(Semantic trajectories)* είναι ακολουθίες στάσεων *(stops)* και κινήσεων *(moves)* ενός ατόμου κατά τη διάρκεια της κίνησής του. Κάθε τοποθεσία της στάσης του μπορεί να συνδεθεί με ορισμένες σημασιολογίες πληροφορίες, είτε από σαφείς ανιχνεύσεις είτε συμπερασματικά (σχήμα 1.1). Στάσεις είναι τα σημαντικά μέρη μιας τροχιάς όπου το κινούμενο αντικείμενο διέμεινε για ένα ελάχιστο χρονικό διάστημα. Κινήσεις είναι οι υποτροχιές που περιγράφουν τις μετακινήσεις μεταξύ δύο διαδοχικών στάσεων. Πιο συγκεκριμένα:

❖ Ορισμός 1.2

Τροχιά δεδομένων με σημασιολογική πληροφορία. Αν έχουμε ένα σύνολο $I$ από σημαντικές τοποθεσίες, μία τροχιά με σημασιολογική πληροφορία $T = p_1, p_2, \ldots\ldots, p_n$ με $p_i \in I$, είναι μια χρονικά διατεταγμένη ακολουθία από σημαντικές τοποθεσίες που έχει επισκεφτεί ένα κινούμενο αντικείμενο.

**Σχήμα 1.1**

Απεικόνιση ενός δείγματος τροχιάς και μιας τροχιάς με σημασιολογική πληροφορία *(Monreale et al., 2011)*

### 1.4 Προστασία δεδομένων-Κίνδυνοι

Οι συλλογές δεδομένων κίνησης, συχνά περιέχουν πληροφορίες που δεν πρέπει να δημοσιευτούν. Οι χειριστές των βάσεων δεδομένων, έχουν την ευθύνη να δημοσιεύουν πληροφορίες χωρίς να θέτουν σε κίνδυνο την προστασία της ιδιωτικότητας, την εμπιστευτικότητα ή ακόμα και γνώσεις εθνικών συμφερόντων. Λειτουργώντας αυτόνομα και συχνά χωρίς ιδιαίτερες γνώσεις, είναι δύσκολο να το καταφέρουν.

Σε πολλές περιπτώσεις, η ύπαρξη της προστασίας της βάσης δεδομένων εξαρτάται από την ικανότητα των χειριστών τους να δημοσιοποιούν ανώνυμα δεδομένα, καθώς οφείλουν να παρέχουν τέτοιο όγκο πληροφοριών που να διατηρεί τη χρησιμότητά τους. Σε αντίθετη περίπτωση, μια αποτυχία προστασίας των δεδομένων μπορεί να επιφέρει συνθήκες που βλάπτουν το κοινό.

Σε αυτό το σημείο θα θέλαμε να εξηγήσουμε τις έννοιες του ευαίσθητου χαρακτηριστικού *(sensitive attribute)* και του οιονεί-αναγνωριστικού *(quasi-identifier)*. Αρχικά, ένα χαρακτηριστικό επισημαίνεται ως ευαίσθητο όταν το άτομο στο οποίο ανήκει δεν επιθυμεί τη δημοσιοποίησή του. Χαρακτηριστικά που δεν σημειώνονται ως ευαίσθητα, θεωρούνται μη-ευαίσθητα. Στη συνέχεια έχουμε τον παρακάτω ορισμό:

- Ορισμός 1.3

  Οιονεί-αναγνωριστικό *(Quasi-identifier)*. Το σύνολο μη ευαίσθητων χαρακτηριστικών ($Q_1$, ….. ,$Q_w$ ) ενός πίνακα, καλείται οιονεί-αναγνωριστικό, αν αυτά τα χαρακτηριστικά μπορούν να συνδεθούν με εξωτερικά δεδομένα και να προσδιορίσουν μοναδικά, τουλάχιστον ένα άτομο από έναν γενικευμένο πληθυσμό Ω.

Όσον αφορά την προστασία των δεδομένων υπάρχουν δύο ειδών κίνδυνοι. Πρώτον, υπάρχει η περίπτωση κάποιος κακόβουλος χρήστης των βάσεων δεδομένων να κάνει πολλές παρατηρήσεις πάνω στα ίδια ανώνυμα δεδομένα και να εξάγει πολλές πληροφορίες. Δεύτερον, μπορεί να πάρει πληροφορίες από διαφορετικές βάσεις δεδομένων και να τις συνδέσει με τέτοιο τρόπο ώστε να εξάγει πολλά χρήσιμα στοιχεία.

### 1.5 Το πρόβλημα της ανωνυμίας

Οι χειριστές των βάσεων δεδομένων, για να προστατέψουν τα σύνολα δεδομένων από τους κακόβουλους χρήστες, πρέπει να τα μετατρέψουν έτσι ώστε κανείς να μην μπορεί:

1. να συνδέσει μια συγκεκριμένη εγγραφή με τα δεδομένα της
2. να συμπεράνει εμπιστευτικές πληροφορίες από κάποιο σύνολο δεδομένων

Το πιο ουσιαστικό πρόβλημα της διαδικασίας της ανωνυμίας, είναι ότι ο μετασχηματισμός των δεδομένων πρέπει να είναι όσο πιο περιορισμένος γίνεται ώστε να διατηρήσει όσο το δυνατόν περισσότερες πληροφορίες. Στόχος είναι η ελαχιστοποίηση της στρέβλωσης των δεδομένων που δημοσιοποιούν.

### 1.6 Κακόβουλοι χρήστες των Βάσεων Δεδομένων

Κακόβουλοι χρήστες θεωρούνται εκείνοι που θέλουν να αποκομίσουν πληροφορίες από τις βάσεις δεδομένων για να βλάψουν την ιδιωτικότητα κάποιου ατόμου και είναι ουσιαστικά η αιτία για την οποία έχουν δημιουργηθεί οι τεχνικές ανωνυμίας. Οι προσπάθειες που κάνουν για να συλλέξουν όσες ευαίσθητες πληροφορίες μπορούν, δημιουργούν όλο και μεγαλύτερη ανάγκη ώστε οι τεχνικές να καλύπτουν όλους τους πιθανούς τρόπους διαρροής σημαντικών πληροφοριών.

### 1.7 Τεχνικές ανωνυμίας

Όσον αφορά την ανωνυμία δεδομένων κίνησης, υπάρχουν πολλές τεχνικές με σκοπό την προστασία της ιδιωτικότητας. Στην παρούσα διπλωματική εργασία, γίνεται ανάλυση μεγάλου μέρος των τεχνικών αυτών. Η ανάλυση ξεκινάει με την αναφορά στις πρώτες τεχνικές ανωνυμίας, οι οποίες δημιουργήθηκαν με στόχο την παύση της διαρροής προσωπικών πληροφοριών από σύνολα δεδομένων που είναι προς δημοσίευση. Η ανάγκη για να καλυφθούν όσες περισσότερες πλευρές της ιδιωτικότητας είναι δυνατόν, οδήγησε στη δημιουργία νέων τεχνικών ανωνυμίας. Στα κεφάλαια που ακολουθούν, οι τεχνικές έχουν διαχωριστεί στις παρακάτω κατηγορίες:

i. τεχνικές που ασχολούνται με χωρικά δεδομένα

ii. τεχνικές που ασχολούνται με δεδομένα τροχιάς

iii. τεχνικές που ασχολούνται με χωρικά δεδομένα με σημασιολογική πληροφορία

iv. τεχνικές που ασχολούνται με δεδομένα τροχιάς με σημασιολογική πληροφορία

Όσον αφορά την πρώτη και την τρίτη ομάδα τεχνικών, αναλύσαμε την Ανωνυμία σε Υπηρεσίες Εντοπισμού Θέσης *(Location Based Services)*. Οι υπηρεσίες εντοπισμού θέσης είναι μια γενική κατηγορία ηλεκτρονικών υπηρεσιών που χρησιμοποιούνται για να συμπεριλάβουν ειδικούς ελέγχους για την τοποθεσία και το χρόνο των δεδομένων. Έχουν ποικίλες χρήσεις στην κοινωνική δικτύωση και είναι προσιτές με συσκευές κινητών τηλεφώνων μέσω των δικτύων τηλεπικοινωνιών. Στη δεύτερη και τέταρτη ομάδα τεχνικών ανωνυμίας δίνουμε έμφαση σε δύο πτυχές: Στην Προστασία Ιδιωτικότητας κατά την Δημοσιοποίηση Δεδομένων Κίνησης και στην Προστασία Ιδιωτικότητας σε Επερωτήσεις Δεδομένων Κίνησης. Η πρώτη πτυχή, αφορά τεχνικές που δημοσιοποιούν ανώνυμα σύνολα δεδομένων με κύριο μέλημά τους τη μέγιστη δυνατή χρησιμότητα των δεδομένων. Η δεύτερη πτυχή αφορά τεχνικές που έχουν σαν βασική αρχή τους να κρατούν τα δεδομένα εντός του οργανισμού που τα φιλοξενεί *(in-house)* και να έχουν μεγαλύτερο έλεγχο στις πληροφορίες που δημοσιοποιούν.

Η εκτενής ανάλυση όσον αφορά τις παραπάνω τεχνικές, έχει σαν στόχο τον σαφή διαχωρισμό των κατηγοριών τους, την παρουσίαση των βασικών τους στοιχείων καθώς και τον εντοπισμό των «δυνατών» και «τρωτών» τους σημείων, σε περίπτωση που υπάρχουν.

Οι τεχνικές ανωνυμίας, θα μας απασχολήσουν και από μία άλλη οπτική. Θα ερευνήσουμε κατά πόσο ακολουθούν τις υποθέσεις και τις αρχές της Σχεδιασμένης Προστασίας της Ιδιωτικότητας *(Privacy by Design)*, έναν όρο που δίνει έμφαση στον τρόπο με τον οποίο οι τεχνικές προσαρτούν προσωπικές πληροφορίες στο σχεδιασμό τους. Με τη διαδικασία αυτή θα μπορέσουμε να ελέγξουμε ποιες τεχνικές έχουν τη δυνατότητα να παρέχουν υψηλή προστασία της ιδιωτικότητας και ταυτόχρονα καλή ποιότητα δεδομένων.

### 1.8 Σκοπός της διπλωματικής εργασίας

Το αντικείμενο της προστασίας ιδιωτικότητας δεδομένων κίνησης είναι ιδιαίτερα απαιτητικό. Τα δεδομένα που καλούνται να προστατέψουν οι τεχνικές ανωνυμίας συνήθως αφορούν προσωπικές πληροφορίες ανθρώπων, συνεπώς οφείλουν να παρέχουν προστασία υψηλού επιπέδου.

Σκοπός της παρούσας εργασίας, είναι αρχικά η μελέτη των περισσότερων τεχνικών ανωνυμίας που ασχολούνται με το αντικείμενο της προστασίας ιδιωτικότητας σε δεδομένα κίνησης. Στα επόμενα κεφάλαια, καταγράφονται τα πιο σημαντικά χαρακτηριστικά της κάθε τεχνικής. Αναλύεται το είδος δεδομένων στο οποίο η κάθε μία απευθύνεται, οι επιθέσεις που μπορεί να αντιμετωπίσει καθώς και η διαδικασία που ακολουθεί για να προσφέρει ανωνυμία.

Στη συνέχεια, ελέγχεται αν οι τεχνικές ανωνυμίας πληρούν τις προϋποθέσεις προκειμένου να είναι ποιοτικές ως προς εκείνους που χρησιμοποιούν τα δεδομένα αλλά και αποτελεσματικές ως προς την παροχή προστασίας. Για το σκοπό αυτό χρησιμοποιήσαμε τον όρο της Σχεδιασμένης Προστασίας της Ιδιωτικότητας. Με αυτή τη διαδικασία αποκτήσαμε περισσότερες γνώσεις για τις δυνατότητες των τεχνικών ανωνυμίας, τις αξιολογήσαμε και καταλήξαμε στα συμπεράσματα που παραθέτουμε στο τελευταίο κεφάλαιο.

# ΚΕΦΑΛΑΙΟ 2

## Βιβλιογραφική Ανασκόπηση

### 2.1 Δημοσιοποίηση Δεδομένων

Η διαδικασία που χρειάζεται για να γίνουν τα δεδομένα ανώνυμα δεν είναι καθόλου εύκολη. Είναι αρκετά πιθανό να επιτευχθεί αναγνώριση ταυτότητας δεδομένων, ακόμα και αν έχουν δημοσιευτεί χωρίς προσωπικές πληροφορίες των εγγεγραμμένων. Σαν παράδειγμα έχουμε την κλινική εγγραφή του Κυβερνήτη της Μασαχουσέτης *William Weld (Sweeney, 2002).*

Ο *William Weld* λοιπόν, ο οποίος ζούσε στο *Cambridge* της Μασαχουσέτης εκείνη την εποχή, είχε τον ιατρικό του φάκελο στα δεδομένα της *GIC*. Σύμφωνα με τη λίστα των ψηφοφόρων του *Cambridge*, έξι άτομα είχαν την ίδια ημερομηνία γέννησης με αυτόν, εκ των οποίων μόνο τρεις ήταν άνδρες και αυτός ήταν ο μοναδικός με 5-ψήφιο ταχυδρομικό κώδικα (σχήμα 2.1).

**Σχήμα 2.1**

Το παράδειγμα της *Sweeney*-Συνδυασμός δύο διαφορετικών συνόλων δεδομένων

Επειδή η *GIC* θεώρησε ότι τα δεδομένα που είχε στην κατοχή της ήταν ανώνυμα, έδωσε αντίγραφα των δεδομένων σε ερευνητές. Επιπλέον, πρέπει να επισημανθεί ότι ήταν πολύ εύκολο για κάποιον ερευνητή, μόλις με 20 δολάρια, να πάρει τον κατάλογο εγγραφής των ψηφοφόρων και τελικά να εξάγει το αποτέλεσμα για την κατάσταση του Κυβερνήτη.

### 2.1.1 Η *k*-ανωνυμία

Η *Latanya Sweeney* (2002), χρησιμοποίησε το μοντέλο προστασίας της *k*-ανωνυμίας προσπαθώντας να σταματήσει τη διαρροή των προσωπικών στοιχείων από τα δεδομένα που δημοσιοποιούνται.

Αρχικά, ξεκινάμε με την υπόθεση ότι ο κάτοχος των δεδομένων αναμένεται να προσδιορίσει όλα τα χαρακτηριστικά των προσωπικών πληροφοριών που θα μπορούσαν να χρησιμοποιηθούν για εξαγωγή περισσότερων συμπερασμάτων. Τα χαρακτηριστικά αυτά δεν περιλαμβάνουν μόνο σαφή αναγνωριστικά στοιχεία, όπως για παράδειγμα το όνομα, η διεύθυνση και το τηλέφωνο αλλά και χαρακτηριστικά όπως η ημερομηνία γέννησης και το φύλο, τα οποία σε συνδυασμό μπορεί να προσδιορίζουν με μοναδικό τρόπο τα άτομα. Το σύνολο αυτών των χαρακτηριστικών είναι τα οιονεί-αναγνωριστικά.

❖ Ορισμός 2.1

*k*-ανωνυμία *(k-anonymity).* Ας είναι ο RT ($A_1$, ..., $A_n$) ένας πίνακας, και το $QI_{RT}$ ένα οιονεί-αναγνωριστικό που συνδέεται με αυτόν. Ο RT λέγεται ότι ικανοποιεί την *k*-ανωνυμία, αν και μόνο αν, κάθε ακολουθία των τιμών στο RT [$QI_{RT}$] εμφανίζεται με τουλάχιστον *k* εγγραφές.

Για να καταστούν σαφείς οι έννοιες της *k*-ανωνυμίας και του οιονεί-αναγνωριστικού ακολουθούν οι πίνακες 2.1 και 2.2.

**Πίνακας 2.1**

Ιατρικά δεδομένα

| | Μη-ευαίσθητα | | | Ευαίσθητα |
|---|---|---|---|---|
| | ΤΑΧ. ΚΩΔΙΚΑΣ | ΗΛΙΚΙΑ | ΕΘΝΙΚΟΤΗΤΑ | ΠΡΟΒΛΗΜΑ ΥΓΕΙΑΣ |
| 1 | 13053 | 28 | Ρώσικη | Καρδιαγγειακό |
| 2 | 13068 | 29 | Αμερικάνικη | Καρδιαγγειακό |
| 3 | 13068 | 21 | Ιαπωνική | Ηπατίτιδα |
| 4 | 13053 | 23 | Αμερικάνικη | Ηπατίτιδα |
| 5 | 14853 | 50 | Ινδική | Άσθμα |
| 6 | 14853 | 55 | Ρώσικη | Καρδιαγγειακό |
| 7 | 14850 | 47 | Αμερικανική | Μόλυνση από ιό |
| 8 | 14850 | 49 | Αμερικανική | Μόλυνση από ιό |
| 9 | 13053 | 37 | Ινδική | Άσθμα |
| 10 | 13053 | 36 | Ιαπωνική | Άσθμα |
| 11 | 13068 | 35 | Αμερικάνικη | Άσθμα |
| 12 | 13068 | 35 | Αμερικάνικη | Άσθμα |

**Πίνακας 2.2**

4-ανωνυμία στα ιατρικά δεδομένα

| | Μη-ευαίσθητα | | | Ευαίσθητα |
|---|---|---|---|---|
| | ΤΑΧ. ΚΩΔΙΚΑΣ | ΗΛΙΚΙΑ | ΕΘΝΙΚΟΤΗΤΑ | ΠΡΟΒΛΗΜΑ ΥΓΕΙΑΣ |
| 1 | 130** | <30 | * | Καρδιαγγειακό |
| 2 | 130** | <30 | * | Καρδιαγγειακό |
| 3 | 130** | <30 | * | Ηπατίτιδα |
| 4 | 130** | <30 | * | Ηπατίτιδα |
| 5 | 148** | ≥40 | * | Άσθμα |
| 6 | 148** | ≥40 | * | Καρδιαγγειακό |
| 7 | 148** | ≥40 | * | Μόλυνση από ιό |
| 8 | 148** | ≥40 | * | Μόλυνση από ιό |
| 9 | 130** | 3* | * | Άσθμα |
| 10 | 130** | 3* | * | Άσθμα |
| 11 | 130** | 3* | * | Άσθμα |
| 12 | 130** | 3* | * | Άσθμα |

Ο πίνακας 2.1 μας δείχνει ιατρικά δεδομένα από ένα νοσοκομείο. Τα χαρακτηριστικά διαιρούνται σε δύο ομάδες: στα ευαίσθητα χαρακτηριστικά, στα οποία ανήκει μόνο το πρόβλημα υγείας και στα μη-ευαίσθητα, που είναι ο ταχυδρομικός κώδικας, η ηλικία και η εθνικότητα. Επιπλέον, το σύνολο των χαρακτηριστικών {ταχυδρομικός κώδικας, ηλικία, εθνικότητα} είναι τα οιονεί-αναγνωριστικά για αυτό το σύνολο δεδομένων.

Ο πίνακας 2.2 δείχνει έναν 4-ανώνυμο πίνακα που προέρχεται από τον πίνακα 2.1 (εδώ το "*" υποδηλώνει μία τιμή που δε θέλουμε να αποκαλυφθεί, έτσι για παράδειγμα αν έχουμε

ταχυδρομικός κώδικας = 1485* σημαίνει ότι ο ταχυδρομικός κώδικας είναι στην περιοχή [14850-14859] και αν έχουμε ηλικία = 3* σημαίνει ότι η ηλικία είναι στην περιοχή [30 έως 39]). Σημειώνουμε ότι στον 4-ανώνυμο πίνακα, κάθε πλειάδα έχει τις ίδιες τιμές για το οιονεί-αναγνωριστικό με τουλάχιστον δύο άλλες πλειάδες του πίνακα.

### 2.1.2 Η *l*-πολυμορφία *(l-diversity)*

Η *k*-ανωνυμία δέχεται δύο είδη επιθέσεων: την επίθεση ομοιογένειας *(homogeneity attack)* και την επίθεση γνωστικού υπόβαθρου *(background knowledge attack)*. Παρακάτω τις εξηγούμε με παραδείγματα.

- Επίθεση ομοιογένειας. Ας υποθέσουμε ότι δύο γείτονες Α και Β έχουν μεταξύ τους κάποιες διαφορές. Ο Α αρρωσταίνει και πηγαίνει με ασθενοφόρο στο νοσοκομείο. Ο Β θέλει να μάθει τι συνέβη στο γείτονά του. Ανακαλύπτει λοιπόν τον 4-ανώνυμο πίνακα 2.2. Γνωρίζοντας την ηλικία του (35 χρονών), την περιοχή που κατοικεί (ταχυδρομικός κώδικας 13068) αλλά και την εθνικότητά του (Αμερικανός), συμπεραίνει ότι είναι ο ασθενής 7, 8 ή 9. Επειδή λοιπόν όλοι αυτοί οι ασθενείς έχουν το ίδιο πρόβλημα υγείας, συμπεραίνει ότι έχει άσθμα.
- Επίθεση γνωστικού υπόβαθρου. Ας υποθέσουμε ότι ο Β έχει ένα γνωστό Γ, ο οποίος τυχαίνει να είναι επίσης σε αυτόν τον πίνακα καταγεγραμμένος καθώς νοσηλεύεται στο ίδιο νοσοκομείο. Για εκείνον γνωρίζει την ηλικία του (21 χρονών), την περιοχή που κατοικεί (ταχυδρομικός κώδικας 13053) και την εθνικότητά του (Ιαπωνική). Συμπεραίνει λοιπόν ότι η πληροφορία που θέλει είναι μεταξύ των ασθενών 1-3. Όμως επειδή είναι γνωστό πως οι Ιάπωνες σπάνια παθαίνουν προβλήματα με την καρδιά τους, καταλαβαίνει ότι πάσχει από ηπατίτιδα.

Για να ξεπεραστούν οι περιορισμοί της *k*-ανωνυμίας, η *l*-πολυμορφία εισήχθη ως μια ισχυρότερη έννοια διασφάλισης της ιδιωτικότητας. Η *l*-πολυμορφία προσφέρει προστασία της ιδιωτικότητας ακόμα και όταν αυτός που δημοσιεύει τα δεδομένα δεν ξέρει τι είδους γνώση έχει ο αντίπαλος. Η κύρια ιδέα, είναι η απαίτηση ότι οι τιμές των ευαίσθητων χαρακτηριστικών σε κάθε ομάδα παρουσιάζονται με τέτοιο τρόπο ώστε να μην αφήνουν περιθώρια ανακάλυψής τους.

❖ Αρχή 2.1

Η αρχή της *l*-πολυμορφίας *(The l-diversity principle).* Μια τάξη ισοδυναμίας λέγεται ότι έχει *l*-πολυμορφία αν υπάρχουν *l* τουλάχιστον «καλά παρουσιασμένες» τιμές για το ευαίσθητο χαρακτηριστικό. Ένας πίνακας λέγεται ότι έχει *l*-πολυμορφία αν κάθε τάξη ισοδυναμίας του πίνακα έχει *l*-πολυμορφία (*Machanavajjhala et al.*, 2006).

Παρακάτω δίνεται ένας πίνακας για την πλήρη κατανόηση της αρχής της *l*-πολυμορφίας.

**Πίνακας 2.3**

3-πολυμορφία στα ιατρικά δεδομένα

| | **Μη-ευαίσθητα** | | | **Ευαίσθητα** |
|---|---|---|---|---|
| | **ΤΑΧ. ΚΩΔΙΚΑΣ** | **ΗΛΙΚΙΑ** | **ΕΘΝΙΚΟΤΗΤΑ** | **ΠΡΟΒΛΗΜΑ ΥΓΕΙΑΣ** |
| 1 | 1305* | <40 | * | Καρδιαγγειακό |
| 4 | 1305* | <40 | * | Ηπατίτιδα |
| 9 | 1305* | <40 | * | Άσθμα |
| 10 | 1305* | <40 | * | Άσθμα |
| 5 | 1485* | ≥40 | * | Άσθμα |
| 6 | 1485* | ≥40 | * | Καρδιαγγειακό |
| 7 | 1485* | ≥40 | * | Μόλυνση από ιό |
| 8 | 1485* | ≥40 | * | Μόλυνση από ιό |
| 2 | 1306* | <40 | * | Καρδιαγγειακό |
| 3 | 1306* | <40 | * | Ηπατίτιδα |
| 11 | 1306* | <40 | * | Άσθμα |
| 12 | 1306* | <40 | * | Άσθμα |

Συνεχίζοντας το παράδειγμα με τους δύο γείτονες, σε αυτή την περίπτωση ο γείτονας Β χρειάζεται *l*-1 δεδομένα του γνωστικού υπόβαθρου που μπορεί να προκαλέσουν κάποια διαρροή σε πληροφορίες της ιδιωτικότητας ενός χρήστη, για να εξαλείψει *l*-1 πιθανά ευαίσθητα χαρακτηριστικά και να κάνει μια αποκάλυψη. Έτσι, ρυθμίζοντας την παράμετρο *l* ο εκδότης των δεδομένων μπορεί να καθορίσει πόση προστασία παρέχεται έναντι του γνωστικού υπόβαθρου ακόμη και αν το γνωστικό υπόβαθρο είναι άγνωστο στον εκδότη.

### 2.1.3 Η *t*-εγγύτητα *(t-closeness)*

Σύμφωνα με τους *Li et al.* (2007), όσον αφορά την *l*-πολυμορφία, ένα πρόβλημα που υφίσταται είναι ότι έχει περιορισμούς στις προϋποθέσεις για τις γνώσεις των αντιπάλων. Ένα άλλο πρόβλημα με τις μεθόδους προστασίας ιδιωτικότητας, σε γενικές γραμμές, είναι ότι

υποτίθεται ότι όλα τα χαρακτηριστικά είναι κατηγορικά, δηλαδή ο αντίπαλος είτε έχει είτε δεν έχει μάθει κάτι ευαίσθητο.

Προτείνεται λοιπόν, μια νέα έννοια για την προστασία της ιδιωτικότητας που ονομάζεται *t*-εγγύτητα, η οποία τυποποιεί την ιδέα του γενικού γνωστικού υποβάθρου απαιτώντας η κατανομή του κάθε ευαίσθητου χαρακτηριστικού στην τάξη ισοδυναμίας, να είναι κοντά στην κατανομή του χαρακτηριστικού στο γενικό πίνακα (δηλαδή, η απόσταση μεταξύ των δύο κατανομών να μην υπερβαίνει ένα όριο *t*). Αυτό περιορίζει αποτελεσματικά την ποσότητα των σημαντικών πληροφοριών ενός χρήστη που μπορεί να μάθει ένας παρατηρητής.

❖ Ορισμός 2.2

Η αρχή της *t*-εγγύτητας *(The t-closeness principle)*. Μια τάξη ισοδυναμίας λέγεται ότι έχει *t*-εγγύτητα, αν η απόσταση μεταξύ της κατανομής ενός ευαίσθητου χαρακτηριστικού αυτής της κλάσης και της κατανομής του χαρακτηριστικού σε ολόκληρο τον πίνακα δεν ξεπερνάει ένα όριο *t*. Ένας πίνακας λέγεται ότι έχει *t*-εγγύτητα, αν όλες οι τάξεις ισοδυναμίας του έχουν *t*-εγγύτητα.

Στις επόμενες ενότητες θα ασχοληθούμε με βασικές έννοιες σχετικά με τη χωρική προστασία ιδιωτικότητας, την προστασία της ιδιωτικότητας για δεδομένα κίνησης και τη λεπτομερή ανάλυση τεχνικών προστασίας ιδιωτικότητας σε χωρικά και χωροχρονικά δεδομένα. Πιο συγκεκριμένα θα αναλύσουμε τις παρακάτω τρεις πτυχές:

1. Ανωνυμία σε Υπηρεσίες Εντοπισμού Θέσης *(Anonymity in Location-Based Services)*
2. Προστασία Ιδιωτικότητας κατά τη Δημοσιοποίηση Δεδομένων Κίνησης *(Privacy-Preserving Mobility Data Publishing)*
3. Προστασία Ιδιωτικότητας σε Επερωτήσεις Δεδομένων Κίνησης *(Privacy-Aware Mobility Data Querying)*

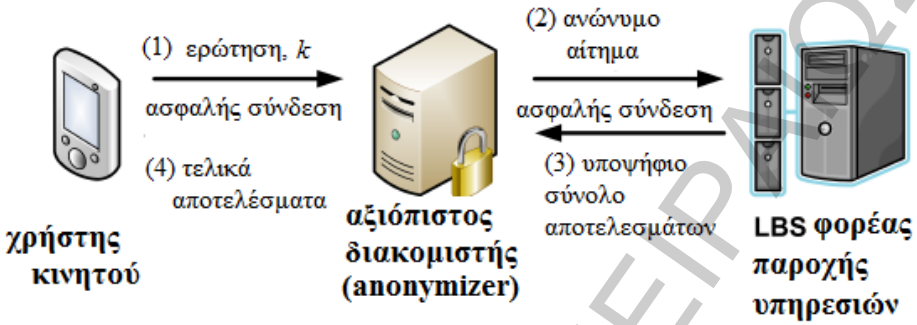
## 2.2 Ανωνυμία σε Υπηρεσίες Εντοπισμού Θέσης *(Anonymity in Location-Based Services)*

Σύμφωνα με τους *Gkoulalas et al.* το 2010, αν ένας χρήστης υποβάλλει αίτημα *LBS* απευθείας σε εκείνον που παρέχει τις υπηρεσίες, τότε η ταυτότητά του μπορεί να αποκαλυφθεί και η προσωπική του ζωή μπορεί να τεθεί σε κίνδυνο. Για αυτό το λόγο, όπως βλέπουμε και στο σχήμα 2.2, το μοντέλο της κεντρικής προστασίας της ιδιωτικότητας προϋποθέτει ότι κάθε αίτημα του χρήστη για *LBS* πρέπει να υποβάλλεται σε έναν αξιόπιστο

διακομιστή του φορέα εκμετάλλευσης των τηλεπικοινωνιών, μέσω ενός ασφαλούς διαύλου επικοινωνίας. Ο ρόλος ενός αξιόπιστου διακομιστή *(anonymizer)* είναι να φιλτράρει τα εισερχόμενα αιτήματα των χρηστών και να παράγει αντίστοιχα ανώνυμα.

**Σχήμα 2.2**

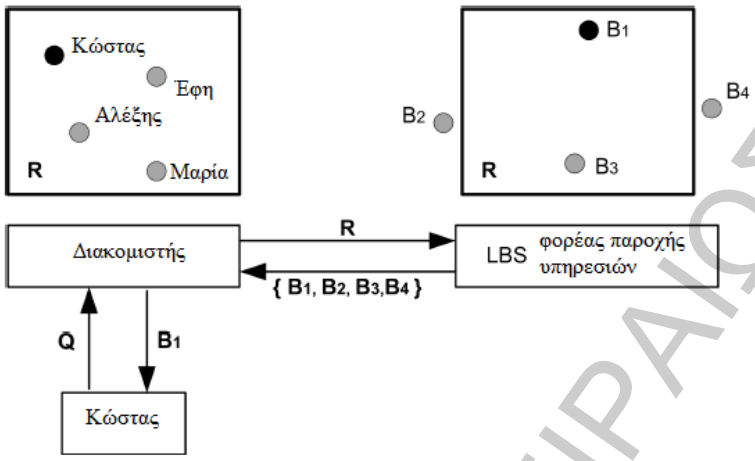
Το συγκεντρωτικό μοντέλο προστασίας ιδιωτικότητας σε *LBS*




Σε αυτή την παράγραφο παραθέτουμε ένα παράδειγμα για το πως λειτουργούν οι υπηρεσίες που είναι βασισμένες στο χώρο. Ο χρήστης Κώστας (σχήμα 2.3) ζητάει το πιο κοντινό περίπτερο $B_i$ και προωθεί το αίτημά του $Q$ στο διακομιστή. Τότε ο διακομιστής ο οποίος έχει γνώση της τρέχουσας θέσης του κάθε χρήστη στο σύστημα, αναγνωρίζει 3 χρήστες που βρίσκονται κοντά στον Κώστα και περικλείει και τους 4 χρήστες σε μια περιοχή *R*. Αντί λοιπόν να στείλει στο *LBS* που βρίσκεται ο Κώστας, στέλνει την περιοχή *R*. Όταν το *LBS* λαμβάνει το *R*, υπολογίζει όλα τα περίπτερα σε οποιοδήποτε σημείο του *R*. Είναι σημαντικό να σημειωθεί ότι παρόλο που το *LBS* ξέρει πως ο Κώστας βρίσκεται στο *R*, δεν έχει κανένα μέσο για να προσδιορίσει την ακριβή του θέση. Χρησιμοποιώντας τη βάση δεδομένων, το *LBS* δίνει ένα υποψήφιο σύνολο αποτελεσμάτων $\{B_1, B_2, B_3, B_4\}$ και τα προωθεί στον διακομιστή. Ο διακομιστής χρησιμοποιεί την πραγματική θέση του Κώστα μέσα στο *R* και φιλτράρει όλες τις εσφαλμένες απαντήσεις ώστε να προωθήσει το πραγματικό πλησιέστερο περίπτερο (σε αυτή την περίπτωση το $B_1$). Αυτό το βήμα ολοκληρώνει την παροχή των *LBS* με ασφαλή τρόπο.

**Σχήμα 2.3**

Ένα παράδειγμα χωρικής *k*-ανωνυμίας




Σε γενικές γραμμές ο επιτιθέμενος θεωρείται ότι έχει τις εξής ικανότητες:

- Μπορεί να υποκλέψει την περιοχή όπου προσφέρεται ανωνυμία στον αιτούντα ενός *LBS*.
- Έχει γνώση των αλγόριθμων που χρησιμοποιούνται από τον αξιόπιστο διακομιστή για να προσφέρει προστασία ιδιωτικότητας στο *LBS*.
- Μπορεί να αποκτήσει την τρέχουσα θέση όλων των χρηστών του συστήματος.
- Προσπαθεί να παραβιάσει την ιδιωτικότητα της θέσης του χρήστη χρησιμοποιώντας μόνο τα τρέχοντα δεδομένα θέσης.

### 2.2.1 Ταξινόμηση των προσεγγίσεων *k*-ανωνυμίας

Το κύριο μέρος της έρευνας για την προστασία της ιδιωτικότητας στο *LBS*, περιλαμβάνει προσεγγίσεις που βασίζονται στην έννοια της *k*-ανωνυμίας. Παρακάτω παρουσιάζονται κάποιες από αυτές.

### 2.2.1.1 Οι τεχνικές της απόκρυψης *(Cloaking Techniques)*

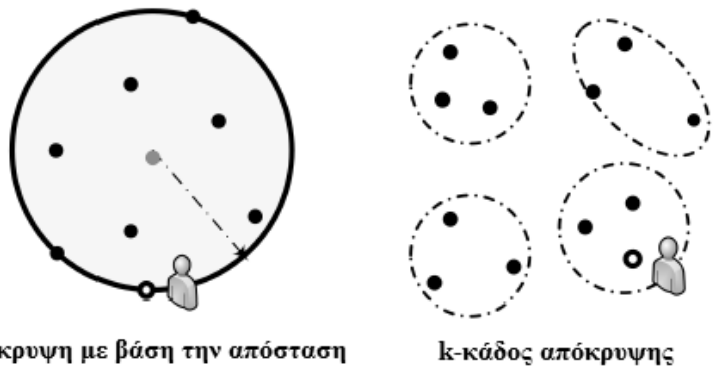
Για να ικανοποιηθεί η *k*-ανωνυμία στο *LBS*, η πιο ευρέως διαδεδομένη τεχνική ανωνυμίας είναι η απόκρυψη *(cloaking)*. Στην απόκρυψη, η πραγματική θέση του αιτήματος μετατρέπεται σε μια φραγμένη περιοχή που είναι αρκετά μεγάλη για να περιέχει τον αιτούντα μαζί με τουλάχιστον *k*-1 άλλους χρήστες. Η απόκρυψη εξασφαλίζει ότι η ταυτότητα του αιτούντος δεν μπορεί να αποκαλυφθεί με πιθανότητα σημαντικά μεγαλύτερη από $\frac{1}{k}$ μεταξύ των *k*-1 άλλων χρηστών.

Στη συνέχεια, ακολουθούν μερικές από τις πιο διαδεδομένες τεχνικές απόκρυψης που γενικεύουν τις πραγματικές θέσεις του αιτήματος σε χωρικά οριοθετημένες περιοχές. Μπορούν να χωριστούν σε δύο ομάδες: της απόκρυψης που εξαρτάται από τα δεδομένα και της απόκρυψης που εξαρτάται από το χώρο.

A. Απόκρυψη βασιζόμενη στα δεδομένα *(Data-based cloaking)*. Οι αλγόριθμοι απόκρυψης που βασίζεται στα δεδομένα, διαμορφώνουν την περιοχή της ανωνυμίας με βάση την πραγματική θέση του κάθε χρήστη στο σύστημα και την απόστασή του από τη θέση του αιτήματος. Πιο συγκεκριμένα, όπως φαίνεται και στο σχήμα 2.4 στην απόκρυψη που βασίζεται στην απόσταση *(distance-based cloaking)* ανακτώνται οι *k*-1 κοντινότεροι γείτονες του αιτούντος και δημιουργούν μια περιοχή που περιλαμβάνει όλους τους *k* χρήστες. Στον *k*-κάδο απόκρυψης *(k-bucket cloaking)* οι χρήστες οργανώνονται σε *k* ομάδες και η περιοχή ανωνυμίας υπολογίζεται ως το Ελάχιστο Περιβάλλον Ορθογωνίου (*MBR*-έκφραση των μέγιστων εκτάσεων ενός αντικειμένου 2-διαστάσεων) που περιέχει τους *k* χρήστες στην περιοχή του αιτούντος.

**Σχήμα 2.4**

Απόκρυψη βασιζόμενη στα δεδομένα

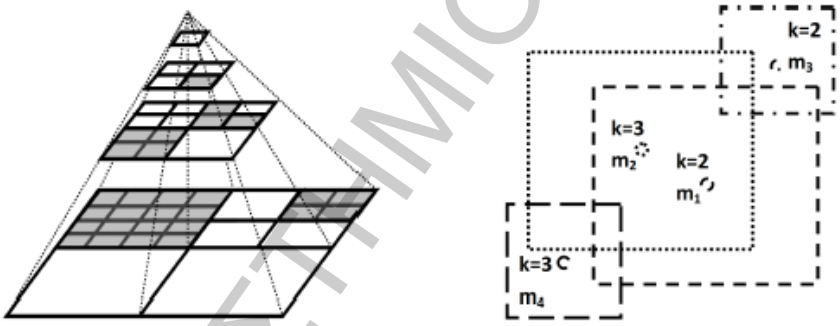


απόκρυψη με βάση την απόσταση  k-κάδος απόκρυψης

Β. Απόκρυψη βασιζόμενη στο χώρο *(Space-dependant cloaking).* Για να διατυπώσει τις περιοχές της ανωνυμίας, λαμβάνει υπόψη τη συνολική έκταση που καλύπτεται από τον αξιόπιστο διακομιστή. Συγκεκριμένα όπως φαίνεται στο σχήμα 2.5, οι τεχνικές απόκρυψης με βάση το πλέγμα *(grid-based cloaking),* χωρίζουν την περιοχή σε διαμερίσματα με ένα πλέγμα. Σκοπός τους είναι να δημιουργήσουν την περιοχή της ανωνυμίας ανακτώντας τους χρήστες από κάθε κελί του πλέγματος. Ξεκινάνε από το κελί του αιτούντος και κινούνται προς γειτονικά κελιά, έως ότου οι χρήστες που βρέθηκαν να είναι τουλάχιστον *k*. Από την άλλη, οι στρατηγικές απόκρυψης με βάση την περιοχή *(region-based cloaking),* για την παραγωγή ορθογώνιων χρειάζονται τις χωρικές ιδιότητες της περιοχής που επικεντρώνονται στη θέση του αιτήματος και τις χρησιμοποιούν για την προσφορά της *k*-ανωνυμίας.

**Σχήμα 2.5**

Απόκρυψη βασιζόμενη στο χώρο




### 2.2.1.2 Χωρική *k*-ανωνυμία *(Location k-anonymity)*

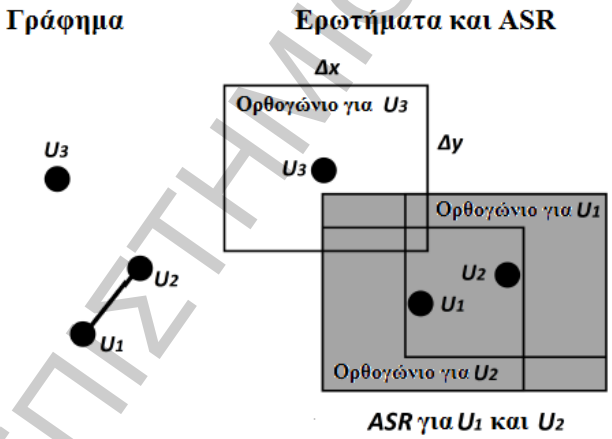
Οι προσεγγίσεις της χωρικής *k*-ανωνυμίας, προστατεύουν την ιδιωτικότητα των χρηστών χρησιμοποιώντας την τρέχουσα θέση του κάθε χρήστη στο σύστημα. Πιο δημοφιλείς προσεγγίσεις είναι οι παρακάτω:

Ι. Η απόκρυψη κλίκας *(clique cloak),* είναι μια τεχνική που βασίζεται στην περιοχή και μετατρέπει σε ανώνυμα, πολλαπλά εισερχόμενα αιτήματα για *LBS*. Για κάθε ερώτημα που λαμβάνει προς εξυπηρέτηση, ο αλγόριθμος δημιουργεί ένα ορθογώνιο με κέντρο τη θέση του αιτούντα, με τις πλευρές του παράλληλες με τους υπό εξέταση *x*, *y* άξονες

και με εκτάσεις *Δx* και *Δy* αντίστοιχα. Το νέο ερώτημα στη συνέχεια, επισημαίνεται ως ένας κόμβος σε ένα γράφημα για όσο χρονικό διάστημα βρίσκεται «σε αναμονή» για την ανωνυμία του. Δύο κορυφές (ερωτήματα) στο γράφημα συνδέονται μεταξύ τους αν συμπέσει ο ένας χρήστης στο τετράγωνο του άλλου. Μια πλευρά του γραφήματος δείχνει ότι ο αιτών καθενός από τα δύο ερωτήματα μπορεί να συμπεριληφθεί στον άξονα ανωνυμίας του άλλου και έτσι ένα *k-clique* του γραφήματος δείχνει ότι όλες οι αντίστοιχες *k* αιτήσεις μπορούν να είναι ανώνυμες μαζί. Τέλος, κοντά σε κάθε αίτημα είναι ένα χρονικό διάστημα *Δt* που ορίζει το μέγιστο χρονικό διάστημα που το αίτημα αυτό μπορεί να διατηρηθεί από το σύστημα για την ανωνυμία του. Αν μία ομάδα *k*-clique αιτημάτων δεν μπορεί να βρεθεί μέσα σε ένα διάστημα *Δt* τότε το αίτημα δεν εξυπηρετείται.

**Σχήμα 2.6**

Ένα παράδειγμα της τεχνικής απόκρυψης κλίκας




Το σχήμα 2.6, δίνει ένα παράδειγμα της τεχνικής της απόκρυψης κλίκας, σε περίπτωση που υπάρχουν 3 ερωτήματα ($U_1$,$U_2$,$U_3$) για *LBS* τα οποία έχουν συγχρόνως υποβληθεί στον αξιόπιστο διακομιστή. Υποθέτοντας λοιπόν ότι *k*=2, τα ορθογώνια που έχουν δημιουργηθεί για $U_1$,$U_2$ εμπίπτουν το ένα στο άλλο και έτσι σχηματίζουν μια *2-clique* στο γράφημα. Ως αποτέλεσμα, το *MBR* που περιβάλλει τα αντίστοιχα ορθογώνια (παρουσιάζονται εδώ σε γκρι) αντιπροσωπεύει την περιοχή χωρικής ανωνυμίας *(Anonymity Spatial Region)*, όπου στην περίπτωση αυτών των

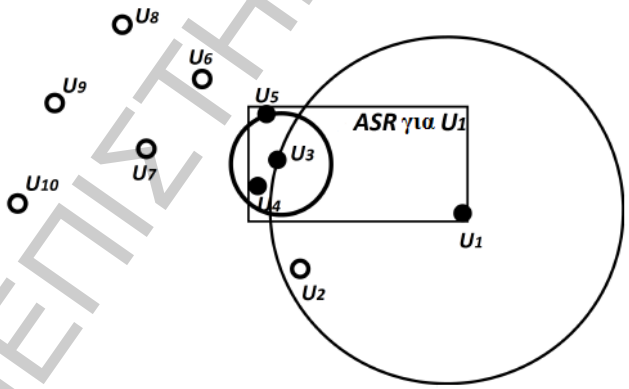
χρηστών είναι 2-ανωνυμία. Από την άλλη πλευρά, το αίτημα του $U_3$ πρέπει να περιμένει στο σύστημα έως ότου ένα νέο αίτημα εμφανιστεί.

II. Η τεχνική της κεντρικής απόκρυψης *(center cloak)*, είναι μια προσέγγιση βασισμένη στην απόσταση που παρέχει μια πιο απλοϊκή λύση της *k*-ανωνυμίας σε *LBS*. Οι *k*-1 κοντινότεροι γείτονες του αιτούντος ανακτώνται και η *ASR* υπολογίζεται ως το *MBR* που περιβάλλει όλους τους *k* χρήστες.

III. Μια παραλλαγή της κεντρικής απόκρυψης, που προσφέρει αυξημένη αβεβαιότητα σχετικά με τη θέση του αιτούντος στο παραγόμενο *ASR*, είναι η τεχνική της απόκρυψης του κοντινότερου γείτονα *(NN-Cloak)*. Στη τεχνική αυτή, το *ASR* διατυπώνεται ως εξής: Δοθέντος του ερωτήματος ενός χρήστη για *LBS*, η τεχνική ανακτά τους *k*-1 πλησιέστερους γείτονες του αιτούντος. Στη συνέχεια, επιλέγει τυχαία έναν από τους χρήστες και εντοπίζει τους *k*-1 κοντινότερους γείτονες. Τέλος, το *k-ASR* κατασκευάζεται ως το *MBR* που περιβάλλει την δεύτερη ομάδα των *k* χρηστών, ενισχυμένο (αν χρειαστεί) για να συμπεριλάβει τον αιτούντα.

**Σχήμα 2.7**

Ένα παράδειγμα της τεχνικής της απόκρυψης του κοντινότερου γείτονα

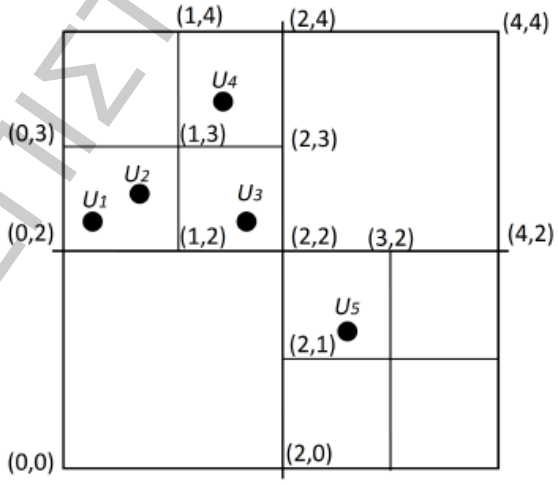



Το σχήμα 2.7, μας δίνει ένα παράδειγμα της λειτουργίας της τεχνικής της απόκρυψης του κοντινότερου γείτονα όπου προσφέρεται 3-ανωνυμία στο χρήστη $U_1$. Αρχικά δημιουργεί ένα σύνολο $S_1=\{U_1, U_2, U_3\}$ με τους δύο κοντινότερους γείτονες του χρήστη. Στη συνέχεια επιλέγει τυχαία τον $U_3$ από το σύνολο και υπολογίζει τους δύο πιο κοντινούς του γείτονες στο σύστημα. Αυτό οδηγεί στο σύνολο $S_2=\{U_3,U_4,U_5\}$. Τελικά το *MBR* του $S_2$ ενισχύεται για να συμπεριλάβει τον αιτούντα $U_1$ οδηγώντας τον σε 4-*ASR*.

IV. Η τεχνική *Casper* είναι μία από τις πιο δημοφιλείς προσεγγίσεις για χωρική *k*-ανωνυμία που βασίζονται σε πλέγμα *(grid-based)*. Στη προσέγγιση αυτή, ολόκληρη η περιοχή που καλύπτεται από τον αξιόπιστο διακομιστή χωρίζεται με ένα είδος πλέγματος και οργανώνεται με μια δομή πυραμίδας δεδομένων σε στρώματα. Κάθε κελί στο χαμηλότερο επίπεδο της πυραμίδας, έχει ελάχιστο μέγεθος που αντιστοιχεί στην επίλυση της ανωνυμίας. Όταν ένα νέο ερώτημα για *LBS* παραλαμβάνεται από τον αξιόπιστο διακομιστή, ο *Casper* το τοποθετεί στο χαμηλότερο επίπεδο του κελιού στην πυραμίδα που περιέχει τον αιτούντα και εξετάζει εάν αυτό το κελί περιέχει *k*-1 άλλους χρήστες. Αν το κελί περιέχει αρκετούς χρήστες τότε δημιουργείται το *k-ASR*. Διαφορετικά, ο *Casper* αναζητά τον οριζόντιο και τον κάθετο γείτονα αυτού του κελιού για να προσδιορίσει εάν ο αριθμός των χρηστών σε κάθε ένα από τα κελιά αυτά σε συνδυασμό με τον αριθμό των χρηστών στο κελί του αιτούντος, αρκεί για την παροχή *k*-ανωνυμίας. Εάν τελικά αρκεί, τότε η αντίστοιχη ένωση κελιών γίνεται *k-ASR*. Διαφορετικά, ο *Casper* κινείται ένα επίπεδο πάνω στην πυραμίδα για να ανακτήσει το μητρικό κελί από το κελί του αιτήματος και επαναλαμβάνει την ίδια διαδικασία μέχρι να βρεθούν οι *k* χρήστες που θα διαμορφώσουν το *ASR*.

**Σχήμα 2.8**

Ένα παράδειγμα της τεχνικής του *Casper*




Το σχήμα 2.8, μας δίνει ένα παράδειγμα της λειτουργίας του *Casper*. Υποθέτοντας ότι ένα αίτημα που προέρχεται από το κελί ⟨(0,2),(1,3)⟩ (όπου η (0,2) είναι η κατώτερη αριστερά και η (1,3) η ανώτερη δεξιά συντεταγμένη του κελιού), με μια απαιτούμενη

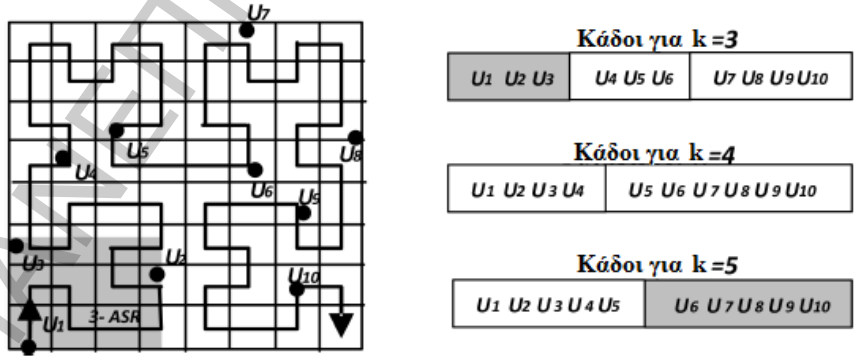
ανωνυμία *k*=2, το επιστρεφόμενο *ASR* είναι το ίδιο κελί. Σε περίπτωση που ένα ερώτημα με τις ίδιες απαιτήσεις ανωνυμίας προέρχεται από το κελί ⟨(1,2),(2,3)⟩, το επιστρεφόμενο *ASR* είναι η ένωση κελιών ⟨(1,2),(2,3)⟩ ∪ ⟨(1,3),(2,4)⟩.

V. Η τεχνική της απόκρυψης διαστήματος *(interval cloak)* είναι παρόμοια με αυτή του *Casper*, καθώς χωρίζει τη συνολική έκταση που καλύπτεται από τον αξιόπιστο διακομιστή σε ίσου μεγέθους τεταρτημόρια και οργανώνει τις πληροφορίες αυτές σε δενδροειδή δομή. Ωστόσο, η τεχνική αυτή δεν εξετάζει τα γειτονικά κελιά του ίδιου επιπέδου κατά τον υπολογισμό του *ASR*, αλλά ανεβαίνει απευθείας στο πρόγονο επίπεδο της πυραμίδας.

VI. Η τεχνική της απόκρυψης του *Hilbert (Hilbert cloak)* δημιουργεί μία τρισδιάστατη χαρτογράφηση της θέσης του κάθε χρήστη (σχήμα 2.9). Στη προτεινόμενη χαρτογράφηση, τοποθεσίες που βρίσκονται κοντά μεταξύ τους στο δύο διαστάσεων πεδίο, αναμένεται να βρίσκονται επίσης κοντά σε ένα άλλο μετασχηματισμό μιας διάστασης. Για κάθε αίτημα με την απαίτηση της *k*-ανωνυμίας, η απόκρυψη του *Hilbert* χωρίζει τις ομάδες *k* χρηστών του συστήματος σε έναν κάδο σύμφωνα με τις *Hilbert* τιμές τους. Μετά από αυτό, η απόκρυψη *Hilbert* ανακτά όλους τους *k*-1 χρήστες που βρίσκονται στον ίδιο κάδο με τον αιτούντα και διαμορφώνει το *k-ASR* ως το *MBR* που τα περικλείει.

**Σχήμα 2.9**

Ένα παράδειγμα της τεχνικής της απόκρυψης του *Hilbert*

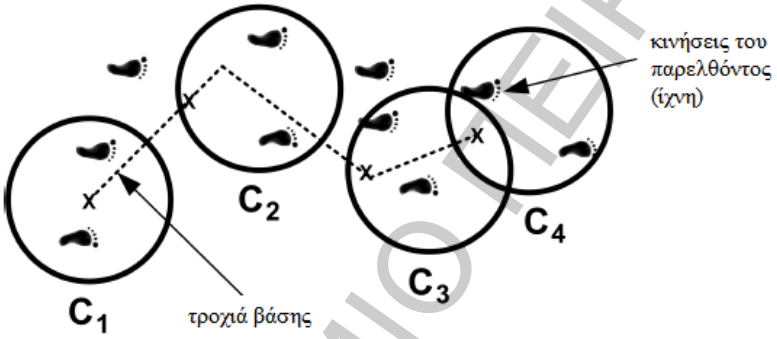


### 2.2.1.3 Ιστορική *k*-ανωνυμία *(Historical k-anonymity)*

Σε σύγκριση με άλλες μεθοδολογίες για την προσφορά της *k*-ανωνυμίας στο *LBS*, η ιστορική *k*-ανωνυμία (σχήμα 2.10) προσεγγίζει τους συμμετέχοντες της ανωνυμίας που έχουν επιλεγεί με βάση το ιστορικό της εξέλιξής τους στο σύστημα, με την προϋπόθεση ότι κάποια στιγμή στην ιστορία της κίνησής τους οι χρήστες ήταν κοντά στο σημείο του αιτήματος.

**Σχήμα 2.10**

*k*-ανωνυμία βασισμένη σε μετακινήσεις του παρελθόντος




### 2.2.1.4 *k*-ανωνυμία σε τροχιές κινούμενων αντικειμένων *(Trajectory k-anonymity)*

Οι προσεγγίσεις της *k*-ανωνυμίας σε τροχιές κινούμενων αντικειμένων, είναι κατάλληλες για τη διατήρηση της ιδιωτικότητας των χρηστών που ζητούν *LBS* και που δεν μπορεί να τους προσφερθεί όταν ο χρήστης έχει επικοινωνήσει μία μόνο φορά με τον πάροχο υπηρεσιών. Σε αντίθεση με τις προσεγγίσεις της ιστορικής *k*-ανωνυμίας οι οποίες μπορούν να προστατεύσουν την τροχιά ενός χρήστη παρέχοντας *k*-ανωνυμία στον αιτούντα των *LBS*, δημιουργούν το *k-ASR* χρησιμοποιώντας την τρέχουσα αντί της ιστορικής κίνησης των χρηστών στο σύστημα, ώστε να καλύπτουν επαρκώς την τροχιά του αιτούντα.

Οι γενικές προσεγγίσεις της *k*-ανωνυμίας σε τροχιές κινούμενων αντικειμένων, δεν λαμβάνουν υπόψη κάποια συγκεκριμένη συμπεριφορά στην κίνηση του αιτούντα ενός *LBS* (καθώς και άλλων χρηστών του συστήματος) όταν του παρέχεται ανωνυμία. Όλα τα αιτήματα για *LBS* γίνονται ακριβώς με τον ίδιο τρόπο από τον αξιόπιστο διακομιστή, χωρίς να έχει

σημασία ποια είναι η θέση του αιτήματος ή η διαδρομή που ακολουθεί ο χρήστης στο σύστημα κατά την παροχή *LBS*.

Οι εξατομικευμένες προσεγγίσεις της *k*-ανωνυμίας σε τροχιές κινούμενων αντικειμένων, χρησιμοποιούν το ιστορικό της κίνησης όλων των χρηστών του συστήματος για την αντιμετώπιση των επιθέσεων συσχέτισης σε συνεχή ερωτήματα των χρηστών. Διαφέρουν από τις γενικές προσεγγίσεις προστασίας ιδιωτικότητας στο *LBS* για τους εξής λόγους:

- θεωρούν ότι οι επιτιθέμενοι γνωρίζουν τη συμπεριφορά της κίνησης των χρηστών στο σύστημα και μπορούν να χρησιμοποιήσουν τη γνώση των συχνών προτύπων κίνησης των χρηστών για την παραβίαση της ιδιωτικής τους ζωής.
- απεικονίζουν την κίνηση του κάθε χρήστη στο σύστημα ως μία συνεχή συνάρτηση αντί για ένα σύνολο ατομικών θέσεων και χρόνων.
- λαμβάνουν αυτόματα μια σειρά από συχνά πρότυπα κίνησης ανά χρήστη βάσει του ιστορικού της κίνησής του στο σύστημα. Στη συνέχεια τα χρησιμοποιούν για την προστασία της προσωπικής του ζωής όταν κάνουν συνεχόμενα αιτήματα στο *LBS*.
- μπορούν να προσφέρουν *k*-ανωνυμία σε τροχιές κινούμενων αντικειμένων στους αιτούντες των *LBS*, υποθέτοντας την τοπολογία του δικτύου των κινήσεων του χρήστη.

## 2.3 Προστασία ιδιωτικότητας σε βάσεις δεδομένων τροχιών

Σε αυτό το υποκεφάλαιο, θα ασχοληθούμε με τις τεχνικές διαφύλαξης ιδιωτικότητας τροχιών, οι οποίες μπορούν να χωριστούν σε δύο κατηγορίες: (α) Προστασία Ιδιωτικότητας κατά τη Δημοσιοποίηση Δεδομένων Κίνησης και (β) Προστασία Ιδιωτικότητας σε Επερωτήσεις Δεδομένων Κίνησης. Η πρώτη κατηγορία αφορά προσεγγίσεις που έχουν σαν σκοπό τη δημοσίευση μιας ανώνυμης εκδοχής του αρχικού συνόλου δεδομένων, ενώ παράλληλα γίνεται προσπάθεια να διατηρηθεί όσο μεγαλύτερη χρησιμότητα των δεδομένων είναι εφικτό κατά τη διάρκεια της μετατροπής. Η δεύτερη κατηγορία βασίζεται στην υπόθεση ότι τα δεδομένα θα πρέπει να παραμείνουν εντός του οργανισμού που τα φιλοξενεί (*in-house).*

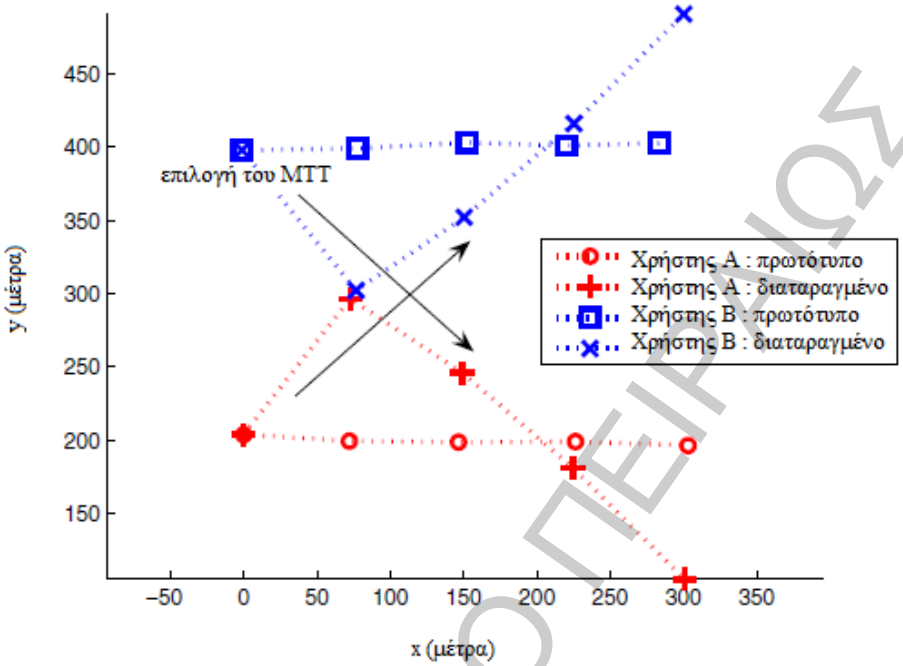
### 2.3.1 Προστασία Ιδιωτικότητας κατά τη Δημοσιοποίηση Δεδομένων Κίνησης *(Privacy-preserving mobility data publishing)*

#### 2.3.1.1 Προστασία Ιδιωτικότητας Τοποθεσίας μέσω Σύγχυσης Διαδρομής *(Protecting Location Privacy Through Path Confusion)*

Η τεχνική της Σύγχυσης Διαδρομής (*Hoh & Gruteser*, 2005) ανήκει στην προσέγγιση της διατάραξης. Σε αυτήν την προσέγγιση οι αλγόριθμοι επιβάλλουν κάποια «λάθη» σε αποδεκτά πλαίσια στα αρχικά δείγματα των τοποθεσιών, με σκοπό τη διατήρηση του επιπέδου ποιότητας των υπηρεσιών που παρέχονται στους χρήστες. Ο αλγόριθμος της διαταραχής διαδρομής *(path perturbation)* ακολουθεί την εξής διαδικασία: όταν δύο τροχιές (που δεν τέμνονται) είναι πολύ κοντά, δημιουργεί μία ψεύτικη διασταύρωση. Έτσι κάθε φορά που δύο μονοπάτια συναντιούνται (ορίζουμε ως συνάντηση μια κοντινή απόσταση), υπάρχει μια πιθανότητα για τον αντίπαλο να μπερδέψει τις τροχιές και να ακολουθήσει το λάθος χρήστη. Με αυτό το τρόπο, μεγιστοποιείται στιγμιαία η προστασία ιδιωτικότητας της τοποθεσίας σε κάθε βήμα, τροποποιώντας το αρχικό σύνολο των δειγμάτων τοποθεσίας εντός μιας ακτίνας διατάραξης *R*. Τα μεγαλύτερα *R* έχουν σαν αποτέλεσμα μεγαλύτερο βαθμό προστασίας, ενώ τα μικρότερα *R* περιορίζουν την επίδραση της διαταραχής η οποία οδηγεί σε υψηλότερη ποιότητα υπηρεσιών και χαμηλότερη προστασία ιδιωτικότητας.

**Σχήμα 2.11**

Παράδειγμα διασταύρωσης διαδρομών μεταξύ δύο τροχιών




### 2.3.1.2 Προσεγγίσεις με βάση την ομαδοποίηση *(Clustering-based approaches)*

Σε αυτές τις προσεγγίσεις ανήκουν οι μέθοδοι *Never Walk Alone-NWA (Abul et al.*, 2010) και *Wait For Me-W4M* (*Abul et al.*, 2010).

Λόγω της δειγματοληψίας και της ασάφειας των συστημάτων εντοπισμού θέσης (π.χ. *GPS*), η τροχιά ενός κινούμενου αντικειμένου δεν είναι πλέον ένα γράφημα με πολλές γραμμές σε τρισδιάστατο χώρο, αλλά ένας κυλινδρικός όγκος όπου η ακτίνα του $\delta$ αντιπροσωπεύει την πιθανή ασαφή θέση (σχήμα 2.12). Γνωρίζουμε ότι η τροχιά του κινούμενου αντικειμένου είναι μέσα στον κύλινδρο, αλλά δεν γνωρίζουμε που ακριβώς. Εάν κάποιο άλλο αντικείμενο κινείται εντός του ίδιου κυλίνδρου είναι δυσδιάκριτο από κάθε άλλο. Παρακάτω αναφέρουμε δύο ορισμούς που χρησιμεύουν για την περαιτέρω κατανόηση αυτών των προσεγγίσεων.

❖ Ορισμός 2.3

Συν-εντοπισμός *(Co-localization).* Δύο τροχιές $\tau_1$, $\tau_2$ που ορίζονται στο $[t_1, t_n]$ λέγεται ότι είναι συν-εντοπισμένες με περιοχή αβεβαιότητας $\delta$, αν και μόνο αν για κάθε σημείο

$(x_1, y_1, t)$ κατά μήκος του $\tau_1$ και $(x_2, y_2, t)$ κατά μήκος του $\tau_2$ και με $t \in [t_1, t_n]$ ισχύει ότι $Dist((x_1, y_1),(x_2, y_2)) \leq \delta$.

Όπου *Dist* είναι η Ευκλείδεια απόσταση :

$$Dist\ ((x_1, y_1), (x_2, y_2)) \leq \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Γράφεται $Coloc_\delta\ (\tau_1, \tau_2)$ παραλείποντας το χρονικό διάστημα $[t_1, t_2]$.

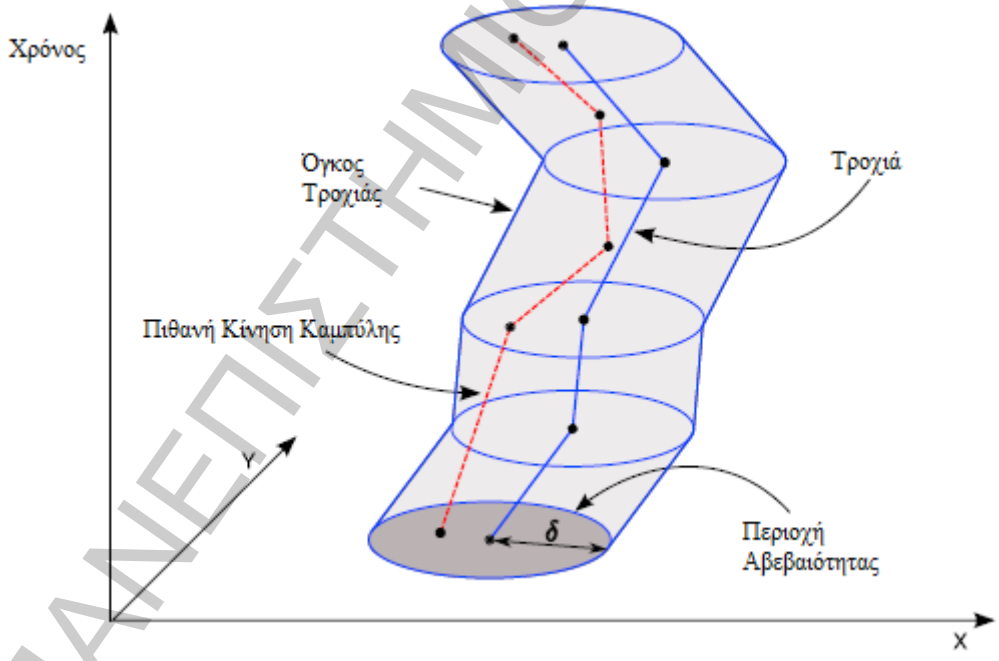
Δοθέντος ενός ορίου αβεβαιότητας *k*, μπορεί να ορισθεί ένα σύνολο ανωνυμίας σαν ένα σύνολο από *k* τουλάχιστον τροχιές που είναι συνεντοπισμένες (σχήμα 2.13).

❖ Ορισμός 2.4

Ανωνυμία ενός συνόλου τροχιών *(Anonymity set of trajectories)*. Δοθέντος ενός ορίου αβεβαιότητας $\delta$ και ενός ορίου ανωνυμίας *k*, ένα σύνολο τροχιών *S* είναι *(k,δ)*-ανώνυμο, αν και μόνο αν $|S| \geq k$ και $\forall\ \tau_i, \tau_j \in S$. $Coloc_\delta\ (\tau_i, \tau_j)$.

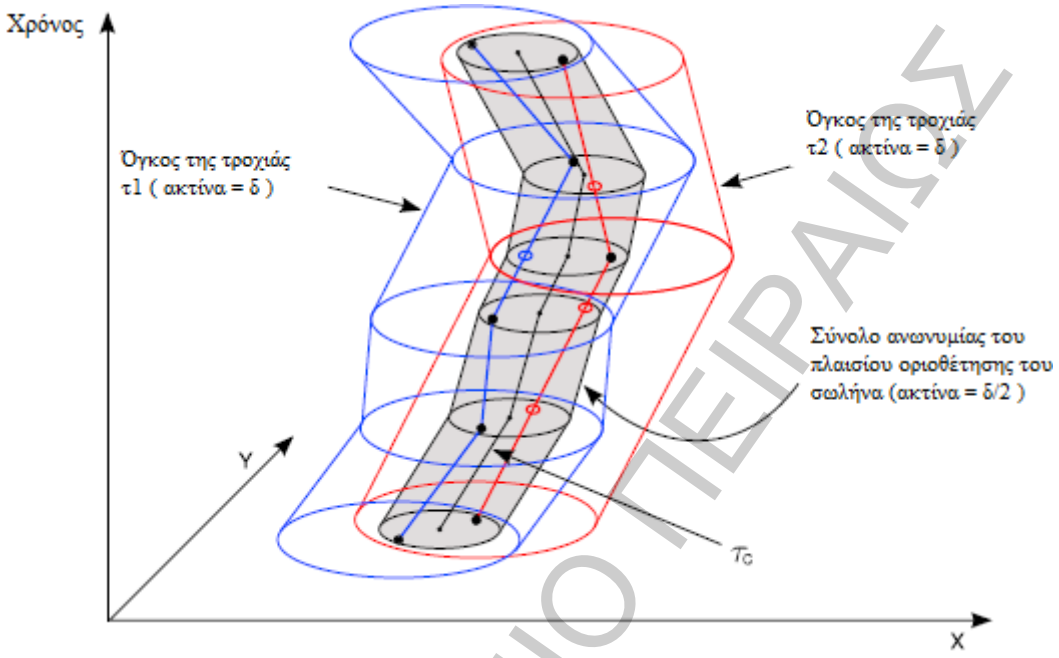
**Σχήμα 2.12**

Περιοχή αβεβαιότητας: η τροχιά, ο όγκος τροχιάς και η πιθανή καμπύλη κίνησης



**Σχήμα 2.13**

Σύνολο ανωνυμίας που σχηματίζεται από δύο συνεντοπισμένες τροχιές, τους αντίστοιχους όγκους αβεβαιότητας και τον κεντρικό κυλινδρικό όγκο της ακτίνας δ/2 που περιέχει και τις δύο τροχιές




Η μέθοδος *NWA*, βασίζεται στην ομαδοποίηση και την χωρική διαταραχή και αναπτύσσεται γύρω από τρεις κύριες φάσεις:

1. Προ-επεξεργασία *(Pre-processing)*: όλες οι τροχιές χωρίζονται σε ισοδύναμες κλάσεις, έτσι ώστε η κάθε μία να περιέχει αντικείμενα που έχουν τον ίδιο αρχικό και τελικό χρόνο.
2. Ομαδοποίηση (*Clustering*): γίνεται με βάση τη μέθοδο *GC* και ενισχυμένη με τεχνικές για να κρατηθεί μικρή η ακτίνα των παραγόμενων συστάδων, με τίμημα την καταστολή κάποιων ακραίων τροχιών. Σύμφωνα με αυτή τη μέθοδο, επιλέγεται μια ακολουθία τροχιών σαν άξονας που παίζουν το ρόλο του κέντρου της συστάδας. Κάθε μία επιλέγεται ως η πιο μακρινή από τον προηγούμενο άξονα. Έτσι δημιουργείται μια συστάδα ακριβώς *k* τροχιών γύρω από κάθε άξονα, με τους *k*-1 πλησιέστερους γείτονες.
3. Μετατροπή Χώρου *(Spatial Translation)*: κάθε συστάδα μετασχηματίζεται σε ένα σύνολο *(k, δ)*-ανωνυμίας. Η διαδικασία αυτή συμπεριλαμβάνει τη μετάβαση κάποιων σημείων της τροχιάς από την αρχική τους θέση σε μια άλλη. Στόχος είναι να επιτευχθεί *(k, δ)*-ανωνυμία, διατηρώντας παράλληλα τις πρωτότυπες τροχιές αλλά και εκείνες που έχουν μετατραπεί, όσο το δυνατόν παρόμοιες.

Το κύριο μειονέκτημα του *NWA* φαίνεται να είναι το γεγονός ότι χρησιμοποιεί τη λειτουργία της Ευκλείδειας απόστασης, καθώς είναι ευαίσθητη σε ακραίες τιμές και στις εναλλαγές της εκάστοτε τοπικής ώρας, ενώ μπορεί να εφαρμοστεί μόνο σε τροχιές που έχουν το ίδιο μέγεθος.

Η μέθοδος *W4M* είναι μια παραλλαγή του *NWA* που χρησιμοποιεί το *EDR*, ένα μέτρο απόστασης με ανοχή στο χρόνο διατηρώντας τη τεχνική της ομαδοποίησης αμετάβλητη. Το σημαντικότερο όφελος είναι ότι η χρήση του *EDR* καταργεί την ανάγκη της ομαδοποίησης των τροχιών σε ισοδύναμες κλάσεις του ίδιου χρονικού διαστήματος πριν το βήμα της ομαδοποίησης. Με λίγα λόγια, το βήμα της προ-επεξεργασίας του *NWA* δεν είναι πλέον απαραίτητο.

Η τεχνική της άπληστης ομαδοποίησης για ανωνυμία δεδομένων τροχιάς *(The greedy clustering-based approach to anonymize trajectory data)*. Σύμφωνα με τους *Mahdavifar et al.* το 2012, παρόλο που έχει αποδειχθεί ότι η εύρεση της βέλτιστης λύσης για την *k*-ανωνυμία είναι πρόβλημα *NP-Hard* (μια ειδική κατηγορία προβλημάτων), παραμένει ένα πρότυπο μοντέλο για την πρόληψη παραβιάσεων της ιδιωτικότητας των σχεσιακών δεδομένων. Όμως, σε περίπτωση δεδομένων τροχιάς η φυσική πολυπλοκότητα των χώρο-χρονικών τροχιών και η εξάρτηση των διαδοχικών σημείων, κάνουν το πρόβλημα της ανωνυμίας ακόμα δυσκολότερο.

Η τεχνική αυτή αποτελείται από δύο κύριες φάσεις :

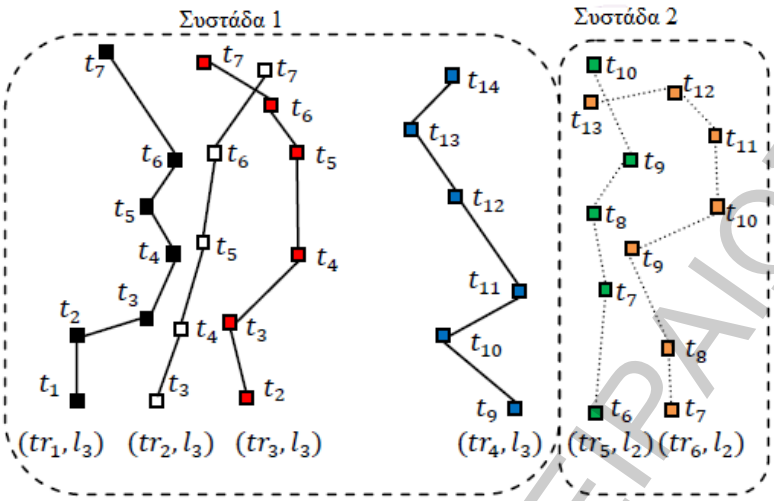
1. Ομαδοποίηση των τροχιών με βάση τα επίπεδα της ιδιωτικότητάς τους.
2. Ανωνυμία των τροχιών εντός κάθε συστάδας με μία ενιαία τροχιά.

Όσον αφορά την πρώτη φάση υπάρχουν 3 στρατηγικές:

Στην πρώτη στρατηγική (σχήμα 2.14), οι τροχιές πρώτα κατηγοριοποιούνται σε ομάδες με βάση τα επίπεδα της ιδιωτικότητας και στη συνέχεια δημιουργούνται μερικές συστάδες για κάθε ομάδα ανεξάρτητα από τις άλλες. Η στρατηγική αυτή όμως δεν είναι η πιο αποτελεσματική. Αυτό συμβαίνει επειδή τροχιές μέσα σε μια ενιαία συστάδα μπορεί να απέχουν από κάθε άλλη, παρόλο που έχουν το ίδιο επίπεδο προστασίας ιδιωτικότητας.

**Σχήμα 2.14**

Ομαδοποίηση τροχιών με χρήση της 1ης στρατηγικής για την τεχνική της άπληστης ομαδοποίησης

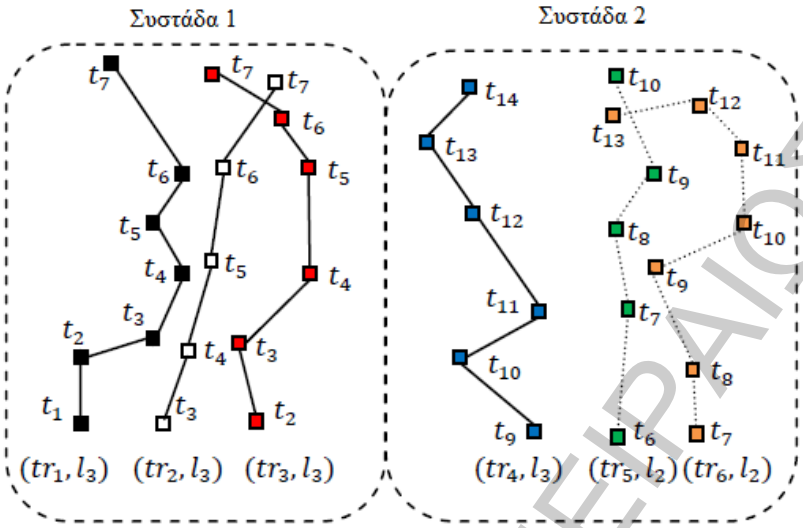



Η δεύτερη στρατηγική, αποσκοπεί στην απομόνωση των τροχιών σε συστάδες με μέγεθος τουλάχιστον ίσο με $l_{max}$, χωρίς να λαμβάνονται υπόψη τα επίπεδα της ιδιωτικότητας. Για το σκοπό αυτό, αρχικά επιλέγεται τυχαία μια τροχιά ως κέντρο βάρους της συστάδας και στη συνέχεια τουλάχιστον $l_{max}$-1 τροχιές που βρίσκονται πλησιέστερα, προστίθενται στη συστάδα. Αυτή η διαδικασία συνεχίζεται έως ότου όλες οι τροχιές μπουν στις ομάδες. Στην πραγματικότητα, αυτή η στρατηγική έχει σαν σκοπό κάθε τροχιά να τοποθετείται σε μια συστάδα με τουλάχιστον $l_{max}$-1 άλλες τροχιές, ανεξάρτητα από το επίπεδο προστασίας της ιδιωτικότητας. Στην περίπτωση της ανωνυμίας των συστάδων, αυτή η στρατηγική αυξάνει την ποσότητα της απώλειας πληροφοριών, ιδίως σε τροχιές με χαμηλά επίπεδα προστασίας της ιδιωτικότητας.

Στην τρίτη στρατηγική (σχήμα 2.15), αρχικά οι τροχιές ταξινομούνται σε ομάδες με βάση το επίπεδο της ιδιωτικότητας. Η ομαδοποίηση ξεκινάει από την ομάδα με το μέγιστο επίπεδο προστασίας της ιδιωτικότητας, με φθίνουσα σειρά, καθώς η διατήρηση της ιδιωτικότητας των τροχιών με υψηλότερα επίπεδα έχει προτεραιότητα. Οι εν λόγω ομάδες δημιουργούνται έτσι ώστε κάθε μέγεθος συστάδας να είναι ανάλογο με το επίπεδο προστασίας της ιδιωτικότητας του κέντρου βάρους της και η απόσταση της κάθε τροχιάς από το κέντρο βάρους της συστάδας να είναι μικρότερη από ένα ορισμένο όριο. Ως εκ τούτου, η στρατηγική αυτή έχει ως αποτέλεσμα την σημαντικά χαμηλότερη απώλεια πληροφοριών σε σχέση με τις άλλες δύο στρατηγικές.

**Σχήμα 2.15**

Ομαδοποίηση τροχιών με χρήση της 3ης στρατηγικής για την τεχνική της άπληστης ομαδοποίησης

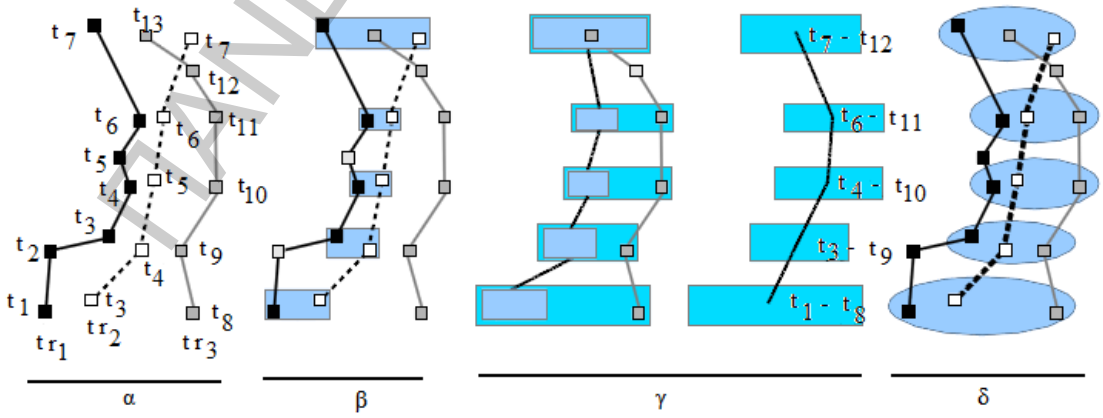



### 2.3.1.3 Προσεγγίσεις με βάση τη γενίκευση *(Generalization-based approaches)*

Μία τέτοια προσέγγιση είναι η τεχνική *Always Walk With Others-AWO* των *Nergiz et al.* το 2008. Η τεχνική αυτή, απαιτεί κατηγοριοποίηση τροχιών σε ομάδες μεγέθους μεγαλύτερου από *k* με βάση κάποιο μέτρο ομοιότητας, έχοντας σαν στόχο τη βέλτιστη ανωνυμία. Υπολογίζει ένα αντίστοιχο σημείο μεταξύ των σημείων των ζευγών των τροχιών που έχουν κατηγοριοποιηθεί. Τα σημεία που ταιριάζουν, αντικαθίστανται από τα αντίστοιχα *MBR*.

**Σχήμα 2.16**

Διαδικασία Ανωνυμίας της τεχνικής *AWO*



Στο σχήμα 2.16, όπως βλέπουμε στο βήμα α έχουμε 3 τροχιές. Στο βήμα β δημιουργείται ένα σύνολο ανωνυμίας (tr*) από τις δύο πρώτες τροχιές. Στη συνέχεια, στο βήμα γ στο σύνολο ανωνυμίας tr* προστίθεται και η τρίτη τροχιά. Στο βήμα δ, δημιουργείται ένα νέο σύνολο ανωνυμίας που συμπεριλαμβάνει και τις τρεις τροχιές.

Τέλος, πρέπει να επισημάνουμε ότι σε αυτή την προσέγγιση προστατεύεται η ιδιωτικότητα των ατόμων από τον αντίπαλο χρησιμοποιώντας τις κάτωθι τεχνικές:

- *k*-ανωνυμία *(k-anonymity)*: ανωνυμία του συνόλου δεδομένων, έτσι ώστε κάθε τροχιά να είναι δυσδιάκριτη από *k* - 1 άλλες τροχιές. Έχει την δυνατότητα να περιορίζει την ικανότητα του αντιπάλου να συνδέσει όλες τις πληροφορίες ενός ατόμου.
- Ανασυγκρότηση *(reconstruction)*: απελευθέρωση τυχαία επιλεγμένων ατομικών τροχιών από την περιοχή που καλύπτεται από ανώνυμες τροχιές. Εμποδίζει την περαιτέρω διαρροή προσωπικών πληροφοριών των χρηστών λόγω της ανωνυμίας.

Μία ακόμα προσέγγιση για την ανωνυμία κινούμενων δεδομένων μέσω γενίκευσης είναι η μέθοδος που συνδυάζει τις έννοιες της γενίκευσης χώρου και της *k*-ανωνυμίας (*Monreale et al.*, 2010). Η βασική ιδέα αυτής της τεχνικής είναι η απόκρυψη τοποθεσιών με μέσα γενίκευσης, αντικαθιστώντας ακριβείς θέσεις των τροχιών με θέσεις κατά προσέγγιση. Το σχέδιο ανωνυμίας γίνεται με βάση: (α) τη δημιουργία ενός διαμερίσματος σε περιοχές της επικράτειας που καλύπτεται από τις τροχιές, (β) την εφαρμογή μιας λειτουργίας για τη χωρική γενίκευση σε όλες οι τροχιές προκειμένου να τις μετατρέψει σε ακολουθίες των σημείων που είναι κέντρα βάρους των συγκεκριμένων περιοχών, (γ) τη μετατροπή των γενικευμένων τροχιών σε εγγυημένη προστασία ιδιωτικότητας.

❖ Ορισμός 2.5

Γενικευμένη τροχιά *(Generalized trajectory)*. Ας είναι $T=\langle x_1, y_1, t_1\rangle,....,\langle x_n, y_n, t_n\rangle$ μια τροχιά. Μια γενικευμένη εκδοχή της T είναι μια ακολουθία ζευγών $T_g=\langle x_{c1}, y_{c1}\rangle,....,\langle x_{cm}, y_{cm}\rangle$, με $m \leq n$. Τα $x_{ci}, y_{ci}$ είναι τα κέντρα βάρους της περιοχής που καλύπτεται από την T.

Παρακάτω, με ένα απλό παράδειγμα θα εξηγήσουμε τη λειτουργία της μεθόδου, προκειμένου να λάβει υπόψη τη χωρική γενίκευση αλλά και το όριο *k* της *k*-ανωνυμίας. Όσον αφορά τη χωρική γενίκευση, αν α και β είναι δύο περιοχές που πρέπει να συνδεθούν και *C* είναι το σύνολο από τα κέντρα βάρους όλων των περιοχών, τότε $c_a \in C$ είναι το κέντρο βάρους του α και $c_\beta \in C$ είναι το κέντρο βάρους του β. Στη συνέχεια, εξάγονται όλα τα

σημεία της τροχιάς που περιέχονται στα α και β και υπολογίζεται το $c_{\alpha+\beta}$ σαν το κέντρο βάρους όλων αυτών των σημείων. Τέλος αφαιρούνται από το *C* τα $c_a$, $c_\beta$ και προστίθεται το $c_{\alpha+\beta}$. Η διαδικασία αυτή ονομάζεται προοδευτική γενίκευση *(progressive generalization).* Όσον αφορά την ικανοποίηση του ορίου της *k*-ανωνυμίας, υπάρχουν δύο παραλλαγές μιας μεθόδου, οι *KAM_CUT* και *KAM_REC* οι οποίες μετατρέπουν τα γενικευμένα σύνολα δεδομένων που τους ζητούνται, σε *k*-ανώνυμα.

#### 2.3.1.4 Προστασία δεδομένων κίνησης με απόκρυψη μονοπατιού *(Path Cloaking)*

Σε αυτό το υποκεφάλαιο, παρουσιάζεται η μέθοδος της απόκρυψης μονοπατιού (*Hoh et al.*, 2007) που ανήκει στις προσεγγίσεις της *k*-ανωνυμίας. Σαν στόχο έχει διατήρηση ιδιωτικότητας σε ίχνη *GPS*, η οποία μπορεί να εγγυηθεί καλό επίπεδο ιδιωτικότητας ακόμα και σε χρήστες που οδηγούν σε περιοχές χαμηλής πυκνότητας.

Ο αλγόριθμος αυτής της μεθόδου ακολουθεί τα παρακάτω βήματα: Πρώτα, προσδιορίζει τα οχήματα που μπορούν να αποκαλυφθούν, επειδή έχει περάσει λιγότερος χρόνος από την τιμή του μέγιστου χρόνου σύγχυσης από το τελευταίο σημείο σύγχυσης. Στη συνέχεια, προσδιορίζει ένα σύνολο οχημάτων που μπορούν να αποκαλυφθούν επειδή η τρέχουσα αβεβαιότητα παρακολούθησης είναι μεγαλύτερη από το όριο αβεβαιότητας. Τέλος, ενημερώνει τον χρόνο του τελευταίου σημείου σύγχυσης και του τελευταίου ορατού δείγματος *GPS* για κάθε όχημα. Αυτό το στάδιο, μπορεί να πραγματοποιηθεί μόνο όταν το σύνολο των δειγμάτων που αποκάλυψε το *GPS* έχει αποφασιστεί, καθώς η σύγχυση θα πρέπει να υπολογίζεται με βάση τα αποκαλυφθέντα δείγματα.

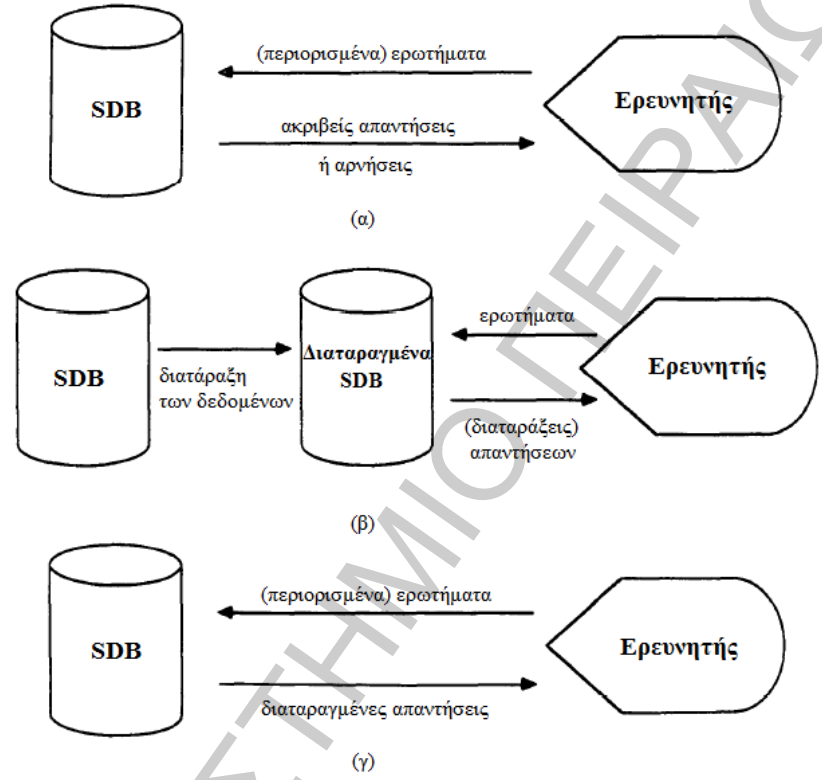
### 2.3.2 Προστασία Ιδιωτικότητας σε Επερωτήσεις Δεδομένων Κίνησης *(Privacy-Aware Mobility Data Querying)*

#### 2.3.2.1 Μέθοδοι ελέγχου ασφαλείας για στατιστικές βάσεις δεδομένων

Οι μέθοδοι ελέγχου ασφαλείας που προτείνονται στη βιβλιογραφία κατατάσσονται σε τέσσερις γενικές προσεγγίσεις: την εννοιολογική, τον περιορισμό ερωτήσεων, τη διατάραξη των στοιχείων και τη διατάραξη της εξόδου.

Η στατιστική βάση δεδομένων *(SDB)* σύμφωνα με τους *Adam & Wortmann* το 1989, είναι ένα σύστημα της βάσης δεδομένων που επιτρέπει στους χρήστες του να ανακτήσουν μόνο τα συγκεντρωτικά στατιστικά στοιχεία (π.χ. μέσο δείγματος) για ένα υποσύνολο των δεδομένων που εκπροσωπούνται στη βάση δεδομένων (σχήμα 2.17).

**Σχήμα 2.17**
(α) περιορισμός ερωτημάτων, (β) διατάραξη δεδομένων, (γ) διατάραξη εξόδου




### 2.3.2.1.1 Τύποι στατιστικών δεδομένων και συστημάτων υπολογιστών

Υπάρχουν οι παρακάτω κατηγορίες:

- Αποσυνδεδεμένη-Συνδεδεμένη. Με συνδεδεμένη *SDB*, υπάρχει άμεση και σε πραγματικό χρόνο αλληλεπίδραση του χρήστη με τα δεδομένα μέσω ενός τερματικού. Με αποσυνδεδεμένη *SDB*, ο χρήστης δεν έχει τον έλεγχο της επεξεργασίας των δεδομένων αλλά ούτε ξέρει και πότε εκτελείται το αίτημά της επεξεργασίας των δεδομένων.
- Στατική-Δυναμική. Η στατική βάση δεδομένων είναι αυτή που δεν αλλάζει ποτέ αφότου έχει δημιουργηθεί. Οι περισσότερες βάσεις απογραφής δεδομένων είναι στατικές. Κάθε φορά που μια νέα έκδοση της βάσης δεδομένων δημιουργείται,

θεωρείται ότι είναι μια ακόμα στατική βάση δεδομένων. Αντίθετα, οι δυναμικές βάσεις δεδομένων μπορεί να αλλάζουν συνεχώς.

- Κεντρική-Αποκεντρωμένη. Σε μια κεντρική *SDB* υπάρχει μία βάση δεδομένων. Σε μια αποκεντρωμένη *SDB*, συσσωρευμένα υποσύνολα της βάσης δεδομένων αποθηκεύονται σε διαφορετικές τοποθεσίες που συνδέονται με ένα δίκτυο επικοινωνίας.
- Αποκλειστική-*SCS* (Κοινόχρηστο Σύστημα Ηλεκτρονικών Υπολογιστών). Σε μια αποκλειστική *SDB*, το σύστημα υπολογιστών χρησιμοποιείται αποκλειστικά για να εξυπηρετήσει τις εφαρμογές της *SDB*. Σε ένα κοινόχρηστο σύστημα, οι εφαρμογές της *SDB* λειτουργούν με το ίδιο σύστημα υλικού άλλων εφαρμογών. Το κοινόχρηστο περιβάλλον είναι πιο δύσκολο να προστατευτεί, αφού άλλες εφαρμογές μπορεί να είναι σε θέση να παρέμβουν στα προστατευμένα δεδομένα απευθείας μέσω του συστήματος λειτουργίας, παρακάμπτοντας τον μηχανισμό ασφαλείας της *SDB*.

### 2.3.2.2 Τεχνική βασισμένη στη μηχανή αναζήτησης *Hermes ++*

Ο *Hermes++ (Pelekis et al.,* 2011) είναι μια μηχανή αναζήτησης για τροχιές κινούμενων αντικειμένων, που επιτρέπει στους εγγεγραμμένους χρήστες να έχουν περιορισμένη πρόσβαση στη βάση δεδομένων, καθώς τα δεδομένα μένουν εντός του οργανισμού που τα φιλοξενεί *(in-house).*

Η μηχανή αυτή:

(i) ελέγχει ερωτήματα σε δεδομένα τροχιών κινούμενων αντικειμένων, ώστε να εμποδίσει πιθανές επιθέσεις στην ιδιωτικότητα των χρηστών,

(ii) υποστηρίζει ερωτήματα: εύρους, απόστασης, k-πλησιέστερων γειτόνων, για χωρικά και χωροχρονικά δεδομένα,

(iii) διατηρεί την ανωνυμία των χρηστών σε απαντήσεις ερωτημάτων δημιουργώντας ένα σύνολο προσεκτικά σχεδιασμένων, ρεαλιστικών, ψεύτικων τροχιών (σχήμα 2.18).

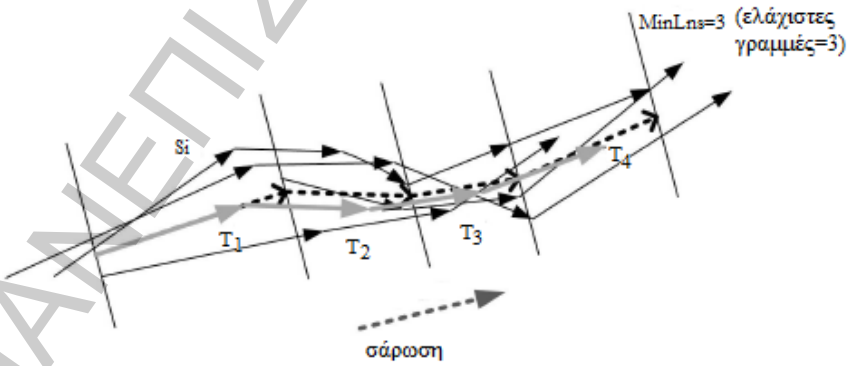
Επιπλέον, ο *Hermes++* μπορεί να προστατεύσει αποτελεσματικά την ιδιωτικότητα των χρηστών, αντιμετωπίζοντας τρεις τύπους επιθέσεων που οι κακόβουλοι χρήστες μπορεί να προσπαθήσουν να εφαρμόσουν στην αρχική βάση δεδομένων:

• Επίθεση αναγνώρισης χρήστη *(User identification attack)*: Σε αυτή την επίθεση, η ταυτότητα του χρήστη μπορεί να εκτεθεί από ερωτήματα που αφορούν περιοχές που αλληλοκαλύπτονται χωροχρονικά. Για να αποφευχθεί, γίνεται έλεγχος ερωτημάτων, που ξεκινούν από κάθε τελικό χρήστη στο σύστημα και δεν δίνονται απαντήσεις σε επικαλυπτόμενα ερωτήματα.

• Επίθεση εντοπισμού ευαίσθητης θέσης *(Sensitive location tracking attack)*: Σε αυτή την επίθεση, ο κακόβουλος χρήστης προσπαθεί να συνδέσει στο χάρτη μία ή περισσότερες θέσεις της τροχιάς του χρήστη σε γνωστές τοποθεσίες, που μπορεί να εκθέσουν αποτελεσματικά την ταυτότητά του. Για να αποτραπούν αυτές οι επιθέσεις, προστατεύεται η αρχική και η τελική θέση των τροχιών, καθώς και οποιαδήποτε άλλη θέση στην διάρκεια της πορείας των χρηστών που μπορεί να θεωρηθεί ως ευαίσθητη για τον χρήστη.

• Επίθεση διαδοχικής παρακολούθησης *(Sequential tracking attack)*: Ο χρήστης έχει εντοπιστεί μέσω της τροχιάς του από ένα σύνολο ερωτήσεων σε περιοχές που βρίσκονται κοντά η μία στην άλλη από άποψη χώρου και χρόνου. Ο εισβολέας μπορεί να «ακολουθήσει» το χρήστη και να μάθει τις θέσεις που έχει επισκεφθεί. Για να σταματήσει αυτή την επίθεση, ο προτεινόμενος αλγόριθμος ελέγχου λαμβάνει τα αναγκαία μέτρα για να συνεχίσει ομαλά η κίνηση των ψεύτικων τροχιών από τις γειτονικές περιοχές προς την τρέχουσα. Σκοπός είναι να απαγορευθεί στους εισβολείς να διακρίνουν τις ψεύτικες τροχιές από τις πραγματικές.

**Σχήμα 2.18**
Ο αλγόριθμος παραγωγής ψεύτικης τροχιάς



## 2.4 Ανωνυμία σε σημασιολογικά ευαίσθητες τοποθεσίες

### 2.4.1 *LBS* για σημασιολογικά ευαίσθητες τοποθεσίες

Υπάρχουν κάποιοι κακόβουλοι χρήστες, που γνωρίζουν τη σημασιολογική τοποθεσία μιας περιοχής. Σε τέτοιες περιπτώσεις, η περαιτέρω σημασιολογική γνώση θα μπορούσε να οδηγήσει σε αποκάλυψη προσωπικών πληροφοριών. Οι παρακάτω μέθοδοι έχουν σαν στόχο την προστασία της θέσης ενός χρήστη που κάνει αίτημα για *LBS*, όταν σταματάει σε ευαίσθητες τοποθεσίες.

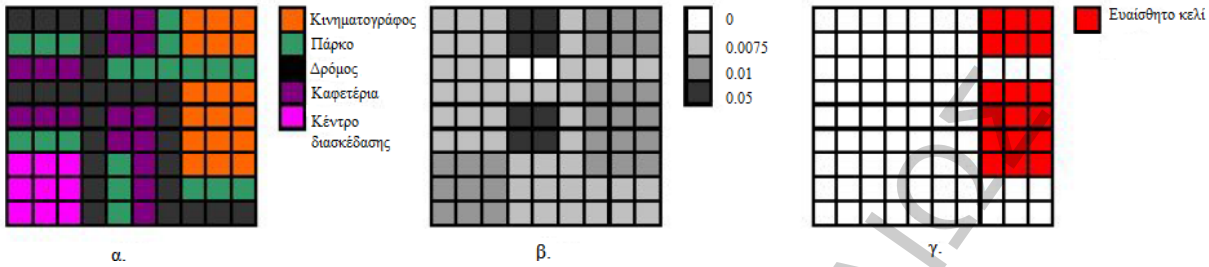
#### 2.4.1.1 Η τεχνική της απόκρυψης θέσεων με σημασιολογική πληροφορία *(Cloaking of semantic locations)*

Για την ανάλυση αυτής της τεχνικής, ξεκινάμε εξηγώντας την έννοια των κρυμμένων περιοχών *(CRs).* Οι αλγόριθμοι απόκρυψης τοποθεσίας, τις παράγουν με σκοπό να καλύπτουν ευαίσθητες θέσεις καθώς ικανοποιούν τις προτιμήσεις του προφίλ των χρηστών για την ιδιωτικότητά τους, ανεξάρτητα από την πραγματική θέση του χρήστη.

Η τεχνική της απόκρυψης θέσεων με σημασιολογική πληροφορία *(Damiani et al.*, 2011), υποθέτει ότι για κάθε χρήστη οι ευαίσθητες τοποθεσίες είναι διαφορετικές. Η διαδικασία που ακολουθείται είναι η εξής: η περιοχή που αφορά το χρήστη χωρίζεται σε κελιά ίσου μεγέθους, καθένα από τα οποία έχει την αντίστοιχη πιθανότητά του (με άθροισμα πιθανοτήτων το 1). Αν κάποιο κελί βρίσκεται σε μη προσβάσιμη τοποθεσία, όπως για παράδειγμα σε λίμνη που δεν επιτρέπονται οι βάρκες, η πιθανότητά του είναι 0 (Σχήμα 2.19).

**Σχήμα 2.19**

(α) Πλέγμα τοποθεσίες με σημασιολογική πληροφορία στον πραγματικό κόσμο, (β) Πλέγμα με τις πιθανότητες τοποθεσιών με σημασιολογική πληροφορία, (γ) Πλέγμα με τις ευαίσθητες πληροφορίες




Κάθε χρήστης έχει τη δυνατότητα να καθορίσει το δικό του προφίλ ιδιωτικότητας λαμβάνοντας υπόψη ποια από τα σημεία ενδιαφέροντος (*POIs)* είναι ευαίσθητα και τον αντίστοιχο βαθμό της ιδιωτικότητας. Το σύνολο των κρυπτογραφημένων περιοχών, προστατεύει τις ευαίσθητες περιοχές και διαμορφώνεται σύμφωνα με τα προφίλ ιδιωτικότητας του κάθε χρήστη και με βάση τις γνώσεις των κακόβουλων χρηστών.

Θα μπορούσαμε να πούμε ότι συνοπτικά όλη η διαδικασία αποτελείται από 3 βήματα:

1. Καθορισμός του προφίλ ιδιωτικότητας κάθε χρήστη.
2. Απόκρυψη των σημασιολογικών τοποθεσιών.
3. Μετασχηματισμός θέσης.

Τελικά, όταν ένας χρήστης κάνει αίτημα για *LBS* εξετάζεται η θέση του. Αν αυτή αντιστοιχεί σε μία *CR* τότε επιστρέφεται σαν απάντηση αυτή, αλλιώς επιστρέφεται η πραγματική θέση του χρήστη.

### 2.4.1.2 Η επέκταση της τεχνικής απόκρυψης θέσεων με σημασιολογική πληροφορία *(The extension of cloaking of semantic locations)*

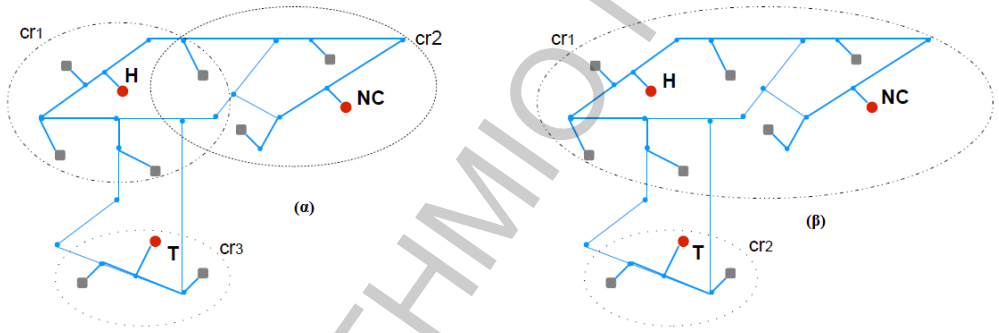
Η επέκταση του μοντέλου της απόκρυψης θέσεων με σημασιολογική πληροφορία (*Yigitoglu et al.*, 2012), γίνεται στο πλαίσιο των οδικών δικτύων και λαμβάνεται υπόψη η επίθεση που βασίζεται σε γνώση της ταχύτητας. Η συχνή ενημέρωση της θέσης στην οποία βρίσκονται οι χρήστες, βοηθούν τους κακόβουλους χρήστες να ανακαλύψουν την ταχύτητά τους. Εκείνοι με τη σειρά τους, με την αξιοποίηση αυτών των πληροφοριών έχουν τη δυνατότητα να προβλέψουν τις επόμενες τοποθεσίες που θα επισκεφτούν οι χρήστες μέσα σε μία *CR*.

Σε αυτή την τεχνική προτείνονται δύο είδη αλγορίθμων:

1. Απόκρυψη εκτός σύνδεσης *(offline cloaking)*. Λειτουργεί σε δύο στάδια: (i) στατική απόκρυψη εκτός σύνδεσης ευαίσθητων περιοχών και (ii) μετασχηματισμός εντός σύνδεσης, που εξασφαλίζει να μην υπάρχει καμία παραβίαση της ιδιωτικότητας από επιθέσεις ταχύτητας. Θεωρούνται δύο μέθοδοι απόκρυψης: η ασυνεχής *(disjoint)* και η επικαλυπτόμενη *(overlapping)*. Η ασυνεχής απόκρυψη, δεν επιτρέπει κανένα κοινό χαρακτηριστικό μεταξύ των κρυμμένων περιοχών, αλλά επιτρέπει περισσότερες από μία ευαίσθητες τοποθεσίες να βρίσκονται μαζί σε μια ενιαία κρυμμένη περιοχή. Από την άλλη, η επικαλυπτόμενη απόκρυψη επιτρέπει επικαλύψεις μεταξύ κρυμμένων περιοχών και εκχωρεί μόνο ένα ευαίσθητο μέρος ανά περιοχή (σχήμα 2.20).

**Σχήμα 2.20**

(α) επικαλυπτόμενη απόκρυψη *(overlapping cloaking)*, (β) ασυνεχής απόκρυψη *(disjoint cloaking)*




2. Απόκρυψη εντός σύνδεσης *(online cloaking)*. Θεωρείται πιο κατάλληλη για τις περιπτώσεις όπου η πλειοψηφία των θέσεων δεν είναι στατική και αλλάζει σημαντικά ανάλογα με την ώρα της ημέρας, την ημέρα της εβδομάδας και ούτω καθεξής.

### 2.4.2 Προστασία ιδιωτικότητας σε τροχιές με σημασιολογική πληροφορία

Η αύξηση των δεδομένων σχετικά με τις τροχιές των προσωπικών μετακινήσεων των ανθρώπων, ανοίγει νέες ευκαιρίες για την ανάλυση και την εξόρυξη της ανθρώπινης κινητικότητας. Ωστόσο, προκύπτουν νέοι κίνδυνοι παρείσφρησης στην ιδιωτική ζωή. Παρακάτω παρουσιάζεται μία μέθοδος για την προστασία ιδιωτικότητας σε σημασιολογικές τροχιές.

### 2.4.2.1 Η τεχνική *C*-ασφάλεια (C-safety)

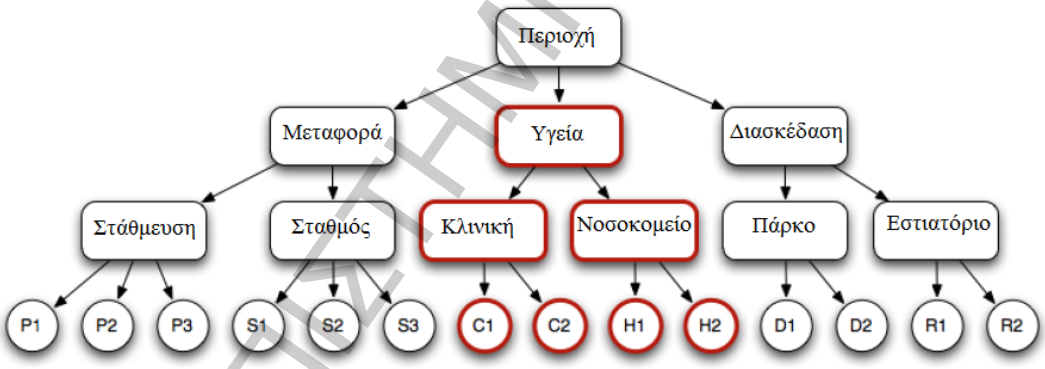
Ξεκινάμε την ανάλυση της μεθόδου της *C*-ασφάλειας (*Monreale et al.*, 2011) παρουσιάζοντας κάποιες βασικές έννοιες που είναι χρήσιμες για την κατανόηση της ιδέας της τεχνικής.

Ευαίσθητη τοποθεσία *(Sensitive place)*: καλείται μια τοποθεσία όταν επιτρέπει να συναχθούν ευαίσθητες προσωπικές πληροφορίες του ατόμου που εντοπίζεται.

Οντολογία *(Ontology)*: είναι ένας τεχνικός όρος που δηλώνει ένα αντικείμενο σχεδιασμένο με σκοπό να καταστεί δυνατή η μοντελοποίηση της γνώσης αναφορικά με κάποιο πραγματικό ή φανταστικό τομέα. Μια οντολογία καθορίζει τι μπορεί να εκπροσωπείται και τι μπορεί να συναχθεί σχετικά με ένα τομέα, χρησιμοποιώντας ένα συγκεκριμένο τύπο εννοιών.

Η ταξινόμηση *(Taxonomy)* των θέσεων ενδιαφέροντος αποτελεί την σημασιολογική ιεραρχία γεωγραφικών σημείων ενδιαφέροντος (σχήμα 2.21).

**Σχήμα 2.21**
Παράδειγμα περιοχών ταξινόμησης




Η μέθοδος που προτείνει η *C*-ασφάλεια, βασίζεται στη γενίκευση. Εστιάζει στις πραγματικές θέσεις της διαδρομής κάποιου χρήστη, οι οποίες αντιπροσωπεύουν τις πιο ευαίσθητες πληροφορίες. Έτσι λοιπόν οι εσωτερικοί κόμβοι της ταξινόμησης, χρησιμοποιούνται για τη γενίκευση της τροχιάς με σημασιολογική πληροφορία, προκειμένου να μειωθούν οι πιθανότητες ώστε κάποιος κακόβουλος χρήστης να εντοπίσει τους πραγματικούς σταθμούς.

Η *C*-ασφάλεια, παρέχει ένα άνω φράγμα *C* στην πιθανότητα του συμπεράσματος ότι ένα συγκεκριμένο πρόσωπο που παρατηρήθηκε σε μια σειρά από μη-ευαίσθητα σημεία έχει επίσης σταματήσει σε οποιοδήποτε ευαίσθητο σημείο. Τα κύρια στάδια της μεθόδου είναι τα

εξής: (α) καταστολή κάθε ευαίσθητης τοποθεσίας από το σύνολο δεδομένων, όταν για το χρήστη αυτό το μέρος είναι ένα οιονεί-αναγνωριστικό, (β) ομαδοποίηση των τροχιών με σημασιολογική πληροφορία σε ομάδες με ένα προκαθορισμένο μέγεθος *m*, (γ) δημιουργία μιας γενικευμένης έκδοσης της κάθε τροχιάς με σημασιολογική πληροφορία του συνόλου, γενικεύοντας τις τοποθεσίες που είναι οιονεί-αναγνωριστικά. Σε κάθε σύνολο, τα οιονεί-αναγνωριστικά των γενικευμένων τροχιών πρέπει να είναι ταυτόσημα. Οι ευαίσθητες τοποθεσίες γενικεύονται, όταν η γενίκευση των οιονεί-αναγνωριστικών δεν είναι αρκετή για ένα *C*-ασφαλές σύνολο δεδομένων.

# ΚΕΦΑΛΑΙΟ 3

## Σχεδιασμένη Προστασία της Ιδιωτικότητας *(Privacy by Design)*

### 3.1 Η ιδέα

Το πρόβλημα προστασίας προσωπικών δεδομένων, απαιτεί σε γενικές γραμμές, να βρούμε τον καλύτερο συμβιβασμό μεταξύ της ιδιωτικότητας και της χρησιμότητας των δεδομένων. Η Σχεδιασμένη Προστασία της Ιδιωτικότητας στην εξόρυξη δεδομένων, είναι ένας όρος που επινοήθηκε στη δεκαετία του 1990 από την *Ann Cavoukian,* Επίτροπο Πληροφοριών και Προστασίας Προσωπικών Δεδομένων του Οντάριο, στον Καναδά. Στον τομέα της έρευνας της ανάλυσης των δεδομένων για τη διατήρηση της ιδιωτικότητας, είναι ένα πρόσφατο παράδειγμα που υπόσχεται ένα ποιοτικό άλμα στη σύγκρουση μεταξύ της προστασίας και της χρησιμότητας των δεδομένων. Ασχολείται με την προσέγγιση της ενσωμάτωσης των προσωπικών δεδομένων στο σχεδιασμό, στη λειτουργία και στη διαχείριση των τεχνολογιών που επεξεργάζονται πληροφορίες και συστήματα. Η γενική αρχή, είναι ότι μπορεί να επιτευχθεί υψηλότερη προστασία και ποιότητα σε μια προσέγγιση προσανατολισμένη στο στόχο.

Σε μια τέτοια προσέγγιση, η διαδικασία ανακάλυψης γνώσης (συμπεριλαμβανομένης της συλλογής δεδομένων), έχει σχεδιαστεί με υποθέσεις *(Pedreschi et al.,* 2012*)* σχετικά με:

- Τα (ευαίσθητα) προσωπικά δεδομένα που είναι το αντικείμενο της ανάλυσης.
- Το μοντέλο επίθεσης. Τις γνώσεις και το σκοπό ενός κακόβουλου μέλους που θέλει να ανακαλύψει τα ευαίσθητα δεδομένα κάποιων συγκεκριμένων ατόμων.
- Τις στοχευμένες αναλυτικές ερωτήσεις που πρέπει να απαντηθούν με τα δεδομένα.

Αυτές οι υποθέσεις είναι θεμελιώδους σημασίας για το σχεδιασμό του πλαισίου της διατήρησης της προστασίας της ιδιωτικότητας, για τους παρακάτω λόγους :

Αρχικά, οι τεχνικές για τη διατήρηση της ιδιωτικότητας εξαρτώνται ιδιαίτερα από την φύση των δεδομένων που θέλουμε να προστατεύσουμε. Επιπλέον, ένα έγκυρο πλαίσιο για την προστασία της ιδιωτικότητας, πρέπει να καθορίσει το γνωστικό υπόβαθρο του αντιπάλου, που εξαρτάται σε μεγάλο βαθμό από το περιβάλλον και από το είδος των δεδομένων. Τέλος,

μια τεχνική ανωνυμίας θα πρέπει να παρέχει μια αποδεκτή ανταλλαγή μεταξύ της προστασίας προσωπικών δεδομένων των χρηστών και της χρησιμότητας τους.

Είναι λοιπόν απαραίτητο, να σχεδιαστεί ένα πλαίσιο διαφύλαξης της ιδιωτικότητας που να είναι σε θέση να μετατρέψει τα δεδομένα σε μια ανώνυμη έκδοση με ποσοτική εγγύηση προστασίας της ιδιωτικότητας. Επιπλέον, να εγγυάται ότι οι στοχευμένες αναλυτικές ερωτήσεις μπορούν να απαντηθούν σωστά, μέσα σε μια ποσοτική προσέγγιση που καθορίζει τη χρησιμότητα των δεδομένων με τη χρήση των μετασχηματισμένων δεδομένων αντί των αρχικών.

### 3.2 Οι 7 θεμελιώδεις αρχές

Για να θεωρείται μια τεχνική ή ένα σύστημα ότι πληροί την ιδέα της Σχεδιασμένης Προστασίας της Ιδιωτικότητας, πρέπει να ικανοποιεί τις παρακάτω θεμελιώδεις αρχές *(Cavoukian A., 2011)*:

1. Προ-ενεργητική όχι Αντιδραστική- Προφυλακτική όχι Διορθωτική *(Proactive not Reactive; Preventative not Remedial)*. Δεν πρέπει να περιμένει να υλοποιηθούν οι κίνδυνοι της προστασίας της ιδιωτικότητας, ούτε να προσφέρει λύσεις για τυχόν παραβιάσεις αφού έχουν συμβεί. Σκοπός της είναι να τις προλαβαίνει πριν συμβούν.
2. Η προστασία της ιδιωτικότητας να είναι προεπιλεγμένη ρύθμιση *(Privacy as the Default Setting)*. Τα δεδομένα πρέπει να προστατεύονται αυτόματα. Αν ένα άτομο δεν κάνει τίποτα, η προστασία της ιδιωτικότητάς του παραμένει άθικτη. Η συλλογή των προσωπικών πληροφοριών πρέπει να είναι δίκαιη, νόμιμη και να περιορίζεται στο ελάχιστο. Επιπλέον να γίνεται μόνο για καθορισμένους σκοπούς οι οποίοι θα πρέπει να είναι σαφείς, συγκεκριμένοι και σχετικοί με τις περιστάσεις. Τέλος, από πλευράς του ατόμου δεν πρέπει απαιτείται να γίνει καμία ενέργεια για την προστασία της ιδιωτικότητάς του.
3. Η ιδιωτικότητα να είναι ενσωματωμένη στο σχεδιασμό *(Privacy Embedded into Design)*. Η προστασία της ιδιωτικότητας πρέπει να αποτελεί ουσιώδες συστατικό της λειτουργικότητας του βασικού σκελετού της τεχνικής. Πρέπει δηλαδή, να είναι αναπόσπαστο μέρος του συστήματος χωρίς να μειώνει τη λειτουργικότητα.

4. Να υπάρχει πλήρης λειτουργικότητα *(Full Functionality- Positive- Sum, not Zero-Sum).* Ιδιωτικότητα εναντίον ασφάλειας, αποδεικνύοντας ότι είναι δυνατόν να συνυπάρχουν, χωρίς η μία να αποκλείει την άλλη.
5. Πλήρης προστασία: Πλήρης προστασία του κύκλου ζωής *(End-to-End Security - Full Lifecycle Protection).* Τα ισχυρά μέτρα προστασίας, είναι βασικά για την προφύλαξη της ιδιωτικότητας από την αρχή μέχρι το τέλος. Έτσι διασφαλίζεται ότι όλα τα δεδομένα κρατούνται με ασφάλεια και στη συνέχεια καταστρέφονται εγκαίρως και με ασφάλεια στο τέλος της διαδικασίας.
6. Να υπάρχει ορατότητα και διαφάνεια *(Visibility and Transparency - Keep it Open).* Τα συστατικά μέρη και οι λειτουργίες πρέπει να παραμένουν ορατά και διαφανή στους χρήστες αλλά και σε αυτούς που παρέχουν το σύστημα.
7. Να υπάρχει σεβασμός στην ιδιωτικότητα του χρήστη και να είναι ο χρήστης το κεντρικό πρόσωπο *(Respect for User Privacy - Keep it User-Centric).* Πρέπει λοιπόν να έχουν ισχυρές προεπιλογές προστασίας της ιδιωτικότητας, κατάλληλη ενημέρωση αλλά και δυνατές και φιλικές προς το χρήστη επιλογές.

Στις επόμενες παραγράφους, γίνεται έλεγχος για την τήρηση των τριών υποθέσεων της Σχεδιασμένης Προστασίας της Ιδιωτικότητας και των επτά θεμελιωδών αρχών. Η μόνη αρχή που δεν θα εξετασθεί, είναι η τέταρτη, αυτή της πλήρης λειτουργικότητας. Ο λόγος είναι ότι οι αρχές που θα σχολιάσουμε δεν ασχολούνται με θέματα ασφάλειας, παρά μόνο με την προστασία της ιδιωτικότητας των χρηστών. Έτσι λοιπόν δε μπορούν να προσαρμοστούν όλα τα νόμιμα συμφέροντα και οι στόχοι κατά τον τρόπο του θετικού αθροίσματος «κέρδισε-κέρδισε».

## 3.3 Έλεγχος τήρησης των τριών υποθέσεων και των επτά θεμελιωδών αρχών της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας

### 3.3.1 Προστασία Ιδιωτικότητας σε Υπηρεσίες Εντοπισμού θέσης *(Privacy in Location Based Services)*

Σε αυτή την κατηγορία θα ασχοληθούμε με δύο από τις πιο δημοφιλείς προσεγγίσεις της χωρικής *k*-ανωνυμίας: τον *Casper* (*Mokbel et al.*, 2006) και την απόκρυψη *Hilbert* (*Kalnis et al.*, 2007).

Α. Έλεγχος τήρησης των τριών υποθέσεων.

- Τα (ευαίσθητα) προσωπικά δεδομένα που είναι το αντικείμενο της ανάλυσης.

Και για τις δύο τεχνικές τα ευαίσθητα προσωπικά δεδομένα είναι το όνομα, η ταυτότητα καθώς και η ακριβής τοποθεσία των χρηστών.

- Το μοντέλο επίθεσης.

Στις τεχνικές αυτές, ο επιτιθέμενος θεωρείται ότι έχει τις εξής ικανότητες:

- ✓ Μπορεί να εμποδίσει την περιοχή στην οποία προσφέρεται ανωνυμία στον αιτούντα ενός *LBS*.
- ✓ Γνωρίζει τους αλγόριθμους που χρησιμοποιούνται από τον αξιόπιστο πάροχο για να προσφέρει προστασία ιδιωτικότητας στο *LBS*.
- ✓ Μπορεί να αποκομίσει όλες τις τρέχουσες τοποθεσίες των χρηστών του συστήματος.
- ✓ Προσπαθεί να σπάσει την ιδιωτικότητα της τοποθεσίας, έχοντας γνώση μόνο για τα τρέχοντα δεδομένα τοποθεσίας. Δε γνωρίζει την προϊστορία της κίνησης των χρηστών.

- Τις στοχευμένες αναλυτικές ερωτήσεις που πρέπει να απαντηθούν με τα δεδομένα.

- Η τεχνική *Casper*, απαντάει σε όλες τις συνήθεις ερωτήσεις με τις οποίες ασχολούνται οι υπηρεσίες εντοπισμού θέσης, όπως: ερωτήσεις εύρεσης καταστημάτων, ερωτήσεις που αφορούν την κίνηση στους δρόμους και ερωτήσεις με στόχο την αποστολή διαφημίσεων μέσω διαδικτύου (π.χ. στείλτε εκπτωτικά κουπόνια σε όλα τα αυτοκίνητα που είναι σε ακτίνα 100 μέτρων από το μαγαζί Α).

  Υποστηρίζει όμως και τρεις καινούριους τύπους ερωτήσεων:

  - ✓ Προσωπικές ερωτήσεις πάνω σε δημόσια δεδομένα (π.χ. Που είναι το πιο κοντινό βενζινάδικο).
  - ✓ Δημόσιες ερωτήσεις πάνω σε προσωπικά δεδομένα (π.χ. Πόσα αυτοκίνητα είναι στην περιοχή Α).
  - ✓ Προσωπικές ερωτήσεις πάνω σε προσωπικά δεδομένα (π.χ. Πού βρίσκεται ο φίλος μου ο Κώστας).

- Η τεχνική της απόκρυψης *Hilbert*, επικεντρώνεται σε ερωτήματα «στιγμιότυπων» *(snapshot queries)*, όπου ο επιτιθέμενος χρησιμοποιεί τα τρέχοντα στοιχεία και όχι πληροφορίες από το παρελθόν σχετικά με τις πρότυπες συμπεριφορές και κινήσεις των χρηστών.

Β. Έλεγχος τήρησης των 7 θεμελιωδών αρχών.

1. Προ-ενεργητική όχι Αντιδραστική- Προφυλακτική όχι Διορθωτική. Αυτή η αρχή ισχύει και για τις δύο τεχνικές.

Και για τις δύο τεχνικές ισχύει ότι εφόσον είναι προσεγγίσεις της *k*-ανωνυμίας, η ισχυρή ανωνυμία θεωρείται δεδομένη. Το κύριο χαρακτηριστικό τους είναι η προστασία ιδιωτικότητας με τη χρήση της τρέχουσας θέσης των χρηστών. Έτσι ενισχύεται η έννοια της πρόληψης καθώς χρησιμοποιούν πολύ πρόσφατα δεδομένα.

2. Η προστασία της ιδιωτικότητας να είναι προεπιλεγμένη ρύθμιση. Αυτή η αρχή ισχύει και στις δύο τεχνικές.

Η ύπαρξη του αξιόπιστου διακομιστή, εγγυάται την ασφαλή μεταφορά δεδομένων από τους χρήστες στον φορέα παροχής υπηρεσιών.

3. Η ιδιωτικότητα να είναι ενσωματωμένη στο σχεδιασμό. Αυτή η αρχή ισχύει και στις δύο τεχνικές.

- Όσον αφορά την τεχνική *Casper*, ολόκληρη η περιοχή που καλύπτεται από τον διακομιστή χωρίζεται με πλέγμα και οργανώνεται σε στρώματα με μια μορφή πυραμίδας. Στόχος αυτής της διαδικασίας είναι να ικανοποιηθεί η *k*-ανωνυμία και να προσφερθεί ενίσχυση της ιδιωτικότητας του κάθε χρήστη.
- Στην τεχνική της απόκρυψης *Hilbert*, η θέση του κάθε χρήστη χαρτογραφείται τρισδιάστατα. Θεωρείται ότι οι τοποθεσίες που βρίσκονται κοντά σε αυτή τη χαρτογράφηση θα είναι κοντά και σε μια μονοδιάστατη. Στόχος αυτής της διαδικασίας είναι να ικανοποιηθεί η *k*-ανωνυμία και να προσφερθεί ενίσχυση της ιδιωτικότητας του κάθε χρήστη.

5. Πλήρης Προστασία: Πλήρης προστασία του κύκλου ζωής. Αυτή η αρχή, υποστηρίζεται και από τις δύο τεχνικές που ανήκουν στην προστασία ιδιωτικότητας σε υπηρεσίες εντοπισμού θέσης.

Τα δεδομένα δημοσιεύονται ανώνυμα, συνεπώς υπάρχει προστασία της ιδιωτικότητας απέναντι στο ενδεχόμενο οι κακόβουλοι χρήστες να αποκομίσουν τις πληροφορίες που θέλουν.

6. Να υπάρχει ορατότητα και διαφάνεια.

Αυτή η αρχή ισχύει και για τις δύο τεχνικές. Τα βήματα των αλγορίθμων είναι σαφή, συνεπώς ο χρήστης μπορεί να καταλάβει απόλυτα τη διαδικασία που ακολουθούν αυτές οι τεχνικές.

7. Να υπάρχει σεβασμός στην ιδιωτικότητα του χρήστη και να είναι ο χρήστης το κεντρικό πρόσωπο.

Αυτή η αρχή ισχύει και στις δύο τεχνικές. Όπως προαναφέραμε κατά τον σχολιασμό της τρίτης αρχής ο διακομιστής παίζει τον βασικότερο ρόλο για την τήρηση της ιδιωτικότητας.

### 3.3.2 Προστασία Ιδιωτικότητας κατά τη Δημοσιοποίηση Δεδομένων Κίνησης *(Privacy-preserving Mobility Data Publishing)*

Παρακάτω θα εξετάσουμε τις εξής τεχνικές :

- Προσέγγιση διατάραξης: Προστασία Ιδιωτικότητας Τοποθεσίας μέσω Σύγχυσης Διαδρομής (*Hoh & Gruteser*, 2005).
- Προσεγγίσεις *k*-ανωνυμίας:
  - Μέθοδος ανωνυμίας με συνδυασμό *k*-ανωνυμίας και χωρικής γενίκευσης (*Monreale et al.*, 2010).
  - Προστασία δεδομένων κίνησης με απόκρυψη μονοπατιού (*Hoh et al.*, 2007).
  - Πιο πρόσφατες τεχνικές: *NWA* (*Abul et al.*, 2010).
    *W4M* (*Abul et al.*, 2010).
    *AWO* (*Nergiz et al.*, 2008).

Α. Έλεγχος τήρησης των τριών υποθέσεων.

- Τα (ευαίσθητα) προσωπικά δεδομένα που είναι το αντικείμενο της ανάλυσης.
  - Στην τεχνική της σύγχυσης διαδρομής ευαίσθητα δεδομένα θεωρούνται όλες οι πληροφορίες για τη θέση του κάθε χρήστη.
  - Στην τεχνική που συνδυάζει την *k*-ανωνυμία με τη χωρική γενίκευση, ευαίσθητα θεωρούνται τα δεδομένα θέσης που μπορούν να βοηθήσουν τον κακόβουλο χρήστη να μάθει πληροφορίες σχετικά με τις συνήθειες και τις προτιμήσεις άλλων χρηστών.
  - Στην τεχνική της απόκρυψης μονοπατιού που ασχολείται με ίχνη από *GPS*, ευαίσθητα δεδομένα θεωρούνται η κατάσταση υγείας, ο τρόπος ζωής, η καθαρή θέση, αλλά και

οι πολιτικές πεποιθήσεις κάποιου χρήστη. Επίσης όσον αφορά την οδική τους συμπεριφορά, ευαίσθητα δεδομένα θεωρούνται κάποιες παραβιάσεις κυκλοφορίας ή στάθμευσης.

- Οι τεχνικές *NWA* και *W4M*, θεωρούν ότι κάθε σημείο μπορεί να είναι ένα οιονεί-αναγνωριστικό.
- Στον *AWO* τα δεδομένα θεωρούνται ακίνητα. Όλες οι πληροφορίες θεωρούνται ευαίσθητες.

o Το μοντέλο επίθεσης.

- Στην τεχνική της σύγχυσης διαδρομής θεωρείται ότι τα δεδομένα μεταβιβάζονται σε φορείς παροχής υπηρεσιών, έτσι ώστε οι κακόβουλοι χρήστες να μην μπορούν να έχουν άμεση επιχειρηματική σχέση και να μην έχουν τη δυνατότητα επιλογής εναλλακτικών παρόχων.
- Στην τεχνική που συνδυάζει την *k*-ανωνυμία με τη χωρική γενίκευση έχει δοθεί ο παρακάτω ορισμός για την γνώση ενός κακόβουλου χρήστη.

❖ Ορισμός 3.1

Γνώση κακόβουλου χρήστη (τεχνική *k*-ανωνυμίας/ χωρικής γενίκευσης). Ο επιτιθέμενος έχει πρόσβαση στο σύνολο των ανώνυμων δεδομένων D* και ξέρει: (α) τις λεπτομέρειες των σχημάτων που χρησιμοποιούνται για την ανωνυμία των δεδομένων, (β) το γεγονός ότι ένας χρήστης U ανήκει στο σύνολο δεδομένων D και (γ) μία υπό-τροχιά S σχετική με τον U.

- Στην τεχνική της απόκρυψης μονοπατιού, κακόβουλος θεωρείται ένας χρήστης που μπορεί να ανακατασκευάσει την πορεία ενός ατόμου χρησιμοποιώντας τα ίχνη από το *GPS*. Δεν θεωρείται όμως ότι έχει προηγούμενη γνώση για κάποιον χρήστη.
- Οι τεχνικές *NWA* και *W4M* θεωρούν όλα τα χαρακτηριστικά οιονεί-αναγνωριστικά, συνεπώς κάθε πληροφορία μπορεί να προσφέρει γνώσεις στον κακόβουλο χρήστη.
- Στη τεχνική *AWO* θεωρείται ότι οι κακόβουλοι χρήστες:
  - ✓ μπορεί να γνωρίζουν ήδη κάποιο τμήμα της τροχιάς ενός ατόμου στο σύνολο δεδομένων και να ενδιαφέρονται για το υπόλοιπο.
  - ✓ μπορεί να γνωρίζουν ήδη όλη την πορεία ενός ατόμου, αλλά να ενδιαφέρονται για κάποιες περαιτέρω ευαίσθητες πληροφορίες.

o Τις στοχευμένες αναλυτικές ερωτήσεις που πρέπει να απαντηθούν με τα δεδομένα.

- Στην τεχνική της σύγχυσης διαδρομής και στη τεχνική που συνδυάζει την *k*-ανωνυμία με τη χωρική γενίκευση, απαντώνται ερωτήσεις που αφορούν την τοποθεσία που βρίσκεται ο κάθε χρήστης.
- Στην τεχνική της απόκρυψης μονοπατιού, σύμφωνα με το πείραμα που έχει εκτελεστεί από τους *Hoh et. al* το 2007, οι στοχευμένες ερωτήσεις που απαντώνται αφορούν το χώρο και το χρόνο.
- Οι τεχνικές *NWA* και *W4M* απαντούν σε ερωτήσεις εύρους.
- Όσον αφορά τη τεχνική *AWO*, οι πιο κοινές χρήσεις για τα στατικά δεδομένα είναι οι στατιστικές αναλύσεις και η εξόρυξη δεδομένων, αν εξαιρέσουμε την υποβολή κάποιων ερωτημάτων. Τα ερωτήματα που απαντώνται αφορούν το χώρο και το χρόνο.

Β. Έλεγχος τήρησης των 7 θεμελιωδών αρχών.

1. Προ-ενεργητική όχι Αντιδραστική- Προφυλακτική όχι Διορθωτική. Η αρχή αυτή τηρείται σε όλες τις παραπάνω τεχνικές.
   - Στην τεχνική της σύγχυσης διαδρομής εξ αρχής περιπλέκονται οι διαδρομές των χρηστών ώστε να μη δοθεί η δυνατότητα σε κανένα κακόβουλο χρήστη να πάρει τις πληροφορίες που θέλει.
   - Η βασική ιδέα της τεχνικής που συνδυάζει την *k*-ανωνυμία με τη χωρική γενίκευση είναι να «κρύψει» τις τοποθεσίες με μέσα γενίκευσης, αντικαθιστώντας τις ακριβείς θέσεις μέσα στις τροχιές με τις κατά προσέγγιση θέσεις. Από την αρχή λοιπόν, αποκρύπτει τις θέσεις που υπάρχουν στις τροχιές ώστε να αποφευχθούν διαρροές πληροφοριών.
   - Στην τεχνική της απόκρυψης μονοπατιού, με τη χρήση της *k*-ανωνυμίας επισημοποιείται η έννοια της ισχυρής ανωνυμίας. Το βασικό σκεπτικό με το οποίο έχουν δημιουργηθεί αυτοί οι αλγόριθμοι είναι η γενίκευση ενός καταγεγραμμένου μονοπατιού μέχρι να είναι δυσδιάκριτο από εκείνα τουλάχιστον *k*-1 άλλων ατόμων. Έχει λοιπόν ισχυρή την έννοια της πρόληψης αφού χρησιμοποιεί έναν εναλλακτικό αλγόριθμο της *k*-ανωνυμίας.
   - Στις τεχνικές *NWA* και *W4M* επιτυγχάνεται προστασία της ιδιωτικότητας μέσω της *(k,δ)*-ανωνυμίας, συνεπώς είναι ισχυρή ένδειξη προστασίας που προφυλάσσει έγκαιρα τις πληροφορίες των χρηστών.

- Στην τεχνική ανωνυμίας τροχιάς *AWO*, οι διαρροές πληροφοριών της ιδιωτικότητας αντιμετωπίζονται με τη δημοσίευση βάσεων δεδομένων στατικών τροχιών επεκτείνοντας την έννοια της *k*-ανωνυμίας σε τροχιές. Αφαιρούνται πληροφορίες από τα δεδομένα κάνοντας χρήση: γενικεύσεων χώρου και χρόνου, παράταξης σημείων κατά σειρά στο χώρο και το χρόνο καθώς και κατάργησης σημείου και τροχιάς. Επιπλέον, με το βήμα της ανακατασκευής, εμποδίζεται η περαιτέρω διαρροή πληροφοριών.

2. Η προστασία της ιδιωτικότητας να είναι προεπιλεγμένη ρύθμιση. Η αρχή αυτή τηρείται σε όλες τις παραπάνω τεχνικές.

- Ο αλγόριθμος της σύγχυσης διαδρομής επιστρέφει διαταραγμένες διαδρομές από το αρχικό σύνολο των διαδρομών δύο χρηστών. Μεγιστοποιεί στιγμιαία την προστασία της τοποθεσίας σε κάθε βήμα με την τροποποίηση της αρχικής σειράς δειγμάτων θέσης εντός της ακτίνας διατάραξης *R*.
- Τα κύρια στάδια που εμπλέκονται στην τεχνική που συνδυάζει την *k*-ανωνυμία με τη χωρική γενίκευση είναι: (α) η κατασκευή μίας κατάλληλης ψηφίδωσης της γεωγραφικής περιοχής σε υποπεριοχές που εξαρτώνται από το σύνολο δεδομένων των εισερχόμενων τροχιών *(Voronoi cells)*, (β) η εφαρμογή μιας χωρικής γενίκευσης των αρχικών τροχιών, (γ) ο μετασχηματισμός του συνόλου δεδομένων των γενικευμένων τροχιών για να εξασφαλισθεί ότι πληροί την έννοια της *k*-ανωνυμίας.
- Για τις πληροφορίες τοποθεσίας, στην τεχνική της απόκρυψης μονοπατιού προτείνονται οι αλγόριθμοι χωρικής απόκρυψης, που μειώνουν τη χωρική ακρίβεια του κάθε δείγματος θέσης μέχρι να συναντήσει τον περιορισμό της *k*-ανωνυμίας.
- Οι τεχνικές *NWA* και *W4M* βασίζονται στη τεχνική ανωνυμίας της μετάφρασης χώρου *(space translation)*, με σκοπό τη διατήρηση των ψεύτικων τροχιών όσο το δυνατόν πιο κοντά στις πρωτότυπες.
- Η τεχνική *AWO* προσφέρει προστασία της ιδιωτικότητας των ατόμων από τον αντίπαλο χρησιμοποιώντας τις κάτωθι τεχνικές:
  - ✓ *k*-ανωνυμία: ανωνυμία του συνόλου δεδομένων, έτσι ώστε κάθε τροχιά να είναι δυσδιάκριτη από *k*-1 άλλες τροχιές.
  - ✓ Ανακατασκευή: απελευθέρωση τυχαία επιλεγμένων ατομικών τροχιών από την περιοχή που καλύπτεται από ανώνυμες τροχιές.

3. Η ιδιωτικότητα να είναι ενσωματωμένη στο σχεδιασμό. Η αρχή αυτή τηρείται σε όλες

τις παραπάνω τεχνικές.

- Η προστασία της ιδιωτικότητας όντως αποτελεί ουσιαστικό συστατικό της λειτουργικότητας στη σύγχυση διαδρομής. Κάθε φορά που δύο μονοπάτια χρηστών συναντιούνται, υπάρχει μια πιθανότητα για τον αντίπαλο να μπερδέψει τις τροχιές και να ακολουθήσει το λάθος χρήστη.
- Το σύστημα ανωνυμίας της τεχνικής που συνδυάζει την *k*-ανωνυμία με τη χωρική γενίκευση βασίζεται: (α) στη δημιουργία χωρίσματος στα πεδία της περιοχής που καλύπτεται από τις τροχιές, (β) στην εφαρμογή μιας λειτουργίας για τη χωρική γενίκευση σε όλες τις τροχιές, προκειμένου να τις μετατρέψει σε ακολουθίες σημείων που είναι κέντρα βάρους στις συγκεκριμένες περιοχές και (γ) στη μετατροπή των γενικευμένων τροχιών ώστε να εγγυηθεί προστασία της ιδιωτικότητας.
- Στην τεχνική της απόκρυψης μονοπατιού, για την ενσωμάτωση της ιδιωτικότητας ακολουθείται η εξής διαδικασία: Προσδιορίζονται πρώτα τα κινούμενα δεδομένα που μπορούν να αποκαλυφθούν επειδή έχει περάσει λιγότερος χρόνος από ότι ο μέγιστος χρόνος της σύγχυσης από το τελευταίο σημείο της σύγχυσης. Δεύτερον, προσδιορίζεται ένα σύνολο κινούμενων δεδομένων που μπορούν να αποκαλυφθούν επειδή η τρέχουσα αβεβαιότητα παρακολούθησης είναι μεγαλύτερη από το σχετικό όριο αβεβαιότητας. Τέλος, ενημερώνεται ο χρόνος του τελευταίου σημείου σύγχυσης και το τελευταίο ορατό δείγμα που δείχνει τη θέση κάθε κινούμενου αντικειμένου.
- Οι τεχνικές *NWA* και *W4M* εφόσον εξ αρχής στον αλγόριθμό τους γεννιούνται νέες τροχιές (παρόμοιες με τις αρχικές) έχουν την ιδιωτικότητα εξασφαλισμένη από το σχεδιασμό τους.
- Στον αλγόριθμο του *AWO* σχηματίζονται οι ομάδες των *k* τροχιών με κάποιες ομοιότητες και απαντώνται όλες οι ερωτήσεις έχοντας ήδη ανωνυμία στις βάσεις δεδομένων τους.

5. Πλήρης Προστασία: Πλήρης προστασία του κύκλου ζωής.

Αυτή η αρχή δεν υποστηρίζεται από καμία από τις τεχνικές που ανήκουν στην προστασία ιδιωτικότητας για τη διαφύλαξη της εξόρυξης βάσεων δεδομένων κινούμενων αντικειμένων. Ο λόγος για τον οποίο συμβαίνει αυτό είναι ότι τα δεδομένα δημοσιεύονται, συνεπώς μετά από αυτό το βήμα δε μπορεί να υπάρξει καμία ενέργεια για περαιτέρω προστασία της ιδιωτικότητας.

6. Να υπάρχει ορατότητα και διαφάνεια.

- Στην τεχνική που συνδυάζει την *k*-ανωνυμία με τη χωρική γενίκευση, αλλά και σε εκείνες της σύγχυσης διαδρομής, της απόκρυψης μονοπατιού και του *AWO*, τηρείται η έκτη αρχή. Κάθε βήμα των αντίστοιχων αλγορίθμων τους είναι ξεκάθαρο και δηλώνει επακριβώς τα βήματα που ακολουθούνται. Αυτό έχει ως αποτέλεσμα ο χρήστης να μπορεί να καταλάβει απόλυτα τη διαδικασία που ακολουθεί η εκάστοτε τεχνική.
- Στις τεχνικές *NWA* και *W4M* τηρείται η έκτη αρχή. Η διαδικασία της ομαδοποίησης των τροχιών εκτελείται επιλέγοντας μια τυχαία, και τις *k*-1 πιο κοντινές σε αυτή, που δεν έχουν επισκεφτεί από κανένα. Σε γενικές γραμμές, τα συστατικά μέρη και οι λειτουργίες τους παραμένουν ορατές και διαφανείς στους χρήστες και στους παρόχους.

7. Να υπάρχει σεβασμός στην ιδιωτικότητα του χρήστη και να είναι ο χρήστης το κεντρικό πρόσωπο. Αυτή η αρχή ισχύει σε όλες τις προαναφερθείσες τεχνικές του υποκεφαλαίου.
   - Στην τεχνική που συνδυάζει την *k*-ανωνυμία με τη χωρική γενίκευση, αλλά και σε εκείνες της σύγχυσης διαδρομής και της απόκρυψης μονοπατιού, εφόσον τηρούνται οι πιο πάνω αρχές που αφορούν την ιδιωτικότητα έχουμε σαν συμπέρασμα ότι τηρούν και αυτή την αρχή.
   - Οι τεχνικές *NWA* και *W4M* τηρούν την έβδομη αρχή καθώς μία από τις φάσεις ανάπτυξής τους είναι η μετάφραση χώρου. Σε αυτή, μετασχηματίζεται κάθε συστάδα τροχιών (που έχει δημιουργηθεί μέσω των αλγορίθμων τους) σε ένα *(k, δ)*-σύνολο ανωνυμίας, γεγονός που διασφαλίζει την ιδιωτικότητα του κάθε χρήστη.
   - Στην τεχνική *AWO* η *k*-ανωνυμία περιορίζει τη δυνατότητα του αντιπάλου να συνδέσει όλες τις πληροφορίες για κάποιο άτομο, ενώ η ανακατασκευή των τροχιών αποτρέπει την περαιτέρω διαρροή.

### 3.3.3 Προστασία Ιδιωτικότητας σε Επερωτήσεις Δεδομένων Κίνησης *(Privacy-aware data querying)*

Σε αυτό το υποκεφάλαιο θα ασχοληθούμε με την τεχνική που είναι βασισμένη στη μηχανή αναζήτησης *Hermes++* (*Pelekis et al.*, 2011).

Α. Έλεγχος τήρησης των τριών υποθέσεων.

- Τα (ευαίσθητα) προσωπικά δεδομένα που είναι το αντικείμενο της ανάλυσης.

Ο *Hermes++* θεωρεί ευαίσθητα δεδομένα, εκείνα τα δεδομένα κίνησης των οποίων η αποκάλυψη σε μη αξιόπιστα συμβαλλόμενα μέλη, μπορεί να θέσει σε κίνδυνο την ιδιωτικότητα των χρηστών των οποίων η κίνηση καταγράφεται. Έτσι ανοίγεται ο δρόμος για σενάρια παρακολούθησης κάποιου χρήστη.

- Το μοντέλο επίθεσης.

Υποστηρίζει τρεις τύπους επιθέσεων από κακόβουλους χρήστες:

- ✓ Αναγνώρισης χρήστη
- ✓ Εντοπισμού ευαίσθητης θέσης
- ✓ Διαδοχικής παρακολούθησης

οι οποίες έχουν αναλυθεί στο υποκεφάλαιο 2.3.2.2 του δεύτερου κεφαλαίου.

- Τις στοχευμένες αναλυτικές ερωτήσεις που πρέπει να απαντηθούν με τα δεδομένα.

Ο *Hermes++* υποστηρίζει ερωτήματα: εύρους, απόστασης, *k*-πλησιέστερων γειτόνων, για χωρικά και χωροχρονικά δεδομένα κίνησης.

Β. Έλεγχος τήρησης των 7 θεμελιωδών αρχών.

1. Προ-ενεργητική όχι Αντιδραστική- Προφυλακτική όχι Διορθωτική. Αυτή η αρχή ισχύει.

Η ειδοποιός διαφορά αυτής της τεχνικής σε σχέση με εκείνες που αναλύσαμε έως τώρα, είναι ότι κρατάει τα δεδομένα εντός του οργανισμού που τα φιλοξενεί *(in-house)* και δεν τα δημοσιοποιεί. Είναι λοιπόν εξ αρχής, βασικό στοιχείο πρόληψης για τη διαρροή πληροφοριών.

2. Η προστασία της ιδιωτικότητας να είναι προεπιλεγμένη ρύθμιση. Αυτή η αρχή ισχύει.

Για να εξασφαλιστεί η προστασία της ιδιωτικότητας στα δεδομένα κίνησης που φυλάσσονται εντός του οργανισμού που τα φιλοξενεί, είναι απαραίτητος ένας μηχανισμός που να ελέγχει τις πληροφορίες που τίθενται στη διάθεση των εξωτερικών μελών κατά την υποβολή ερωτημάτων στη βάση δεδομένων. Αυτός ο μηχανισμός είναι ο *Hermes++*. Έτσι λοιπόν, μόνο μη ευαίσθητες πληροφορίες εξέρχονται από τις εγκαταστάσεις που τις φιλοξενούν.

3. Η ιδιωτικότητα να είναι ενσωματωμένη στο σχεδιασμό. Αυτή η αρχή ισχύει.

Ο *Hermes++* λειτουργεί γεννώντας νέες ψεύτικες τροχιές, τις οποίες επιστρέφει σαν απαντήσεις στις ερωτήσεις των χρηστών. Σκοπός είναι η μείωση της εμπιστοσύνης των επιτιθέμενων σχετικά με τις πραγματικές διαδρομές στο αποτέλεσμα του ερωτήματος.

5. Πλήρης Προστασία: Πλήρης προστασία του κύκλου ζωής. Η αρχή αυτή ισχύει.

Η προστασία είναι εξασφαλισμένη από την αρχή σε αυτή τη τεχνική καθώς κρατάει τα δεδομένα στον οργανισμό που φιλοξενούνται και δεν αφήνει περιθώρια στον κακόβουλο χρήστη για περεταίρω επεξεργασία των δεδομένων που θα οδηγήσει στην εξαγωγή κάποιας προσωπικής ευαίσθητης πληροφορίας.

6. Να υπάρχει ορατότητα και διαφάνεια. Ισχύει αυτή η αρχή.

Οι αλγόριθμοι είναι σαφής και δεν υπάρχει κάποια διαδικασία που να μη είναι ορατή ή που να αφήνει αμφιβολίες για τη λειτουργία τους.

7. Να υπάρχει σεβασμός στην ιδιωτικότητα του χρήστη και να είναι ο χρήστης το κεντρικό πρόσωπο. Ισχύει αυτή η αρχή.

Η ιδιωτικότητα είναι εξασφαλισμένη καθώς ο *Hermes++* επιτυγχάνει: (α) να ελέγξει τις τελικές ερωτήσεις των χρηστών και να μπλοκάρει αποτελεσματικά ένα εκτεταμένο σύνολο επιθέσεων στην ιδιωτικότητά τους, εξασφαλίζοντας τη βάση δεδομένων, έναντι της ταυτοποίησης του χρήστη, του εντοπισμού ευαίσθητης θέσης και των διαδοχικών επιθέσεων παρακολούθησης, (β) να δημιουργήσει ομαλές και πιο ρεαλιστικές ψεύτικες τροχιές που διατηρούν τη τάση των αρχικών δεδομένων να τις χρησιμοποιήσει για να ενισχύσουν τα πραγματικά δεδομένα στις απαντήσεις που επιστρέφονται στα ερωτήματα και (γ) να εξασφαλίσει ότι οι ευαίσθητες περιοχές που θα μπορούσαν να οδηγήσουν σε ταυτοποίηση χρήστη δεν έχουν συμπεριληφθεί ως μέρος των επιστρεφόμενων τροχιών.

### 3.3.4 Έλεγχος τήρησης των τριών υποθέσεων και των επτά θεμελιωδών αρχών της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας για δεδομένα με σημασιολογική πληροφορία

Σε αυτή την κατηγορία θα ασχοληθούμε με την τεχνική *C*-ασφάλεια *(C-safety)* των *Monreale et al.* (2011) και την επέκταση της τεχνικής της απόκρυψης σημασιολογικής τοποθεσίας *(cloaking of semantic locations)* των *Yigitoglu et al.* (2012).

Α. Έλεγχος τήρησης των τριών υποθέσεων.

- Τα (ευαίσθητα) προσωπικά δεδομένα που είναι το αντικείμενο της ανάλυσης.
  - Το μοντέλο *C*-ασφάλεια θεωρεί ευαίσθητα τα δεδομένα της ταυτότητάς του, καθώς και όλες τις τοποθεσίες από τις οποίες μπορούν να εξαχθούν συμπεράσματα για τις προσωπικές πληροφορίες κάποιου χρήστη. Κάθε μη ευαίσθητο χαρακτηριστικό θεωρείται οιονεί-αναγνωριστικό.
  - Η επέκταση της τεχνικής της απόκρυψης, θεωρεί ευαίσθητα δεδομένα όλες τις τοποθεσίες από τις οποίες μπορούν να εξαχθούν προσωπικές πληροφορίες ενός χρήστη, για παράδειγμα ένα νοσοκομείο.
- Το μοντέλο επίθεσης.
  - Το μοντέλο *C*-ασφάλεια δίνει τον παρακάτω ορισμό :
    - Ορισμός 3.2

    Γνώση κακόβουλου χρήστη στη τεχνική *C*-ασφάλεια. Ο επιτιθέμενος έχει πρόσβαση στα γενικευμένα δεδομένα ST* και γνωρίζει: (α) τον αλγόριθμο που χρησιμοποιείται για να προσφέρει ανωνυμία στα δεδομένα, (β) την ταξινόμηση PTax της περιοχής της προστασίας του ιδιωτικότητας και (γ) την ακολουθία μιας περιοχής $S_Q$ που είναι οιονεί-αναγνωριστικό και που έχει επισκεφθεί από το δοθέν χρήστη.
  - Η επέκταση της τεχνικής της απόκρυψης, θεωρεί ότι ο επιτιθέμενος ξέρει:
    - ✓ το δίκτυο της πόλης
    - ✓ το προφίλ της ιδιωτικότητας του χρήστη
    - ✓ τους αλγόριθμους της ιδιωτικότητας
    - ✓ όλες τις προηγούμενες αλλά και τις τρέχουσες τοποθεσίες που έχουν αναφερθεί
- Τις στοχευμένες αναλυτικές ερωτήσεις που πρέπει να απαντηθούν με τα δεδομένα.
  - Σύμφωνα με τα πειράματα που εφαρμόζονται στο άρθρο, η τεχνική *C*-ασφάλεια απαντάει σε ερωτήματα που αφορούν την απόσταση.
  - Η επέκταση της τεχνικής της απόκρυψης δεν αναφέρει ποιες είναι οι στοχευμένες αναλυτικές ερωτήσεις που απαντάει.

Β. Έλεγχος τήρησης των 7 θεμελιωδών αρχών.

1. Προ-ενεργητική όχι Αντιδραστική- Προφυλακτική όχι Διορθωτική. Αυτή η αρχή τηρείται και στις δύο τεχνικές.
   - Το μοντέλο της *C*-ασφάλειας, έχει σαν αρχή να κρατάει ιδιωτικά όλα τα ευαίσθητα μέρη που επισκέφθηκε ένας χρήστης. Συνεπώς, το μοντέλο επίθεσης δεν έχει την ικανότητα να συνδέσει τα δεδομένα που έχουν δημοσιευτεί, με άλλες εξωτερικές πληροφορίες ώστε να επιτρέπονται συμπεράσματα για ευαίσθητες τοποθεσίες που έχουν επισκεφθεί.
   - Η επέκταση της τεχνικής της απόκρυψης, βασίζεται στο μοντέλο απόκρυψης σημασιολογικής τοποθεσίας. Σκοπός του είναι να δημιουργήσει κρυπτογραφημένες περιοχές (CRs) που να μην είναι ευαίσθητες στις επιθέσεις ταυτοποίησης κατά των σημασιολογικών τοποθεσιών. Έτσι λοιπόν έχει μια μέθοδο που αποτρέπει την διαρροή πληροφοριών άρα λειτουργεί προληπτικά.
2. Η προστασία της ιδιωτικότητας να είναι προεπιλεγμένη ρύθμιση. Και στις δύο τεχνικές αυτή η αρχή ισχύει.
   - Το μοντέλο της *C*-ασφάλειας εξ αρχής λαμβάνει υπόψη ένα σύνολο δεδομένων από τροχιές με σημασιολογική πληροφορία και ένα όριο πιθανότητας προστασίας με το οποίο πιστοποιείται ότι το σύνολο δεδομένων που θα προκύψει θα είναι *C*-ασφαλές.
   - Η επέκταση της τεχνικής της απόκρυψης, ακολουθεί την παρακάτω διαδικασία:

   Ας πούμε ότι ένας χρήστης βρίσκεται στο κελί *c* που μπορεί είτε να είναι ευαίσθητο είτε όχι και θέλει να μοιραστεί την τοποθεσία του με γνωστούς του. Το βασικό ερώτημα είναι ποια είναι η τοποθεσία η οποία πρέπει να δημοσιευτεί. Έτσι λοιπόν:
   - ✓ Ελέγχεται αν το κελί *c* είναι μέρος μιας ευαίσθητης τοποθεσίας.
   - ✓ Αν είναι, παράγεται μια κρυπτογραφημένη *(cloaked region) r* που περιέχει το κελί και απελευθερώνει την *r*.
   - ✓ Διαφορετικά, αν ο χρήστης δεν είναι σε ένα ευαίσθητο κελί, ελευθερώνεται το *c*.
3. Η ιδιωτικότητα να είναι ενσωματωμένη στο σχεδιασμό. Αυτή η αρχή ισχύει και στις δύο τεχνικές.
   - Η μέθοδος που προτείνεται από την τεχνική *C*-ασφάλεια βασίζεται στη γενίκευση, ωθούμενη από την ταξινόμηση των θέσεων που επισκέπτεται ο χρήστης. Η χρήση ενός πεδίου ταξινόμησης ώστε κάποιες θέσεις να υποστούν γενίκευση, επιτρέπει τη

διατήρηση κάποιου βαθμού σημασιολογικής πληροφορίας στο ανώνυμο σύνολο δεδομένων. Για να αποφευχθεί ο προσδιορισμός των ευαίσθητων περιοχών που επισκέφθηκε ένας χρήστης, πρώτα απ 'όλα θα πρέπει να προσδιορισθεί ποιά σημεία είναι ευαίσθητα και ποιά μη-ευαίσθητα.

- Στην επέκταση της τεχνικής της απόκρυψης θεωρούνται δύο είδη αρχιτεκτονικής: εκτός σύνδεσης *(offline)* και εντός σύνδεσης *(online)*. Στην εκτός σύνδεσης, όλες οι κρυπτογραφημένες περιοχές είναι προϋπολογισμένες και τα αιτήματα για υπηρεσίες που είναι εντός σύνδεσης ελέγχονται για παραβίαση της ιδιωτικότητας σε περίπτωση που η αντίστοιχη κρυπτογραφημένη περιοχή έχει αποκαλυφθεί. Εάν όχι, η αντίστοιχη κρυπτογραφημένη περιοχή αποκαλύπτεται στον πάροχο *LBS*, αλλιώς είναι απαραίτητος ένας μετασχηματισμός. Θεωρούμε δύο είδη των μετασχηματισμών: με χρονική καθυστέρηση και μεταγενέστερος. Στη χρονική καθυστέρηση, η αίτηση αναβάλλεται με βάση το χρόνο. Στο μεταγενέστερο μετασχηματισμό, αντί να αποκαλυφθεί η πραγματική κρυπτογραφημένη περιοχή, αποκαλύπτεται μια προηγούμενη ασφαλής θέση. Στην εντός σύνδεσης αρχιτεκτονική, τόσο η κρυπτογραφημένη περιοχή όσο και μετασχηματισμός γίνονται όταν ζητούνται οι υπηρεσίες από τον χρήστη.

5. Πλήρης Προστασία: Πλήρης προστασία του κύκλου ζωής.
   - Η τεχνική της *C*-ασφάλειας με τη μέθοδο της γενίκευσης φροντίζει ώστε να μην διαρρεύσει καμία ευαίσθητη πληροφορία σε κακόβουλους χρήστες.
   - Η επέκταση της τεχνικής της απόκρυψης με τη μέθοδο της απόκρυψης και τη χρήση των *CRs* παρέχει πλήρη προστασία προς τους χρήστες.
6. Να υπάρχει ορατότητα και διαφάνεια. Αυτή η αρχή ισχύει και για τις δύο τεχνικές.
   - Στην τεχνική *C*-ασφάλεια, η ταξινόμηση μπορεί να χρησιμοποιηθεί για να μετατρέψει μια τροχιά με σημασιολογική πληροφορία σε μια γενικευμένη εκδοχή της. Διαισθητικά, τα σημεία στάσης μπορεί να αντικατασταθούν από άλλα σημεία που ανήκουν στο προηγούμενο επίπεδο στο σχήμα της ταξινόμησης (σχήμα 2.21). Για παράδειγμα, το «νοσοκομείο Ευαγγελισμός» μπορεί να αντικατασταθεί από το «νοσοκομείο». Προφανώς, η γενικευμένη εκδοχή χάνει τις χωροχρονικές πληροφορίες από την αρχική τροχιά, αλλά τείνει να διατηρήσει κάποιο επίπεδο της σημασιολογίας.
   - Στην επέκταση της τεχνικής της απόκρυψης, κάθε βήμα του αλγορίθμου είναι ξεκάθαρο και λεπτομερές και δηλώνει ακριβώς τα βήματα που ακολουθούνται. Ο

χρήστης λοιπόν μπορεί να καταλάβει απόλυτα τη διαδικασία που ακολουθεί αυτή η τεχνική.

7. Να υπάρχει σεβασμός στην ιδιωτικότητα του χρήστη και να είναι ο χρήστης το κεντρικό πρόσωπο. Ισχύει αυτή η αρχή και στις δύο τεχνικές.

- Τα κύρια στάδια της μεθόδου *C*-ασφάλεια είναι συνοπτικά τα παρακάτω:
  - ✓ Απομάκρυνση από το σύνολο δεδομένων κάθε ευαίσθητης περιοχής όταν για το χρήστη αυτό το μέρος είναι ένα οιονεί-αναγνωριστικό.
  - ✓ Ομαδοποίηση σημασιολογικών τροχιών σε ομάδες με προκαθορισμένο μέγεθος *m*.
  - ✓ Δημιουργία μιας γενικευμένης έκδοσης της κάθε τροχιάς με σημασιολογική πληροφορία σε ομάδα, γενικεύοντας τους χώρους των οιονεί-αναγνωριστικών. Σε κάθε ομάδα τα οιονεί-αναγνωριστικά των γενικευμένων τροχιών πρέπει να είναι ταυτόσημα. Τέλος, τα ευαίσθητα σημεία γενικεύονται, όταν η γενίκευση των οιονεί-αναγνωριστικών δεν είναι αρκετή για να παρέχει c-ασφάλεια στο σύνολο δεδομένων.
- Εφόσον η επέκταση της τεχνικής της απόκρυψης ακολουθεί και την αρχή της ιδιωτικότητας που είναι ενσωματωμένη στο σχεδιασμό ($3^{η}$ αρχή), συμπεραίνουμε ότι σέβεται και τις απαιτήσεις του χρήστη όσον αφορά την προστασία των προσωπικών του δεδομένων.

### 3.4 Πίνακες συνοπτικών αποτελεσμάτων

Σε αυτό το υποκεφάλαιο παρουσιάζονται 4 πίνακες με τα συνοπτικά αποτελέσματα του κεφαλαίου 3. Σε κάθε κελί, χρησιμοποιείται το σύμβολο ✓ όταν η τεχνική τηρεί τις εκάστοτε υποθέσεις ή αρχές που εξετάζονται, ενώ χρησιμοποιείται το σύμβολο ✗ σε περίπτωση που η τεχνική δεν τις τηρεί.

**Πίνακας 3.1**

Συνοπτικά αποτελέσματα ελέγχου τήρησης των τριών υποθέσεων της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας για δεδομένα κίνησης

| Τεχνικές Ανωνυμίας / 3 υποθέσεις του PbD | Casper | Απόκρυψη Hilbert | Σύγχυση Διαδρομής | k-ανωνυμία /χωρική γενίκευση | Απόκρυψη Μονοπατιού | NWA | W4M | AWO | Hermes++ |
|---|---|---|---|---|---|---|---|---|---|
| Ευαίσθητα προσωπικά δεδομένα | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Μοντέλο επίθεσης | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Στοχευμένες αναλυτικές ερωτήσεις | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

**Πίνακας 3.2**

Συνοπτικά αποτελέσματα ελέγχου τήρησης των 7 θεμελιωδών αρχών της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας για δεδομένα κίνησης

| Τεχνικές Ανωνυμίας / 7 Θεμελιώδεις Αρχές | Casper | Απόκρυψη Hilbert | Σύγχυση Διαδρομής | k-ανωνυμία /χωρική γενίκευση | Απόκρυψη Μονοπατιού | NWA | W4M | AWO | Hermes++ |
|---|---|---|---|---|---|---|---|---|---|
| Προληπτικά -όχι αντιδραστικά -όχι ενισχυτικά | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Προστασία προσωπικών δεδομένων σαν προεπιλεγμένη ρύθμιση | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Ιδιωτικότητα ενσωματωμένη στο σχεδιασμό | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Ταυτόχρονα ασφάλεια και ιδιωτικότητα | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Προστασία από την αρχή έως το τέλος | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Ορατότητα και διαφάνεια | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Σεβασμός ιδιωτικότητας χρήστη | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

**Πίνακας 3.3**

Συνοπτικά αποτελέσματα ελέγχου τήρησης των τριών υποθέσεων της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας για δεδομένα κίνησης με σημασιολογική πληροφορία

| Τεχνικές Ανωνυμίας / 3 υποθέσεις του PbD | *C*-ασφάλεια | Επέκταση Απόκρυψης Σημασιολογικής Τοποθεσίας |
|---|---|---|
| Ευαίσθητα προσωπικά δεδομένα | ✓ | ✓ |
| Μοντέλο επίθεσης | ✓ | ✓ |
| Στοχευμένες αναλυτικές ερωτήσεις | ✓ | ✗ |

**Πίνακας 3.4**

Συνοπτικά αποτελέσματα ελέγχου τήρησης των 7 θεμελιωδών αρχών της Σχεδιασμένης Προστασίας της Ιδιωτικότητας σε τεχνικές ανωνυμίας για δεδομένα κίνησης με σημασιολογική πληροφορία

| Τεχνικές Ανωνυμίας / 7 Θεμελιώδεις Αρχές | *C*-ασφάλεια | Απόκρυψη Σημασιολογικής Τοποθεσίας |
|---|---|---|
| Προληπτικά -όχι αντιδραστικά -όχι ενισχυτικά | ✓ | ✓ |
| Προστασία προσωπικών δεδομένων σαν προεπιλεγμένη ρύθμιση | ✓ | ✓ |
| Ιδιωτικότητα ενσωματωμένη στο σχεδιασμό | ✓ | ✓ |
| Ταυτόχρονα ασφάλεια και ιδιωτικότητα | ✗ | ✗ |
| Προστασία από την αρχή έως το τέλος | ✓ | ✓ |
| Ορατότητα και διαφάνεια | ✓ | ✓ |
| Σεβασμός ιδιωτικότητας χρήστη | ✓ | ✓ |

# ΚΕΦΑΛΑΙΟ 4

## Συμπεράσματα-Συζήτηση

### 4.1 Συμπεράσματα

Στην παρούσα διπλωματική εργασία, παρουσιάσαμε και αναλύσαμε ένα μεγάλο μέρος των τεχνικών που ασχολούνται με την προστασία της ιδιωτικότητας σε δεδομένα κίνησης. Μέσω της βιβλιογραφικής ανασκόπησης ελέγξαμε αν οι υπάρχουσες τεχνικές είναι επαρκείς για να αποτρέψουν κάθε πιθανή προσπάθεια των κακόβουλων χρηστών να ανακαλύψουν τις ευαίσθητες πληροφορίες που επιθυμούν. Μπορούμε, σε γενικές γραμμές, να αναφέρουμε ότι η ιδιωτικότητα καλύπτεται σε μεγάλο βαθμό από τις υπάρχουσες τεχνικές παρόλο που οι περισσότερες από αυτές έχουν σημεία προς βελτίωση. Πιο συγκεκριμένα, η μόνη κατηγορία δεδομένων για την οποία δεν υπάρχει μεγάλος αριθμός τεχνικών είναι τα δεδομένα με σημασιολογική πληροφορία.

Ο όρος της Σχεδιασμένης Προστασίας της Ιδιωτικότητας *(Privacy by Design)* χρησιμοποιήθηκε σαν μέσο αξιολόγισης κάποιων από τις τεχνικές ανωνυμίας. Στην πλειοψηφία, τους οι τεχνικές που μελετήθηκαν τηρούν τις τρεις υποθέσεις και τις επτά θεμελιώδεις αρχές που προτείνει το *PbD*. Πιο συγκεκριμένα, η μόνη αρχή που δεν ακολουθείται από μεγάλο μέρος των τεχνικών, είναι αυτή της «Προστασίας από την αρχή έως το τέλος». Οι τεχνικές αυτές ανήκουν στην κατηγορία της Προστασίας Ιδιωτικότητας κατά τη Δημοσιοποίηση Δεδομένων Κίνησης. Η μη τήρηση αυτής της αρχής, υποδηλώνει ότι σε κάποιο σημείο της διαδικασίας της ανωνυμίας τα δεδομένα δεν είναι επαρκώς προστατευμένα και υπάρχει κίνδυνος διαρροής τους που μπορεί να οδηγήσει σε προσβολή της ιδιωτικότητας. Συνεπώς οι τεχνικές αυτής της κατηγορίας θεωρείται ότι έχουν ένα αδύναμο σημείο, που όμως μπορεί να οδηγήσει σε άμεση προσβολή της ανωνυμίας. Επιπρόσθετα, η μόνη υπόθεση που δεν καλύπτεται, αφορά τη μία από τις τεχνικές για δεδομένα τροχιάς με σημασιολογική πληροφορία. Η επέκταση της τεχνικής απόκρυψης σημασιολογικής τοποθεσίας λοιπόν, δεν αναφέρεται στις στοχευμένες-αναλυτικές ερωτήσεις που μπορούν να απαντηθούν και συνεπώς δε δίνει τη δυνατότητα στο χρήστη να κατανοήσει απόλυτα τους σκοπούς και τη λειτουργία της.

Μία επιπρόσθετη επισήμανση, αφορά το σημαντικό ρόλο της τεχνικής της *k*-ανωνυμίας σε μεταγενέστερες τεχνικές. Όπως μπορούμε να κρίνουμε από τις τεχνικές που αξιολογήθηκαν με γνώμονα τον όρο της Σχεδιασμένης Προστασίας της Ιδιωτικότητας, η τεχνική της *k*-ανωνυμίας εγγυάται την τήρηση της αρχής ώστε οι τεχνικές να είναι προφυλακτικές και προενεργηκτικές και όχι αντιδραστικές και διορθωτικές, αλλά και της αρχής ώστε οι τεχνικές να έχουν την ιδιωτικότητα ενσωματωμένη στο σχεδιασμό, ενώ σε κάποιες ακόμα τεχνικές πιστοποιεί και κάποιες άλλες αρχές όπως αυτή του σεβασμού στην ιδιωτικότητας του χρήστη.

## 4.2 Παρεμβάσεις

Σε αυτή την παράγραφο προτείνονται κάποιες παρεμβάσεις που θα μπορούσαν να βελτιώσουν τις υπάρχουσες τεχνικές ανωνυμίας.

Αρχικά, θα θέλαμε να επισημάνουμε ότι οι τεχνικές ανωνυμίας πρέπει να γίνουν περισσότερο προσωποκεντρικές, καθώς το κεντρικό πρόσωπο όσον αφορά την προστασία της ιδιωτικότητας είναι ο χρήστης και πιο συγκεκριμένα οι προσωπικές του πληροφορίες που δημοσιεύονται. Συνεπώς πρέπει να λαμβάνονται περισσότερο υπόψη οι επιλογές του χρήστη σχετικά με τα δεδομένα που τον αφορούν. Όσον αφορά τις τεχνικές που ανήκουν στην κατηγορία της Προστασίας Ιδιωτικότητας κατά τη Δημοσιοποίηση Δεδομένων Κίνησης, όπως προαναφέραμε στο υποκεφάλαιο 4.1, πρέπει να δοθεί μεγαλύτερη βάση στη διατήρηση της προστασίας σε όλη τη διαδικασία ανωνυμίας. Είναι πολύ σημαντικό για μία τεχνική να μπορεί να επιβεβαιώσει στους χρήστες ότι προφυλάσσει την ιδιωτικότητά τους από την αρχή ως το τέλος της διαδικασίας.

Στη συνέχεια, όσον αφορά τις τεχνικές που ασχολούνται με δεδομένα με σημασιολογική πληροφορία, είτε σε χωρικά δεδομένα, είτε σε δεδομένα τροχιάς, πρωτίστως θεωρούμε ότι πρέπει να αυξηθούν. Επιπλέον, για τις τεχνικές που ασχολούνται με χωρικά δεδομένα θα πρέπει να ληφθεί υπόψη η περίπτωση που οι πάροχοι υπηρεσιών δεν είναι έμπιστοι. Μια τέτοια περίπτωση καθιστά αναγκαίους επιπλέον ελέγχους για τυχόν διαρροή πληροφοριών μέσω των παρόχων. Για τις τεχνικές που ασχολούνται με δεδομένα τροχιάς, θα θέλαμε να προτείνουμε να συμπεριλαμβάνονται και οι επιλογές των ίδιων των χρηστών όσον αφορά τα οιονεί-αναγνωριστικά.

### 4.3 Μελλοντικές επεκτάσεις

Μία μελλοντική επέκταση που έχουμε να προτείνουμε, αφορά τις τεχνικές που κρατούν τα δεδομένα εντός του οργανισμού που τα φιλοξενεί *(in-house)*. Θεωρούμε λοιπόν, ότι πρέπει να αποκτήσουν όσον το δυνατόν μεγαλύτερη ποικιλία στις στοχευμένες ερωτήσεις που μπορούν να απαντήσουν.

Μια ακόμα μελλοντική επέκταση σχετίζεται με τις τεχνικές που ασχολούνται με τροχιές με σημασιολογική πληροφορία. Η εξέλιξη της τεχνολογίας όσον αφορά τα συστήματα πλοήγησης *(GPS)*, έχει ωθήσει όλο και περισσότερους ανθρώπους στις μέρες μας να τα χρησιμοποιούν όχι μόνο από ειδικές συσκευές αλλά και μέσω κινητών τηλεφώνων. Συνεπώς, έχει αυξηθεί ο αριθμός των χρηστών που δημοσιοποιούν την τοποθεσία που βρίσκονται μια δεδομένη χρονική στιγμή αλλά και την τοποθεσία στην οποία θέλουν να βρεθούν. Προτείνουμε λοιπόν, τη δημιουργία τεχνικών που να ασχολούνται με την προστασία ιδιωτικότητας των χρηστών αυτών, οι οποίοι παρόλο που δημοσιοποιούν προσωπικές τους πληροφορίες, επιθυμούν η ιδιωτικότητά τους να είναι υπό προστασία.

Εν κατακλείδι, λόγω της σοβαρότητας του αντικειμένου της προστασίας της ιδιωτικότητας και λόγω της ραγδαίας εξέλιξης της τεχνολογίας, επιβάλλεται να δημιουργούνται συνεχώς νέες τεχνικές ανωνυμίας καθώς θα έχουν να αντιμετωπίσουν καινούριους κινδύνους.

# ΒΙΒΛΙΟΓΡΑΦΙΑ

## Ξένη


Abul O., Bonchi F., Nanni M. (2010). Anonymization of moving objects databases by clustering and perturbation. *Information Systems*, *35*(8), 884-910.

Adam N. R., Wortmann J. C. (December 1989). Security-Control Methods for Statistical Databases: A Comparative Study. *ACM Computing Surveys (CSUR)*, *21*(4), 515-556.

Cavoukian A. (2011). Privacy by Design in Law, Policy and Practice.

Damiani M. L., Silvestri C., Bertino E. (2011). Fine-Grained Cloaking of Sensitive Positions in Location-Sharing Applications. IEEE Pervasive Computing, vol. 10(4), pp. 64–72.

Gkoulalas A., Divanis P., Kalnis V., Verykios S. (2010). Providing k–Anonymity in Location Based Services. *ACM SIGKDD Explorations Newsletter*, *12*(1), 3-10.

Hoh B., Gruteser M., Xiong H. Alrabady A. (2007). Preserving Privacy in GPS Traces via Uncertainty-Aware Path Cloaking. In *Proceedings of the 14th ACM conference on Computer and communications security* (pp. 161-171). ACM.

Hoh B., Gruteser M. (2005). Protecting Location Privacy Through Path Confusion. In *Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on* (pp. 194-205). IEEE.

Kalnis P., Ghinita G., Mouratidis K., Papadias D. (2007). Preventing location-based identity inference in anonymous spatial queries. *Knowledge and Data Engineering, IEEE Transactions on*, *19*(12), 1719-1733.

Li N., Li T., Venkatasubramanian S. (2007). t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on* (pp. 106-115). IEEE.

Machanavajjhala A., Gehrke J., Kifer D., Venkitasubramaniam M. (2006). l- Diversity: Privacy Beyond k-Anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *1*(1), 3.

Mahdavifar S., Abadi M., Kahani M., Mahdikhani H. (2012). A clustering-based approach for personalized privacy preserving publication of moving object trajectory data. In *Network and System Security* (pp. 149-165). Springer Berlin Heidelberg.

Mokbel M. F., Chow C. Y., Aref W. G. (2006, September). The new Casper: query processing for location services without compromising privacy. In *Proceedings of the 32nd international conference on Very large data bases* (pp. 763-774). VLDB Endowment.

Monreale A., Trasarti R., Pedreschi D., Renso C., Bogorny V. (2011). C-safety: a framework for the anonymization of semantic trajectories. *Transactions on Data Privacy, 4*(2), 73-101.

Monreale A., Andrienko G., Andrienko N., Giannotti F., Pedreschi D., Rinzivillo S., Wrobel1 S (2010). Movement Data Anonymity through Generalization. *Transactions on Data Privacy, 3*(2), 91-121.

Nergiz M. E., Atzori M., Saygin Y. (2008). Towards Trajectory Anonymization: A Generalization-Based Approach. In *Proceedings of the SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy in GIS and LBS* (pp. 52-61). ACM.

Pedreschi D., Monreale A., Giannotti F. (2012). Privacy by design in data mining.

Pelekis N., Gkoulalas-Divanis A., Vodas M., Kopanaki D., Theodoridis Y. (2011). Privacy-Aware Querying over Sensitive Trajectory Data. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 895-904). ACM.

Sweeney L.(2002). k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 557-570.

Yigitoglu E., Damiani M. L., Abul O., Silvestri C. (2012). Privacy-preserving sharing of sensitive semantic locations under road-network constraints. In *Mobile Data Management (MDM), 2012 IEEE 13th International Conference on* (pp. 186-195). IEEE.