Βόλος 2007

Τμήμα μηχανικών Η/Υ Τηλεπικοινωνιών και δικτύων

Θέμα Διπλωματικής Εργασίας:
Σχεδιασμός και Υλοποίηση Αλγορίθμων προστασίας της ιδιωτικότητας σε χώρο-χρονικά δεδομένα.

Επιβλέπων καθηγητής:
Βερύκιος Βασίλειος

Γιαννακόπουλος Σπύρος

Περιεχόμενα

1. Εισαγωγή

Ο επιστημονικός τομέας των ασύρματων τηλεπικοινωνιών τα τελευταία χρόνια σημειώνει αλματώδη ανάπτυξη, με αποτέλεσμα οι ασύρματες συσκευές να έχουν τη δυνατότητα παροχής προηγμένων υπηρεσιών. Αυτές οι προηγμένες υπηρεσίες συμπεριλαμβάνουν υπηρεσίες για την αποστολή εικόνας και βίντεο, υπηρεσίες πλοήγησης, οι οποίες παρέχονται με την χρήση της τεχνολογίας GPS, υπηρεσίες αναζήτησης(π.χ εύρεση των πέντε πιο κοντινών φαρμακείων) κτλ. Αυτές οι υπηρεσίες παρέχονται από έναν παροχέα υπηρεσιών, ο οποίος λαμβάνει αιτήσεις που κάνουν οι χρήστες, μέσω ασύρματων συσκευών που διαθέτουν. Όταν ένας χρήστης κάνει αίτηση προς έναν παροχέα υπηρεσιών, θα πρέπει να στείλει στον παροχέα κάποιες πληροφορίες. Για παράδειγμα σε υπηρεσίες πλοήγησης, θα πρέπει να σταλεί στον παροχέα υπηρεσιών, η θέση από όπου ο χρήστης έκανε την αίτηση, αλλά και η χρονική στιγμή που έγινε η συγκεκριμένη αίτηση. Αυτές οι πληροφορίες είναι αναγκαίες, για να μπορέσει ο παροχέας υπηρεσιών να παρέχει υπηρεσίες πλοήγησης. Για άλλου είδους υπηρεσίες πρέπει να σταλούν διαφορετικές πληροφορίες.

Το ζήτημα που προκύπτει, είναι ότι οι πληροφορίες που στέλνει κάθε χρήστης στον παροχέα υπηρεσιών, όταν κάνει μια αίτηση, δεν πρέπει να προσδιορίζουν ευαίσθητα προσωπικά δεδομένα, όπως για παράδειγμα το όνομα και το επώνυμο του αιτούντα, το αν πάσχει από κάποια ασθένεια κτλ. Έτσι για να διασφαλισθεί η ιδιωτικότητα των χρηστών, πρώτα απ' όλα δεν πρέπει να αποστέλλονται στον παροχέα υπηρεσιών το όνομα και το επώνυμο του κάθε χρήστη. Αντίθετα για τον μονοσήμαντο προσδιορισμό του κάθε χρήστη, αποστέλλεται στον παροχέα υπηρεσιών ένα αναγνωριστικό το οποίο είναι διαφορετικό για τον κάθε χρήστη.

Η παραπάνω προσέγγιση δεν λύνει εντελώς το πρόβλημα, καθώς υπάρχει και πάλι κίνδυνος να αποκαλυφθούν προσωπικά δεδομένα του χρήστη όπως: το όνομα και το επώνυμο του, η διεύθυνση κατοικίας του, το επάγγελμα του, το αν πάσχει από κάποια ασθένεια κτλ. Κάτι τέτοιο μπορεί να γίνει με συνδυασμό διαφόρων στοιχείων. Για παράδειγμα αν ένας χρήστης κάνει καθημερινά την διαδρομή από το σπίτι του προς το νοσοκομείο, μπορεί να βγει το συμπέρασμα

ότι πάσχει από κάποια σοβαρή ασθένεια. Αυτό το στοιχείο, αν ήταν γνωστό σε κάποιον εργοδότη, τότε ίσως αυτό θα αποτελούσε αντικίνητρο για την πρόσληψη αυτού του ατόμου και τελικά να μην γινόταν η πρόσληψή του.

Με βάση τα όσα προαναφέρθηκαν, γίνεται φανερό ότι για να διασφαλισθεί η ιδιωτικότητα των χρηστών, εκτός από την απόκρυψη των ευαίσθητων δεδομένων τους, είναι απαραίτητη και η απόκρυψη της θέσης όπου βρισκόταν ο χρήστης όταν έκανε την αίτηση, αλλά και των διαδρομών που κάνει συχνά. Σε αυτή την εργασία, θα προταθούν κάποιες τεχνικές όπου διασφαλίζουν την ιδιωτικότητα των χρηστών που στέλνουν αιτήσεις σε έναν παροχέα υπηρεσιών.

Στα πλαίσια αυτής της εργασίας, θεωρούμε ότι οι χρήστες που κάνουν αιτήσεις βρίσκονται σε έναν γεωγραφικό χώρο που αναπαριστά μια πόλη. Για την δημιουργία και αποθήκευση ενός χάρτη, ο οποίος αναπαριστά μια πόλη, χρησιμοποιήθηκαν ειδικές δομές που παρέχονται, προκειμένου να υπάρχει η δυνατότητα χειρισμού και αποθήκευσης χωρικών δεδομένων. Με τον όρο χωρικά δεδομένα, εννοούμε τα δεδομένα που χρησιμοποιούνται για τον προσδιορισμό της θέσης ενός χρήστη στον χώρο που έχουμε ορίσει. Από την άλλη μεριά, τα χρονικά δεδομένα μοντελοποιούν την παράμετρο του χρόνου. Ο ορισμός της παραμέτρου του χρόνου είναι απαραίτητος, διότι πρέπει να ξέρουμε ποια χρονική στιγμή ο εκάστοτε χρήστης έκανε μια αίτηση προς τον παροχέα υπηρεσιών.

Η δομή που έχει η παρούσα εργασία είναι η εξής: Στο κεφάλαιο 2 θα παρουσιαστεί σχετική δουλειά που έχει γίνει πάνω στον τομέα της ιδιωτικότητας των χώρο-χρονικών δεδομένων. Στο κεφάλαιο 3 θα γίνει μια περιγραφή των χωρικών αλλά και των χρονικών δεδομένων, καθώς και πως αυτού του είδους τα δεδομένα αποθηκεύονται στο ΣΔΒΔ της oracle. Στο κεφάλαιο 4 θα περιγράψουμε τα χώρο-χρονικά δεδομένα, αλλά και πως μπορούμε να παράγουμε τέτοια δεδομένα με την χρήση κάποιων ειδικών προγραμμάτων(Generators). Στο κεφάλαιο 5 θα διατυπώσουμε με ακρίβεια το πρόβλημα της ιδιωτικότητας των χώρο-χρονικών δεδομένων, με ορισμούς και παραδείγματα. Στο κεφάλαιο 6 θα παρουσιάσουμε την λύση του προβλήματος. Τέλος στο κεφάλαιο 7 θα εφαρμόσουμε τη λύση που προτείναμε σε πειραματικά δεδομένα. Επίσης θα παρουσιάσουμε και θα αξιολογήσουμε τα αποτελέσματα που προέκυψαν.

ΚΕΦΑΛΑΙΟ 2

2. Σχετική δουλειά

Σχετικά με το θέμα της ιδιωτικότητας χώρο-χρονικών δεδομένων, έχουν προταθεί πολλές τεχνικές. Οι τεχνικές αυτές έχουν σαν στόχο να διασφαλισθεί η ιδιωτικότητα των χρηστών, των οποίων τα στοιχεία αποθηκεύονται σε μια βάση δεδομένων. Ένας αλγόριθμος που διασφαλίζει την ιδιωτικότητα των χρηστών, είναι ο αλγόριθμος της κ-ανωνυμίας. Αυτός ο αλγόριθμος προτάθηκε από τις Samarati και Sweeney στο [1]. Η πρόταση που έγινε από τις Samarati και Sweeney, εξασφαλίζει κ-ανωνυμία για τους χρήστες, αν και μόνο αν η βάση είναι κ-ανώνυμη. Η συγκεκριμένη τεχνική βασίζεται στους προσδιοριστές προστασίας (QIs), των οποίων τυπικός ορισμός και περιγραφή βρίσκεται στο [2] . Εν' συντομία αναφέρουμε ότι σαν προσδιοριστής προστασίας, ορίζεται το σύνολο των γνωρισμάτων της βάσης, τα οποία σε συνδυασμό με πληροφορίες από εξωγενείς πηγές μπορούν να προσδιορίσουν την ταυτότητα συγκεκριμένων ατόμων.

Επίσης η Sweeney στο [3], προτείνει έναν αλγόριθμο ελάχιστης γενίκευσης, σύμφωνα με τον οποίο ο πίνακας που περιέχει τα προσωπικά στοιχεία των χρηστών τροποποιείται κατάλληλα, έτσι ώστε να διασφαλίζεται κ-ανωνυμία. Η τροποποίηση του πίνακα με τα στοιχεία των χρηστών, γίνεται με τέτοιο τρόπο ώστε τα στοιχεία των χρηστών να δέχονται την ελάχιστη δυνατή τροποποίηση-παραποίηση(distortion). Τέλος η Samarati στο [4], προτείνει έναν αλγόριθμο, ο οποίος υπολογίζει πολλαπλές ελάχιστες γενικεύσεις, και επιλέγει από αυτές, την γενίκευση με την ελάχιστη παραποίηση(distortion) των δεδομένων των χρηστών.
Οι Agrawal και Srikant στο [5], πρότειναν έναν αλγόριθμο για τη διασφάλιση κ-ανωνυμίας που βασίζεται στην θεωρία των γράφων. Ειδικότερα το πρόβλημα της διασφάλισης κ-ανωνυμίας, το μοντελοποίησαν με ένα γράφο, οι ακμές του οποίου συμβολίζουν, τις συσχετίσεις μεταξύ των γνωρισμάτων των χρηστών. Από αυτόν τον γράφο διαγράφονται σταδιακά ακμές μέχρι να βρεθεί ένα ελάχιστο επικαλύπτον δέντρο, το οποίο να εξασφαλίζει κ-ανωνυμία των χρηστών. Στο [6] οι Ashwin Machanavajjhala και Daniel Kifer, προτείνουν μια τεχνική διασφάλισης

κ-ανωνυμίας, η οποία βασίζεται στους προσδιοριστές προστασίας και στην ελάχιστη γενίκευση. Πιο συγκεκριμένα, με βάση την προτεινόμενη τεχνική, τα γνωρίσματα των χρηστών χωρίζονται σε δύο κατηγορίες ανάλογα με το αν αυτά μπορούν να προσδιορίσουν μονοσήμαντα ένα χρήστη ή όχι. Στην συνέχεια για αυτά τα γνωρίσματα που αποτελούν μέρος των προσδιοριστών προστασίας, εφαρμόζεται γενίκευση, μέχρι ο πίνακας με γνωρίσματα των χρηστών να γίνει κ-ανώνυμος.

Οι προτάσεις που προαναφέραμε σχετικά με την διασφάλιση της ιδιωτικότητας των δεδομένων των χρηστών, βασίζονται κατά κύριο λόγο στην μεταβολή-τροποποίηση των γνωρισμάτων τα οποία αποτελούν μέρος ενός προσδιοριστή προστασίας. Επιπρόσθετα όλες οι προαναφερθείσες τεχνικές, δε λαμβάνουν υπ' όψη τα χώρο-χρονικά στοιχεία των χρηστών. Τα χώρο-χρονικά στοιχεία των χρηστών όμως, σε πολλές περιπτώσεις είναι δυνατόν να αποκαλύψουν την ταυτότητα, καθώς και άλλα προσωπικά δεδομένα των χρηστών. Άρα επιβάλλεται τα χώρο-χρονικά δεδομένα να προστατεύονται κατάλληλα, έτσι ώστε να διασφαλίζεται η ιδιωτικότητα των χρηστών.

Τεχνικές οι οποίες ικανοποιούν την λειτουργικότητα σχετικά με την προστασία των χώρο-χρονικών στοιχείων των χρηστών, προτείνονται στα [7] & [8]. Ωστόσο αυτές οι τεχνικές που προτείνονται είναι κάπως περιοριστικές για δύο λόγους. Πρώτον γιατί θεωρείται ότι η τιμή του 'κ' (η οποία καθορίζει το επίπεδο ανωνυμίας), δεν είναι δυνατόν να μεταβάλλεται, αλλά ορίζεται εξ' αρχής. Δεύτερον διότι για να θεωρηθεί ότι ένα χρήστης μπορεί να ενταχτεί στο σύνολο ανωνυμίας του αιτούντα, πρέπει να έχει κάνει μια αίτηση από ένα σημείο που είναι αρκετά κοντά στην θέση από όπου έγινε η αίτηση. Αυτό είναι αρκετά περιοριστικό, διότι όταν ένας χρήστης κάνει μια αίτηση, προκειμένου να διασφαλισθεί κ-ανωνυμία, πρέπει να βρεθούν τουλάχιστο 'κ' χρήστες που να καλύπτουν τους χώρο-χρονικούς περιορισμούς της γενίκευσης και παράλληλα να έχουν κάνει και αιτήσεις, πράγμα που είναι σχεδόν απίθανο.

Από την άλλη πλευρά οι C. Bettini, X.S. Wang και S. Jajodia στα [9] & [10], προτείνουν μια τεχνική για την διασφάλιση κ-ανωνυμίας που είναι αρκετά πιο ικανοποιητική. Ειδικότερα προεκτείνεται η έννοια των προσδιοριστών προστασίας([2]), και γίνεται λόγος για προσδιοριστές προστασίας που βασίζονται στην τοποθεσία(LBQIDs). Επίσης προτείνεται ο αλγόριθμος της γενίκευσης αλλά και ο αλγόριθμος αποσύνδεσης. Επιγραμματικά αναφέρεται ότι ο πρώτος αλγόριθμος προκειμένου να διασφαλίσει κ-ανωνυμία προσπαθεί να βρει τους 'κ' κοντινότερους γείτονες του αιτούντα. Αν ο αλγόριθμος γενίκευσης αποτύχει τότε εκτελείται ο αλγόριθμος αποσύνδεσης, ο οποίος σαν στόχο έχει να αλλάξει το αναγνωριστικό του αιτούντα, προκειμένου να διασφαλιστεί η ανωνυμία του.

Στα πλαίσια αυτής τη εργασία υλοποιήθηκαν οι αλγόριθμοι γενίκευσης και αποσύνδεσης. Στα [9] & [10] δεν προτείνεται υλοποίηση για τον αλγόριθμο αποσύνδεσης και έτσι προτάθηκε μια νέα ενδεικτική υλοποίηση αυτού του αλγορίθμου. Και οι δύο αυτοί αλγόριθμοι υλοποιήθηκαν, έτσι ώστε τα στοιχεία των χρηστών να αποθηκεύονται σε μία βάση χώρο-χρονικών δεδομένων. Για αυτό τον σκοπό χρησιμοποιήθηκε το ΣΔΒΔ της Oracle.

Σχετικά με την παραγωγή των χρηστών της εφαρμογής που αναπτύχθηκε, χρησιμοποιήθηκαν προγράμματα παραγωγής χώρο-χρονικών αντικειμένων (Generators) που περιγράφονται στα [11], [12] & [13]. Επίσης σχετικά με τον ορισμό χωρικών αντικειμένων, αλλά και την κατασκευή χωρικών δικτύων στο ΣΔΒΔ της Oracle, πληροφορίες παρέχονται στα [14], [15] & [16]. Τέλος για τη διασύνδεση του run time environment της java με το ΣΔΒΔ της Oracle χρησιμοποιήθηκε το πρωτόκολλο JDBC, σχετικές πληροφορίες για το οποίο βρίσκονται στα [17] & [18].

ΚΕΦΑΛΑΙΟ 3

3. Χωρικά και χρονικά δεδομένα

Τα χωρικά αλλά και τα χρονικά δεδομένα αποτελούν αντικείμενο αυτής της εργασίας και θα χρησιμοποιηθούν προκειμένου να υλοποιηθούν κάποιοι αλγόριθμοι που σχετίζονται με την ιδιωτικότητα τους. Έτσι σε αυτό το κεφάλαιο θα γίνει μια περιγραφή των χωρικών και των χρονικών δεδομένων. Τα χωρικά δεδομένα χρησιμοποιούνται προκειμένου να προσδιοριστεί η θέση ενός χωρικού αντικειμένου. Από την άλλη πλευρά, η έννοια του χρόνου πρέπει να ορισθεί έτσι ώστε να ξέρουμε πότε ο εκάστοτε χρήστης έκανε μια αίτηση προς τον παροχέα υπηρεσιών. Για την μοντελοποίηση της παραμέτρου του χρόνου, γίνεται χρήση των χρονικών δεδομένων. Τα χρονικά δεδομένα είναι δεδομένα που αλλάζουν με την πάροδο του χρόνου. Αυτά τα δεδομένα, αν συνδυαστούν με κάποια άλλα αντικείμενα, παρέχουν πλήθος πληροφοριών. Για παράδειγμα αν συσχετίσουμε τα χρονικά δεδομένα, με τους χρήστες ενός συστήματος (π.χ με τους χρήστες που κάνουν login σε ένα σύστημα διαχείρισης βάσεων δεδομένων) μπορούμε να πούμε τι ώρα έκανε ο κάθε χρήστης login.

3.1 Χωρικά δεδομένα

Ένα αντικείμενο χωρικών δεδομένων χαρακτηρίζεται από την θέση του στον χώρο, αλλά και από τα όριά του. Τα χωρικά δεδομένα χωρίζονται σε δυο μεγάλες κατηγορίες. Η πρώτη είναι τα ***σημειακά δεδομένα*** και η δεύτερη τα ***δεδομένα περιοχής***. Τα *σημειακά δεδομένα* προσδιορίζονται πλήρως από το στίγμα τους και για τον ορισμό τους αρκεί να προσδιορίσουμε την θέση τους στον χώρο. Για παράδειγμα ένα σημείο **A** ορίζεται ως εξής: **A(x,y)**, όπου **x**,**y** είναι οι συντεταγμένες του. Από την άλλη πλευρά τα *δεδομένα περιοχής* δεν είναι σημεία αλλά γεωμετρικά σχήματα όπως: πολύγωνα, κύκλοι κτλ. Για τον προσδιορισμό της θέσης των *δεδομένων περιοχής* απαιτούνται πιο πολύπλοκες δομές. Τα *σημειακά δεδομένα*, χαρακτηρίζονται και σαν δομικά στοιχεία αφού με

βάση αυτά μπορούμε να ορίσουμε πιο πολύπλοκες γεωμετρίες. Με αυτή την λογική μπορούμε να αναπαραστήσουμε οποιοδήποτε γεωμετρικό σχήμα. Πιο κάτω ορίζεται τι είναι γεωμετρία.

Ορισμός 3.1: Ως γεωμετρία ορίζουμε την αναπαράσταση των χωρικών χαρακτηριστικών, τα οποία μπορεί να έχουν οριστεί από απλά σημειακά αντικείμενα ή και από πιο πολύπλοκα αντικείμενα όπως πολύγονα και κύκλους.

3.1.1 Συστήματα συντεταγμένων

Για να μπορούμε να ορίσουμε σχέσεις μεταξύ των αντικειμένων που δημιουργούμε, για παράδειγμα για να μπορούμε να αποφανθούμε στο ερώτημα πόσο απέχει ένα αντικείμενο από ένα άλλο, θα πρέπει να ορίσουμε ένα σύστημα συντεταγμένων. Τα συστήματα συντεταγμένων εντάσσονται σε δύο βασικές κατηγορίες. Πρώτον είναι αυτά που σχετίζονται με την αναπαράσταση της γης. Σε αυτά τα συστήματα υπάρχουν προκαθορισμένα μέτρα μέτρησης όπως για παράδειγμα το μέτρο, το μίλι κτλ. Η δεύτερη κατηγορία είναι αυτή που δεν σχετίζεται με την αναπαράσταση της γης.
Στα πλαίσια αυτής της εργασίας, χρησιμοποιήθηκε το ΣΔΒΔ της Oracle για τη αποθήκευση των χωρικών δεδομένων. Η Oracle υποστηρίζει τέσσερα συστήματα συντεταγμένων:

1)*Καρτεσιανό σύστημα συντεταγμένων*:

Είναι το πιο γνωστό σύστημα συντεταγμένων, στο οποίο η θέση ενός αντικείμενου προσδιορίζεται από ένα ζεύγος αριθμών **(x,y)**. Αυτοί οι αριθμοί προκύπτουν από την προβολή του σημείου σε δυο κάθετους μεταξύ τους άξονες. Αυτό είναι το προκαθορισμένο σύστημα που χρησιμοποιεί η oracle και εμείς αυτό υιοθετήσαμε σε αυτή την εργασία.

2)*Γεωδαιτικό σύστημα συντεταγμένων*:

Είναι ένα γωνιακό σύστημα, το οποίο στηρίζεται στο γεωγραφικό μήκος και πλάτος. Βάση του συστήματος αυτού είναι οι πολικές συντεταγμένες. Το

γεωδαιτικό σύστημα συντεταγμένων, σχετίζεται άμεσα με την μέτρηση αποστάσεων, ανάμεσα σε γεωμετρίες που βρίσκονται πάνω στην γη.

3)*Σύστημα συντεταγμένων προβολής*:

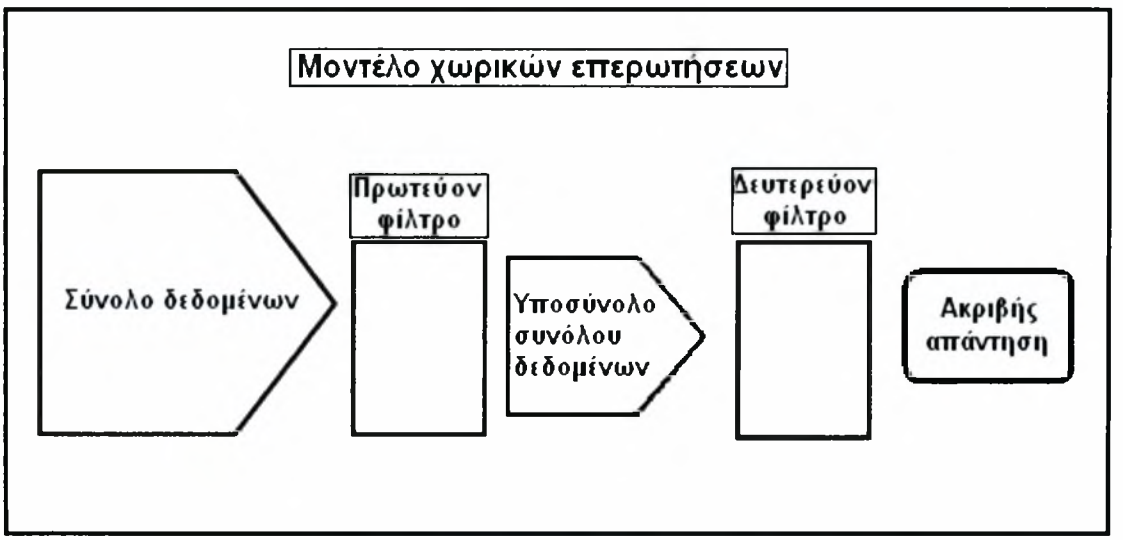
Οι συντεταγμένες σε αυτό το σύστημα είναι στον δισδιάστατο χώρο. Οι συντεταγμένες αυτές είναι καρτεσιανές, οι οποίες και έχουν προκύψει από μια μαθηματική απεικόνιση των γεωδαιτικών συντεταγμένων σε ένα δισδιάστατο επίπεδο.

4)*Σύστημα τοπικών συντεταγμένων*:

Είναι καρτεσιανές συντεταγμένες οι οποίες δεν έχουν καμία σχέση με την αναπαράσταση της γης. Είναι ένα σύστημα το οποίο έχει ανταλλακτικές χρήσεις(π.χ σε κατασκευή VLSI κυκλωμάτων).

3.1.2 Μοντέλο χωρικών επερωτήσεων

Τα χωρικά δεδομένα χρησιμοποιούν ένα μοντέλο επερωτήσεων με δυο επίπεδα φίλτρων. Το ***πρωτεύον φίλτρο*** και το ***δευτερεύον φίλτρο***. Το *πρωτεύον φίλτρο* έχει σαν είσοδο όλο το σύνολο δεδομένων και σαν έξοδο ένα μικρότερο σύνολο δεδομένων. Η έξοδος του είναι ένα υπερσύνολο της απάντησης που έχει το επερώτημα που έχουμε διατυπώσει. Το *δευτερεύον φίλτρο* έχει σαν είσοδο την έξοδο του πρωτεύοντος φίλτρου και σαν έξοδο την απάντηση στην επερώτηση που έχουμε διατυπώσει. Προκειμένου το δευτερεύον φίλτρο να έχει τα ακριβή αποτελέσματα της επερώτησης που έχουμε θέσει στο σύστημα, χρειάζεται μεγάλη υπολογιστική ισχύς αλλά και πολύς χρόνος. Για τον λόγο αυτό και χρησιμοποιείται το πρωτεύον φίλτρο, το οποίο μειώνει το σύνολο των δεδομένων και άρα μειώνει την απαιτούμενη υπολογιστική ισχύ που θα χρειαζόταν στο δευτερεύον φίλτρο αν δεν είχαμε το πρωτεύον. Η αρχιτεκτονική αυτή εικονίζεται στο πιο κάτω σχήμα (σχήμα 3.1).



Σχήμα 3.1: Μοντέλο χωρικών επερωτήσεων.

3.1.3 Ευρετηριοποίηση χωρικών δεδομένων.

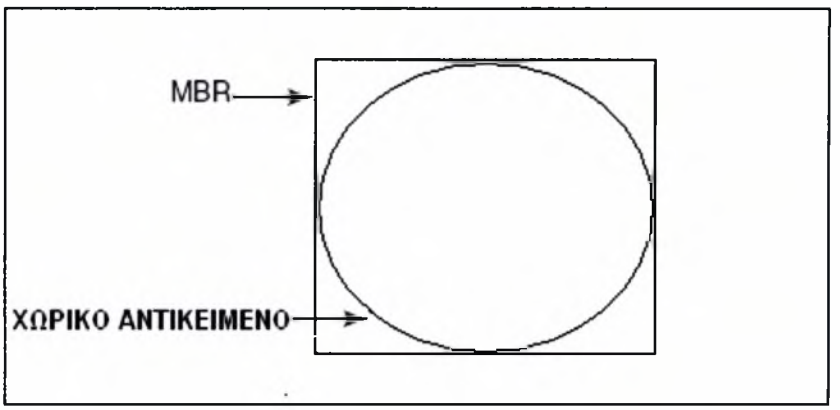
Σε αυτή την ενότητα, θα αναλυθεί η διαδικασία της ευρετηριοποίησης των χωρικών δεδομένων, δεδομένου ότι για την υλοποίηση του πρωτεύοντος φίλτρου χρησιμοποιείται ένα ευρετήριο. Ένα χωρικό ευρετήριο είναι μια δομή για την δεικτοδότηση των χωρικών δεδομένων. Το ευρετήριο είναι ένας μηχανισμός που έχει σαν στόχο, τον περιορισμό του αριθμού των αναζητήσεων. Στη δική μας περίπτωση αυτός ο μηχανισμός βασίζεται σε χωρικά κριτήρια όπως στο αν δυο αντικείμενα τέμνονται, αν το ένα αντικείμενο περιέχει το άλλο, ή κατά πόσο τα δύο αντικείμενα είναι κοντά.

Η θέση ενός χωρικού αντικειμένου προσδιορίζεται με ένα ζευγάρι συντεταγμένων (**x**,**y**). Αυτές οι συντεταγμένες τις περισσότερες φορές είναι ακέραιοι οι οποίοι αποθηκεύονται στην μνήμη του συστήματός μας. Υπάρχουν μερικές επερωτήσεις οι οποίες δεν μπορούν να απαντηθούν μέσω του μηχανισμού της ταυτοποίησης, όπου γίνεται αναζήτηση στο σύνολο των δεδομένων μέχρι να βρεθεί μια συμβατή πλειάδα. Τέτοιες επερωτήσεις είναι οι εξής :

1)Ποια είναι τα πέντε πιο κοντινά φαρμακεία από το σημείο στο οποίο βρίσκεται ένας χρήστης;

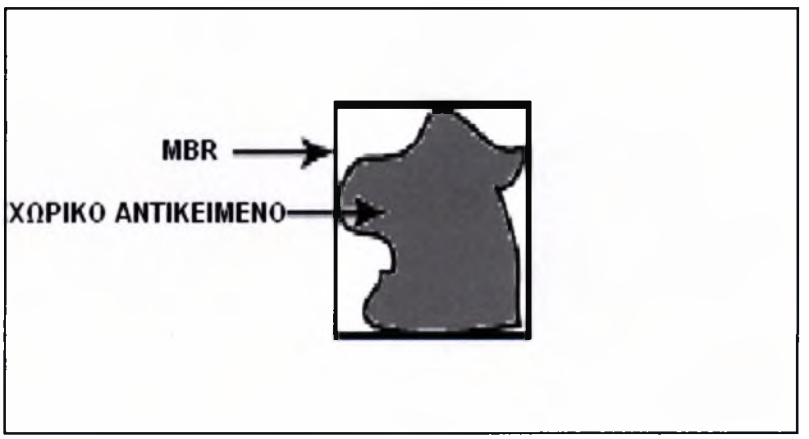
2)Είναι ένα συγκεκριμένο κατάστημα πάνω στην πλατεία; Με άλλα λόγια περιέχει το χωρικό αντικείμενο "πλατεία" το χωρικό αντικείμενο "κατάστημα" ;
Βλέπουμε ότι για τέτοιου είδους επερωτήσεις παίζει ρόλο η σχετική θέση των αντικειμένων στον χώρο.
Έχει αποδειχτεί ότι χωρικές επερωτήσεις όπως αυτές που κάναμε πιο πάνω, σπαταλούν άδικα πόρους του συστήματος μας και ιδιαίτερα πολύ χρόνο, αν οι πληροφορίες των χωρικών αντικείμενων αποθηκεύονται στην μνήμη με τυχαίο τρόπο. Αντίθετα, αν τα χωρικά αντικείμενα ομαδοποιούνται με βάση το πόσο το ένα απέχει από το άλλο, ή με βάση αν το ένα περιέχει το άλλο και αποθηκεύονται στην μνήμη κατά ομάδες, τότε η απόδοση αυξάνεται σημαντικά. Με βάση αυτό το σκεπτικό τα χωρικά αντικείμενα που είναι κοντά, ή το ένα περιέχει το άλλο, αποθηκεύονται σε γειτονικές θέσεις στην μνήμη με αποτέλεσμα η απόδοση να αυξάνεται.
Στα πλαίσια αυτής της εργασίας, για την αποθήκευση των χωρικών δεδομένων, χρησιμοποιήθηκε το ΣΔΒΜ της oracle. Στην oracle υποστηρίζονται δύο τρόποι ευρετηριοποίησης χωρικών αντικείμενων. Ο πρώτος τρόπος είναι με την χρήση των **R-δέντρων**, ο οποίος είναι και ο προκαθορισμένος τρόπος ευρετηριοποίησης (αυτός ο τρόπος χρησιμοποιήθηκε και στην παρούσα εργασία). Ο δεύτερος τρόπος είναι η ευρετηριοποίηση με βάση το τετραδικό δέντρο. Εμείς θα αναλύσουμε τον πρώτο τρόπο, μιας και αυτόν χρησιμοποιήσαμε.

Η ευρετηριοποίηση με βάση τα R-δέντρα, προσεγγίζει το κάθε χωρικό αντικείμενο με ένα ορθογώνιο ελάχιστου εμβαδού, το οποίο περιέχει όλο το χωρικό αντικείμενο. Αυτό το ορθογώνιο λέγεται **"minimum bounding rectangle"** (**MBR**).
Αν το χωρικό αντικείμενο είναι ένας κύκλος τότε το **MBR** είναι το ορθογώνιο με το ελάχιστο εμβαδό που περιέχει τον κύκλο. Ένα κατατοπιστικό παράδειγμα φαίνεται στο ακόλουθο σχήμα(σχήμα 3.2).



Σχήμα 3.2: MBR για το χωρικό αντικείμενο 'κύκλος'.

Το χωρικό αντικείμενο μπορεί να είναι μια οποιαδήποτε γεωμετρία. Πιο κάτω παραθέτουμε ένα παράδειγμα (σχήμα 3.3) όπου η γεωμετρία μας είναι ένα 'σύνθετο' σχήμα.

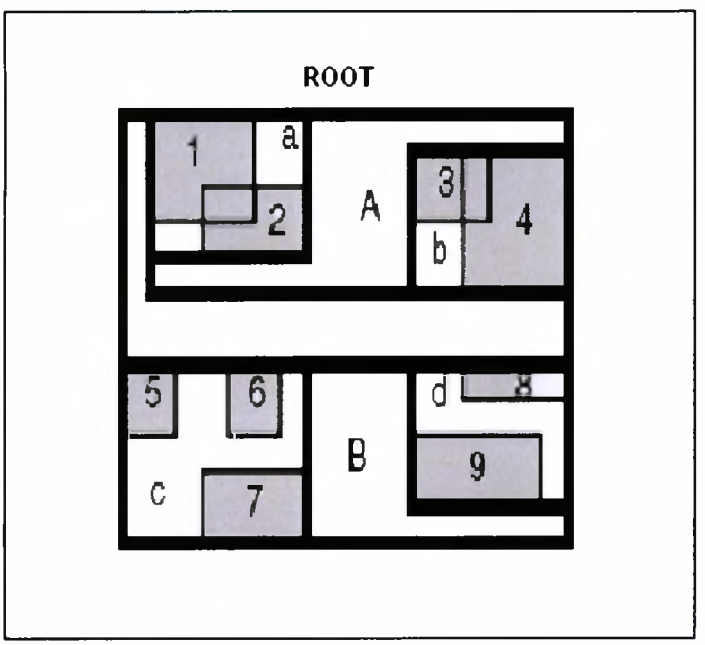


Σχήμα 3.3: MBR για ένα ανομοιόμορφο σχήμα.

Παράδειγμα με R-δέντρο

Σε αυτό το παράδειγμα, υποθέτουμε ότι έχουμε έναν γεωγραφικό χώρο, ο οποίος περιέχει κάποια χωρικά αντικείμενα, και στόχος μας είναι να κατασκευάσουμε το R-δέντρο.

Έστω λοιπόν ότι ο γεωγραφικός χώρος, είναι αυτός που φαίνεται στο πιο κάτω σχήμα (σχήμα 3.4):

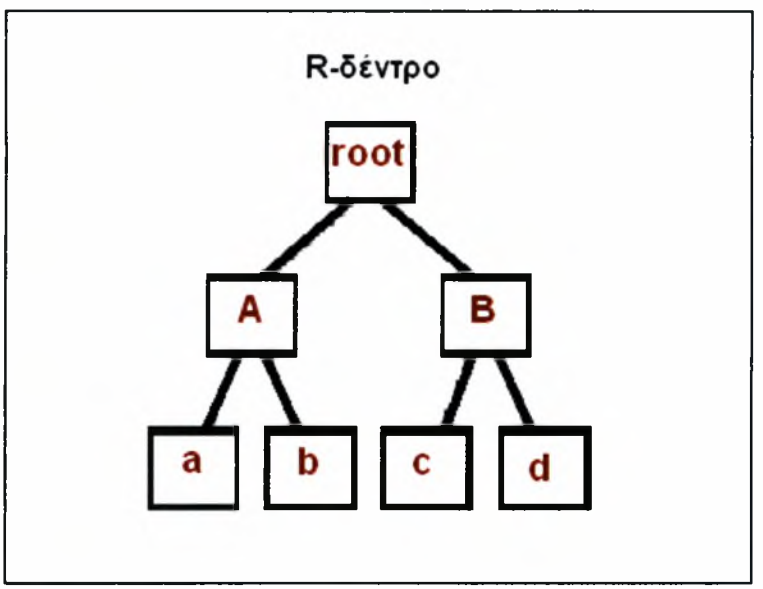


Σχήμα 3.4: Γεωγραφικός χώρος με όνομα: ROOT

Ο γεωγραφικός χώρος ονομάζεται **ROOT** και περιέχει τα χωρικά αντικείμενα **A** και **B**. Το χωρικό αντικείμενο **A** περιέχει τα χωρικά αντικείμενα **a** και **b**. Το **a** περιέχει τα **1** και **2** και το **b** περιέχει τα **3** και **4**. Όμοια και για το χωρικό αντικείμενο **B** και όσα αυτό περιέχει. Το κάθε ένα από όλα τα πιο πάνω χωρικά αντικείμενα μπορεί να έχει οποιαδήποτε γεωμετρία και το κάθε ένα έχει περιγράφει από ένα MBR.

Το R-δέντρο, στην ρίζα του έχει όλο τον γεωγραφικό χώρο **(ROOT)**. Το αμέσως χαμηλότερο επίπεδο στο δέντρο περιέχει εκείνα τα χωρικά αντικείμενα τα οποία

καλύπτουν (με βάση το MBR) όλα τα υπόλοιπα (δηλαδή τα **A** και **B**). Όμοια συνεχίζουμε στο επόμενο χαμηλότερο επίπεδο του δέντρου και σε κάθε ένα από τα χωρικά αντικείμενα **A** και **B**. Στο **A** επεκτείνουμε το δέντρο με τα αντικείμενα **a** και **b** τα οποία περιέχονται στο **A**. Ομοίως και για το **B**.
Το R-δέντρο που προκύπτει εικονίζεται στο σχήμα 3.5:



Σχήμα 3.5: R-δέντρο για τον γεωγραφικό χώρο 'ROOT'.

3.1.4 Υλοποίηση δικτύου στην Oracle

Πριν δώσουμε τον ορισμό του τι είναι δίκτυο και πριν αρχίσουμε την περιγραφή των χαρακτηριστικών και των ιδιοτήτων των δικτύων που μπορούμε να κατασκευάσουμε με την χρήση του ΣΔΒΔ της oracle, θα ορίσουμε με συντομία από τι αποτελείται ένα δίκτυο.

Τα κυριότερα στοιχεία ενός δικτύου, τα οποία μπορούν και να ορίσουν ένα δίκτυο μονοσήμαντα, είναι οι ***κόμβοι(N)*** και οι ***ακμές(E).*** Οι *κόμβοι* είναι αντικείμενα, τα οποία μπορεί να αναπαριστούν κτίρια, πλατείες, πόλεις κτλ. Από την άλλη πλευρά οι *ακμές* αναπαριστούν συσχετίσεις μεταξύ δύο κόμβων. Για να ορίσουμε μια ακμή πρέπει να ορίσουμε τον *κόμβο έναρξης* αλλά και τον *κόμβο προορισμού*. Η κατηγοριοποίηση των ακμών, γίνεται με βάση το αν η φορά κίνησης των αντικειμένων πάνω στην ακμή, παίζει ρόλο ή όχι. Έτσι οι ακμές κατηγοριοποιούνται σε κατευθυνόμενες και μη-κατευθυνόμενες αντίστοιχα.
Ένα ακόμη στοιχείο των δικτύων είναι το μονοπάτι. Το μονοπάτι είναι μια ακολουθία από κόμβους και ακμές, όπου η αρχή και το τέλος του μονοπατιού είναι κόμβοι. Σε ένα μονοπάτι δεν απαγορεύεται να επαναλαμβάνονται κάποιοι κόμβοι ή κάποιες ακμές, ωστόσο κάτι τέτοιο είναι σπάνιο στις περισσότερες εφαρμογές.
Σε αυτό το σημείο, έχοντας περιγράψει τα βασικά συστατικά των δικτύων, είμαστε έτοιμοι να ορίσουμε τι είναι δίκτυο.

Ορισμός 3.2*:** Με τον όρο ***δίκτυο εννοούμε έναν κατευθυνόμενο γράφο, ο οποίος αποτελείται από ***κόμβους(N)*** και ***ακμές(E)***. Οι συσχετίσεις μεταξύ των κόμβων και των ακμών βασίζεται στην έννοια της συνδεσιμότητας. Η συνδεσιμότητα μπορεί να βασίζεται στην έννοια της χωρικής εγγύτητας ή και όχι.

Με την χρήση ενός παραδείγματος θα προσπαθήσουμε να διαχωρίσουμε την περίπτωση που η συνδεσιμότητα βασίζεται στην έννοια της χωρικής εγγύτητας, από την περίπτωση που η συνδεσιμότητα δεν βασίζεται σε αυτή. Για παράδειγμα αν υποθέσουμε ότι δύο πόλεις βρίσκονται απέναντι η μία στη άλλη και ανάμεσα τους έχουν μία λίμνη, τότε το συντομότερο μονοπάτι μεταξύ αυτών των δύο πόλεων, μπορεί να προσδιοριστεί με βάση την χωρική εγγύτητα. Σε αυτή την περίπτωση, το συντομότερο μονοπάτι είναι η ευθεία γραμμή που διασχίζει την λίμνη και συνδέει τις δύο πόλεις. Αντίθετα, αν οι δύο πόλεις συνδέονται με οδικό δίκτυο χρειαζόμαστε να έχουμε πληροφορίες που σχετίζονται με την

συνδεσιμότητα. Με άλλα λόγια θα πρέπει να ξέρουμε πως συνδέονται οι δρόμοι μεταξύ τους, πόσο είναι το κόστος της διάσχισης του κάθε δρόμου-ακμής κτλ.
Για τις ανάγκες αυτής της εργασίας ήταν απαραίτητη η δημιουργία ενός δικτυού. Η λειτουργικότητα που πρέπει να υλοποιηθεί, ταιριάζει περισσότερο με τις εσωτερικές δομές που παρέχει η oracle για την κατασκευή δικτύου, παρά με αυτές που παρέχει για να αποθηκεύει απλές γεωμετρίες (όπως σημεία, γραμμές, πολύγωνα, κτλ).
Σε πολλές εφαρμογές, όπως και σε αυτή που θα αναπτυχθεί στην παρούσα εργασία, είναι πολύ βολικό να μοντελοποιούμε τα αντικείμενα μας σαν ακμές και κόμβους μέσα σε ένα δίκτυο. Ένα δίκτυο περιέχει πληροφορίες όπως: η συνδεσιμότητα των ακμών, το κόστος- βάρος κάθε ακμής και άλλα. Αυτές οι επιπλέον πληροφορίες που ενσωματώνονται στην κατασκευή ενός δικτύου, βοηθούν στην απάντηση μερικών ερωτημάτων όπως :
Α) Ποιος είναι ο πιο σύντομος δρόμος μεταξύ δυο κόμβων του δικτύου μας. Αυτό στον πραγματικό κόσμο μπορεί να αναπαριστά τον πιο σύντομο δρόμο μεταξύ δυο πόλεων.
Β) Ποιοι είναι οι 5 πιο κοντινοί κόμβοι σε σχέση με το σημείο στο οποίο βρισκόμαστε. (π.χ ποια είναι τα 5 πιο κοντινά φαρμακεία).
Γ) Ποια διαδρομή-μονοπάτι μεταξύ της αφετηρίας και του κόμβου προορισμού έχει το ελάχιστο κόστος.

Στο ΣΔΒΔ της oracle οι πληροφορίες σχετικά με τους κόμβους, τις ακμές και τα μονοπάτια αποθηκεύονται σε αντίστοιχους πίνακες. Έτσι έχουμε τον πίνακα των ακμών, τον πίνακα των κόμβων και τον πίνακα των μονοπατιών. Πρέπει να πούμε ότι ο ορισμός των πινάκων ακμών και κόμβων, είναι αρκετός έτσι ώστε να ορίζεται μονοσήμαντα ένα δίκτυο. Ωστόσο υπάρχει η δυνατότητα σε ένα δίκτυο, να ορίσουμε και έναν πίνακα για τα μονοπάτια. Οι πληροφορίες που αποθηκεύονται σε αυτό τον πίνακα, προκύπτουν από τη ανάλυση του δικτύου. Ο πίνακας μονοπατιών πρέπει πάντα να συνοδεύεται από τον ορισμό ενός πίνακα μονοπατιών-ακμών. Αυτό είναι απαραίτητο διότι πρέπει να ορίσουμε ποιες ακμές περιέχονται σε κάθε μονοπάτι. Έτσι ο πίνακας μονοπατιών-ακμών, έχει ένα γνώρισμα στο οποίο αποθηκεύεται η ακολουθία των ακμών που περιέχει κάθε μονοπάτι.

3.1.5 Είδη και Παραδείγματα δικτύων.

Τα δίκτυα χωρίζονται σε δύο μεγάλες κατηγορίες ανάλογα με το είδος των ακμών που υποστηρίζουν. Πιο συγκεκριμένα ένα δίκτυο ανάλογα με το αν αποτελείται από κατευθυνόμενες ή από μη-κατευθυνόμενες ακμές, ονομάζεται κατευθυνόμενο και μη-κατευθυνόμενο αντίστοιχα.

Μια άλλη κατηγοριοποίηση των δικτύων είναι σε λογικά δίκτυα και σε χωρικά δίκτυα (**spatial networks**). Τα λογικά δίκτυα περιέχουν πληροφορίες συνδεσιμότητας αλλά όχι γεωμετρικές πληροφορίες. Αυτού του τύπου τα δίκτυα χρησιμοποιούνται κυρίως για ανάλυση και μόνο των δικτύων. Από την άλλη πλευρά τα χωρικά δίκτυα εμπεριέχουν τόσο πληροφορίες συνδεσιμότητας, όσο και γεωμετρικές πληροφορίες. Σε αυτού του είδους τα δίκτυα, οι κόμβοι και οι ακμές, προκειμένου να αποθηκεύουν γεωμετρικές πληροφορίες, αναπαρίστανται σαν γεωμετρικά αντικείμενα τύπου: **(SDO_GEOMETRY)***, και το δίκτυο ονομάζεται **SDO network.**

Τα χωρικά δίκτυα έχουν εφαρμογή σε πολλούς τομείς. Ένα πολύ κοινό παράδειγμα χωρικού δικτύου είναι ένα οδικό δίκτυο. Ένα τέτοιο δίκτυο, αποτελείται από κόμβους που αναπαριστούν κτίρια, πλατείες, κτλ, αλλά και από ακμές που αναπαριστούν δρόμους. Οι πίνακες των ακμών και των κόμβων ορίζονται όπως είπαμε πιο πάνω και έχουν χωρική υπόσταση. Ένα άλλο παράδειγμα είναι

ένα καλωδιακό δίκτυο. Σε αυτά τα δίκτυα σημαντική παράμετρος είναι η ελαχιστοποίηση του κόστους. Αυτό μπορεί να υπολογιστεί με βάση το κόστος της κάθε ακμής και βρίσκοντας το ελάχιστο επικαλύπτον δέντρο. Τέλος ένα δίκτυο μπορεί να χρησιμοποιηθεί για να προσομοιώσει βιοχημικές αντιδράσεις σε διάφορους οργανισμούς. Για παράδειγμα δύο κόμβοι να αναπαριστούν δυο πρωτεΐνες και η ακμή που τις ενώνει μπορεί να αναπαριστά την αλληλεπίδραση μεταξύ των δυο πρωτεϊνών.

* SDO_GEOMETRY είναι ένας τύπος δεδομένων που χρησιμοποιεί το ΣΔΒΔ oracle για να αποθηκεύει χωρικά δεδομένα. Όταν ένα δίκτυο έχει τέτοιου είδους δεδομένα λέγεται χωρικό δίκτυο (**spatial network**) .

3.2 Χρονικά δεδομένα

Είναι γεγονός ότι τα χρονικά δεδομένα χρησιμοποιούνται σε πολλές εφαρμογές. Για παράδειγμα σε Τραπεζικές εφαρμογές, χρησιμοποιούνται για την καταγραφή των συναλλαγών που γίνονται καθημερινά, καταγράφοντας την χρονική στιγμή που έγινε κάθε συναλλαγή. Χρησιμοποιούνται επίσης σε πολλές επιχειρήσεις αλλά και οργανισμούς. Για παράδειγμα, τα χρονικά δεδομένα χρησιμοποιούνται σε επιχειρήσεις για την καταγραφή και διαχείριση των αποθεμάτων στις αποθήκες. Τέλος τα χρονικά δεδομένα σχετίζονται και με επιστημονικές εφαρμογές, προκείμενου να γίνουν μετρήσεις, αλλά και να βγουν συμπεράσματα για ήδη υπάρχοντα συστήματα, ή και για ενδεχομένως μελλοντικά.

Για την αποθήκευση και χειρισμό των χρονικών δεδομένων, χρησιμοποιήθηκε το ΣΔΒΔ oracle. Για τις επερωτήσεις πάνω σε αυτού του τύπου τα δεδομένα, χρησιμοποιήθηκε η γλωσσά SQL.

Πιο συγκεκριμένα χρειάστηκε ο συνδυασμός των χρονικών δεδομένων με τα χωρικά δεδομένα, προκειμένου να υλοποιηθεί η επιθυμητή λειτουργικότητα. Ειδικότερα, υπήρξε η ανάγκη πάνω στο δίκτυο που κατασκευάστηκε, να παραχθούν κινούμενοι χρήστες οι οποίοι να έχουν χωρική και χρονική υπόσταση. Άρα προέκυψε η ανάγκη παραγωγής χώρο-χρονικών αντικειμένων. Την έννοια αυτών των αντικειμένων την περιγράφουμε στο επόμενο κεφάλαιο.

ΚΕΦΑΛΑΙΟ 4

4. Χώρο-χρονικά δεδομένα

Σε αυτό το κεφάλαιο θα οριστούν και θα αναλυθούν τα χώρο-χρονικά δεδομένα. Αυτού του είδους τα δεδομένα συνδυάζουν τις πληροφορίες που παρέχουν τα χωρικά αλλά και τα χρονικά δεδομένα. Έτσι θα είναι δυνατόν να οριστούν κινούμενοι χρήστες, οι οποίοι αναφέρονται σαν χώρο-χρονικά αντικείμενα. Τα χώρο-χρονικά αντικείμενα περιγράφονται αναλυτικά στην επόμενη ενότητα.

4.1 Χώρο-χρονικά αντικείμενα

Σε αυτή την ενότητα, θα αναλυθεί η σημασία των ***χώρο-χρονικών αντικείμενων***, τα οποία χρησιμοποιούνται στην παρούσα εργασία και η κατανόηση τους είναι σημαντική. Τα *χώρο-χρονικά αντικείμενα* χαρακτηρίζονται από:

1) Ένα αναγνωριστικό: **(obj.id)**

2) Για κάθε χρονική στιγμή **(obj.time**$_i$**)** έχουν και μια διαφορετική θέση στον χώρο **(obj.location**$_i$**).**

Βλέπουμε λοιπόν ότι αυτού του είδους τα δεδομένα εμπεριέχουν και την έννοια του χρόνου. Άρα τα δεδομένα αυτού του είδους, δεν έχουν μια και μοναδική θέση στον χώρο, αλλά μια θέση που μεταβάλλεται σταδιακά με την πάροδο του χρόνου. Το μοντέλο που χρησιμοποιείται για την περιγραφή των χρονικών περιορισμών των χώρο-χρονικών αντικείμενων έχει ως εξής: Πρώτα απ' όλα ορίζεται η περίοδος **(T).** Ο ορισμός της περιόδου, έχει σαν στόχο να προσδιοριστεί το χρονικό διάστημα μέσα στο οποίο οι χρήστες έχουν την δυνατότητα να κινούνται.

Η περίοδος (T) καθορίζεται από δυο αριθμούς: **(t_{min} t_{max}).** Με t_{min} ορίζουμε την μικρότερη χρονοσφραγίδα που μπορεί να πάρει ένα αντικείμενο και με t_{max} την μέγιστη. Πρέπει εδώ να σημειώσουμε ότι ο χρόνος, αν και στον πραγματικό κόσμο είναι συνεχής, εδώ θεωρείται σαν ένα διακριτό γνώρισμα.

Κάθε αντικείμενο (**object**$_i$) κινείται μεταξύ δύο διαδοχικών χρονοσφραγίδων t_i και t_{i+1}, και κινείται από την παρούσα θέση του **(obj.location**$_i$**)** στην καινούρια του θέση **(obj.location**$_{i+1}$**)**. Στη γενική περίπτωση, αυτές οι θέσεις που υπολογίζονται, δεν ταυτίζονται με τους κόμβους του δικτύου. Με άλλα λόγια οι συντεταγμένες **(x,y)** που υπολογίζονται, μερικές φορές δεν αντιστοιχούν σε κόμβους του δικτύου. Προκειμένου να αντιστοιχίσουμε αυτές τις συντεταγμένες σε πραγματικούς κόμβους, υπάρχουν διάφορες τεχνικές, τις οποίες θα αναλύσουμε διεξοδικά στη συνέχεια .

4.2 Παραγωγή χώρο-χρονικών αντικείμενων με την χρήση γνωστών προγραμμάτων(Generators).

Μετά την κατασκευή του χωρικού δικτύου στην oracle, το επόμενο βήμα είναι πάνω σε αυτό το δίκτυο που κατασκευάστηκε, να δημιουργηθούν κινούμενα αντικείμενα. Αυτά τα κινούμενα αντικείμενα αντιπροσωπεύουν τους χρήστες της εφαρμογής που θα αναπτυχθεί. Καθώς οι χρήστες αποτελούν κινούμενα αντικείμενα πάνω στο χωρικό δίκτυο που ορίσαμε, η κίνηση του κάθε χρήστη στο δίκτυο, περιγράφεται με την βοήθεια της θέσης που έχει την εκάστοτε χρονική στιγμή. Αυτό καθίσταται δυνατό με την χρήση των χώρο-χρονικών δεδομένων. Για την παραγωγή αυτών των χώρο-χρονικών δεδομένων, χρησιμοποιούνται ειδικά προγράμματα(Generators).

Υπάρχουν πολλές προτάσεις σχετικά με την παραγωγή χώρο-χρονικών αντικειμένων και είναι πολλοί αυτοί που έχουν προτείνει προγράμματα για την παραγωγή χώρο-χρονικών αντικειμένων. Ένα από τα πιο γνωστά προγράμματα για την παραγωγή χώρο-χρονικών δεδομένων, έχει προταθεί από τον **Thomas Brinkhoff.** Αυτό το πρόγραμμα χρησιμοποιείται ευρέως από πολλούς επιστήμονες για ερευνητικές εργασίες. Αυτό είναι και το πρόγραμμα που θα χρησιμοποιηθεί και στα πλαίσια αυτής της εργασίας, προκειμένου να παραχθούν τα χώρο-χρονικά αντικείμενα της εφαρμογής που θα αναπτυχθεί. Σαν είσοδο αυτό το πρόγραμμα δέχεται ένα δίκτυο, το οποίο είναι σε κατάλληλη μορφή-κωδικοποίηση. Τέτοια δίκτυα παρέχονται στο διαδίκτυο. Ένα από αυτά είναι το

δίκτυο **'oldenburg'** (σχήμα 4.6) το οποίο και θα χρησιμοποιηθεί σε αυτή την εργασία.

Ένα ακόμη γνωστό πρόγραμμα παραγωγής χώρο-χρονικών δεδομένων, είναι το ονομαζόμενο:'**Oporto generator'** το οποίο δημοσιεύτηκε το 1999. Αυτό το πρόγραμμα έχει αναπτυχθεί από τους José Moreira & Jean-Marc Saglio. Αυτό το πρόγραμμα αναπτύχθηκε για να μοντελοποιήσει την κίνηση των αλιευτικών πλοίων. Η κίνηση των πλοίων δεν είναι τυχαία, αλλά καθορίζεται από το σημείο εκκίνησης τους(λιμάνι), καθώς και από τον τελικό προορισμό τους. Ο τελικός προορισμός των πλοίων, όπως είναι φυσικό είναι μια περιοχή στην οποία υπάρχουν πολλά κοπάδια ψαριών. Για να είναι όσο το δυνατόν πιο ρεαλιστικές οι τροχιές που διαγράφουν τα κινούμενα αντικείμενα(πλοία) που παράγει το πρόγραμμα, έχουν γίνει πειράματα σε πραγματικές συνθήκες και έχουν εξαχθεί αποτελέσματα τα οποία χρησιμοποιούνται κατά την παραγωγή των χώρο-χρονικών αντικειμένων(πλοία). Τα πειράματα που έχουν γίνει έχουν διεξαχθεί κάτω από όλες τις δυνατές συνθήκες, έτσι ώστε να είναι δυνατή η πρόβλεψη της κίνησης των ψαριών και κατ' επέκταση και των πλοίων. Μερικοί παράμετροι οι οποίοι επηρεάζουν την κίνηση των ψαριών είναι για παράδειγμα: οι καιρικές συνθήκες, η εποχή, η θερμοκρασία, η μορφολογία του υποθαλάσσιου κόσμου κτλ. Όπως είναι προφανές, το συγκεκριμένο πρόγραμμα, έχει σχεδιαστεί για την παραγωγή ενός πολύ εξειδικευμένου τύπου χώρο-χρονικών αντικειμένων(πλοία), και άρα δεν ταιριάζει με την λειτουργικότητα που επιθυμούμε να αναπτύξουμε σε αυτή την εργασία.

Σε σχέση με την πιο πάνω πρόταση για την παραγωγή χώρο-χρονικών αντικειμένων, η πρόταση των Daniel Krajzewicz & Christian Rössel ήταν πολύ πιο ελκυστική. Το πρόγραμμα που αναπτύχθηκε από τους παραπάνω ονομάστηκε: **'SUMO generator'** και η πρώτη του έκδοση παρουσιάστηκε το 2000. Η πρώτη έκδοση είχε αναπτυχθεί με στόχο να προσομοιώσει την εναέρια κίνηση σε αεροδρόμια, στην συνέχεια όμως το πρόγραμμα αναπτύχθηκε περαιτέρω και προστέθηκαν πακέτα προκειμένου να είναι δυνατή η προσομοίωση της κίνησης οχημάτων στα όρια μια πόλης. Αυτό το πρόγραμμα, είναι open source έτσι ώστε να αναβαθμίζεται συνεχώς και για αυτό τον λόγο έχει και πολύ καλές ιδιότητες.

Το συγκεκριμένο πρόγραμμα παραγωγής χώρο-χρονικών αντικειμένων, αρχικά είχε γραφτεί σε C++, και έχουν χρησιμοποιηθεί βιβλιοθήκες οι οποίες είναι συμβατές με τα περισσότερα συστήματα, έτσι ο κώδικας έχει υψηλή μεταφερσιμότητα. Επιπλέον το πρόγραμμα αυτό υποστηρίζει πολλούς τύπους κινούμενων αντικειμένων, έτσι για παράδειγμα μπορούν να παράγονται τροχιές παράλληλα για αυτοκίνητα, φορτηγά, μηχανές κτλ . Επίσης τα δίκτυα που δέχεται σαν είσοδο, μπορούν να είναι σε πολλές εναλλακτικές μορφές-κωδικοποιήσεις (π.χ ArcView, XML-Descriptions κτλ). Τέλος ένα από τα βασικά του πλεονεκτήματα είναι ότι υποστηρίζει δυναμική δρομολόγηση των κινούμενων αντικειμένων. Αυτό σημαίνει ότι δεν δίνεται σε κάθε όχημα εξ' αρχής η διαδρομή που πρέπει να ακολουθήσει προκειμένου να φτάσει στον τελικό του προορισμό, αλλά αντίθετα δίνονται σταδιακά οδηγίες, σχετικά με το ποιο μονοπάτι πρέπει να ακολουθήσει ένας οδηγός, προκειμένου να φτάσει στον προορισμό του.

Σημαντικό ρόλο στην παραγωγή των χώρο-χρονικών αντικειμένων, με την χρήση του συγκεκριμένου προγράμματος, παίζουν οι δυνατότητες του κάθε οχήματος, αλλά και του κάθε οδηγού. Αυτά τα στοιχεία μπορούν να προκύψουν από παρατηρήσεις που γίνονται σε κάθε όχημα ξεχωριστά. Πέρα από αυτό, στην παραγωγή την χώρο-χρονικών αντικειμένων, σημαντικό ρόλο παίζει ο χρόνος που θα κάνει το κάθε όχημα να φτάσει στον τελικό του προορισμό. Ο χρόνος αυτός σχετίζεται με πολλές παραμέτρους όπως: α)Ποιο είναι το συντομότερο μονοπάτι, β)Αν το όχημα θα διασχίσει δρόμους ταχείας κυκλοφορίας και άρα θα αναπτύξει μεγάλη ταχύτητα, γ)Αν κάποιοι δρόμοι έχουν κίνηση κτλ.

Είναι προφανές ότι το συγκεκριμένο πρόγραμμα είναι αρκετά ικανοποιητικό, παρόλα αυτά δεν χρησιμοποιήθηκε σε αυτή την εργασία για τους εξής λόγους. Πρώτα απ' όλα το πρόγραμμα αυτό, είναι open source και αναβαθμίζεται συνεχώς με αποτέλεσμα να έχουν προστεθεί αρκετά καινούρια κομμάτια κώδικα που δεν είναι γραμμένα σε C++. Για να τρέξουν χρειάζεται να υπάρχουν εγκατεστημένα τα συστήματα Python και Perl. Το γεγονός ότι το πρόγραμμα για να τρέξει χρειάζεται τρία πακέτα λογισμικού εγκατεστημένα αυξάνει αρκετά την πολυπλοκότητα. Επιπρόσθετα το συγκεκριμένο πρόγραμμα, έχει σχεδιαστεί με βασικό στόχο την απόδοση. Για αυτόν τον λόγο και είναι γραμμένος σε ανεξάρτητα κομμάτια κώδικα που το κάθε ένα έχει διαφορετικό ρόλο. Για να εκτελεστεί το κάθε ανεξάρτητο κομμάτι κώδικα, πρέπει να γίνει compile σε κάθε

κομμάτι ξεχωριστά και να εκτελεστεί ξεχωριστά. Αυτή η λογική από την μια μεριά ελαττώνει την απαιτούμενη υπολογιστική ισχύ, αφού κάθε φορά εκτελούνται μόνο τα απολύτως αναγκαία κομμάτια κώδικα, αλλά από την άλλη πλευρά όμως δεν μπορούν να γίνουν πολλά πράγματα παράλληλα, με αποτέλεσμα η διαδικασία να γίνεται σημαντικά πολύπλοκη.

4.3 Παραγωγή χώρο-χρονικών αντικείμενων με την χρήση του προγράμματος του Thomas Brinkhoff.

Αφού περιγράψαμε κάποια προγράμματα παραγωγής χώρο-χρονικών δεδομένων και εξηγήσαμε τους λόγους για τους οποίους δεν ταιριάζουν με την εφαρμογή που θα αναπτυχθεί σε αυτή την εργασία, σε αυτή την ενότητα θα περιγράψουμε το πρόγραμμα που κρίθηκε ως καταλληλότερο για την εφαρμογή που θα αναπτυχθεί. Ως καταλληλότερο πρόγραμμα παραγωγής χώρο-χρονικών δεδομένων λοιπόν, κρίθηκε το πρόγραμμα του **Thomas Brinkhoff**, του οποίου η πρώτη έκδοση δημοσιεύτηκε το 2002.

Ο τρόπος παραγωγής των χώρο-χρονικών αντικειμένων, που παράγονται από το πρόγραμμα αυτό, βασίζεται κυρίως σε μετρήσεις που έχουν γίνει πάνω σε πραγματικά συστήματα. Για παράδειγμα σε ένα σύστημα GPS έχουν γίνει μετρήσεις σχετικά με την θέση από όπου οι χρήστες κάνουν τις αιτήσεις τους, καθώς και για τα μεσοδιαστήματα μεταξύ δυο διαδοχικών αιτήσεων. Με βάση αυτά τα στοιχεία, το πρόγραμμα προσπαθεί να παράγει κινούμενα αντικείμενα των οποίων τα χαρακτηριστικά προσεγγίζουν τον πραγματικό κόσμο. Πέρα από αυτά τα πειραματικά δεδομένα, τα οποία έχουν προκύψει από μετρήσεις, χρησιμοποιούνται και κάποιες βασικές κατανομές από την θεωρία πιθανοτήτων, όπως η Gaussian.

Όσον αφορά τις μετρήσεις που έχουν γίνει, θα χρησιμοποιήσουμε ένα παράδειγμα για να δείξουμε πως αυτές βοηθούν έτσι ώστε τα δεδομένα που παράγονται να είναι ρεαλιστικά, όσον αφορά την προσομοίωση της κίνησης των αντικειμένων.

Έστω ότι έχουμε ορίσει ένα δίκτυο το οποίο αναπαριστά μια πόλη και τα κινούμενα αντικείμενα αντιπροσωπεύουν οχήματα. Το κάθε όχημα δεν επιλέγει τυχαία την πορεία του, αλλά η πορεία του καθορίζεται με βάση τον προορισμό του. Έτσι με βάση κάποια κριτήρια όπως: **1)**Ποιος είναι ο πιο σύντομος δρόμος για να φτάσει στον προορισμό του ένα όχημα, **2)**Αν από την διαδρομή που θα ακολουθήσει θα διασχίσει έναν δρόμο ταχείας κυκλοφορίας και έτσι θα αναπτύξει μεγαλύτερη ταχύτητα,(το πόσο ταχύτητα μπορεί να αναπτύξει ένα όχημα σε ένα συγκεκριμένο δρόμο-ακμή του δικτύου μπορεί να καθορίζεται από ένα κατώφλι που έχει οριστεί για κάθε ακμή), **3)** Αν σε αυτό το μονοπάτι από την αφετηρία προς τον προορισμό υπάρχει μεγάλη κίνηση και άρα δεν είναι δυνατόν το όχημα να αναπτύξει μεγάλη ταχύτητα(η κίνηση μπορεί να προσδιοριστεί από το capacity της κάθε ακμής), **4)**Εξωγενείς παράγοντες, όπως για παράδειγμα οι συνθήκες καιρού που επικρατούν, οι οποίες επηρεάζουν την κίνηση των οχημάτων. Αυτοί και άλλοι πολλοί παράγοντες μπορούν να ληφθούν υπόψη, έτσι ώστε σε μια προσομοίωση ενός συστήματος, το πρόγραμμα(generator) να παράγει κινούμενα αντικείμενα των οποίων η κίνηση να προσεγγίζει όσον το δυνατόν περισσότερο την πραγματικότητα.

Ένας από τους πιο κύριους αλγόριθμους που χρησιμοποιεί το πρόγραμμα του Thomas Brinkhoff, είναι ο **GSTD algorithm** (**Generate SpatioTemporal Data**). Αυτός ο αλγόριθμος αρχίζει την παραγωγή των δεδομένων με βάση την Gaussian κατανομή . Στην συνέχεια, αυτά τα δεδομένα τροποποιούνται με την χρήση κάποιων random συναρτήσεων, οι οποίες είναι παραμετροποιημένες κατάλληλα με βάση τις μετρήσεις από τα πειράματα . Τέλος αν ένα αντικείμενο έχει κινηθεί έξω από τα προκαθορισμένα όρια θεωρείται άκυρο και απορρίπτεται. Το ίδιο γίνεται και αν ένα αντικείμενο έχει κινηθεί έκτος των χωρικών ορίων, δηλαδή αν έχει για κάποια χρονική στιγμή συντεταγμένες που δεν είναι επιτρεπτές με βάση τον χάρτη που έχουμε ορίσει.

Είναι αναγκαίο να τονίσουμε ότι η ενσωμάτωση στοιχείων από την στατιστική είναι αναγκαία, διότι αν βασιζόμαστε μόνο σε πειραματικά δεδομένα από ένα συγκεκριμένο σύστημα, τα δεδομένα τα οποία θα παράγονταν για την δική μας εφαρμογή ίσως να μην ήταν ικανοποιητικά. Έτσι το πρόγραμμα συνδυάζει τα πειραματικά δεδομένα- μετρήσεις με κάποια στοιχεία από την στατιστική και

προσπαθεί να παράγει δεδομένα, τα οποία να είναι κατάλληλα για κάθε είδους εφαρμογή και παράλληλα να είναι ρεαλιστικά.

4.3.1 Καθορισμός του κόμβου έναρξης.

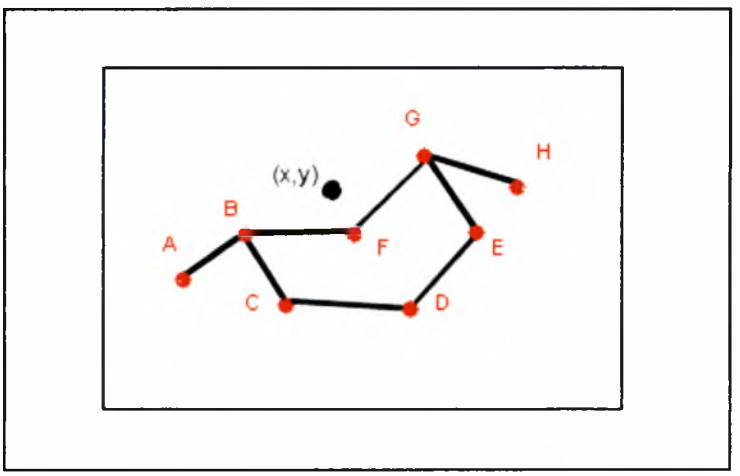
Πριν ένα αντικείμενο αρχίσει να κινείται και αρχίσει να δημιουργεί τα διάφορα στιγμιότυπα του, πρέπει πρώτα να καθοριστεί η αρχική θέση του αντικειμένου. Η βασική ιδέα για τον καθορισμό της αρχικής "θέσης-κόμβου" είναι να επιλεγεί ο πλησιέστερος γείτονας. Η αρχική του θέση είναι συνήθως ένας κόμβος του δικτύου και γι' αυτό ονομάζεται κόμβος έναρξης. Υπάρχουν τρεις προσεγγίσεις σχετικά με τον προσδιορισμό του κόμβου έναρξης:

1. Προσέγγιση που βασίζεται στον χώρο δεδομένων (DSO)

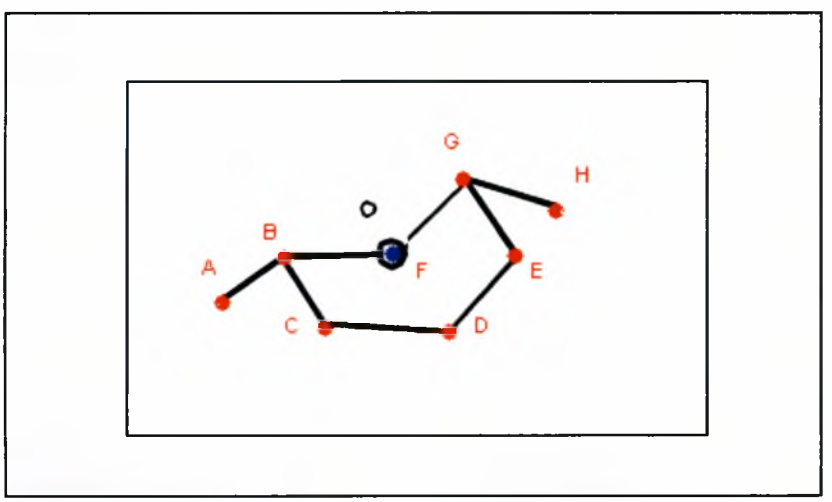
Σύμφωνα με αυτή την προσέγγιση ο κόμβος έναρξης καθορίζεται από την πυκνότητα του δικτύου που έχει δοθεί σαν είσοδο στο πρόγραμμα. Έτσι **τα περισσότερα αντικείμενα** έχουν σαν κόμβους έναρξης, κόμβους όπου το δίκτυο είναι **αραιό** και αντίθετα λιγότερα αντικείμενα θεωρούν σαν κόμβους έναρξης κάποιον κόμβο, όπου εκεί το δίκτυο είναι πυκνό.

Έτσι αν έχει παραχθεί για ένα σημείο το ζευγάρι συντεταγμένων **(x,y)** το οποίο δεν αναπαριστά έναν κόμβο του δικτύου μας, υπολογίζεται ο κοντινότερος γείτονας(υπαρκτός κόμβος του δικτύου) και θεωρείται σαν κόμβος έναρξης.

Στο σχήμα 4.1 απεικονίζεται ο υπολογισμός της θέσης ενός σημείου με συντεταγμένες **(x,y),** το οποίο δεν αποτελεί κόμβο του δικτύου. Στο σχήμα 4.2 απεικονίζεται η διαδικασία σύμφωνα με την οποία ένα σημείο με συντεταγμένες **(x,y)** το οποίο δεν αποτελεί κόμβο του δικτύου, αντιστοιχίζεται με τον πιο κοντινό-γειτονικό κόμβο**('F').**



Σχήμα 4.1: Υπολογισμός της θέσης ενός σημείου.



Σχήμα 4.2: Αντιστοίχηση στον πιο κοντινό γείτονα.

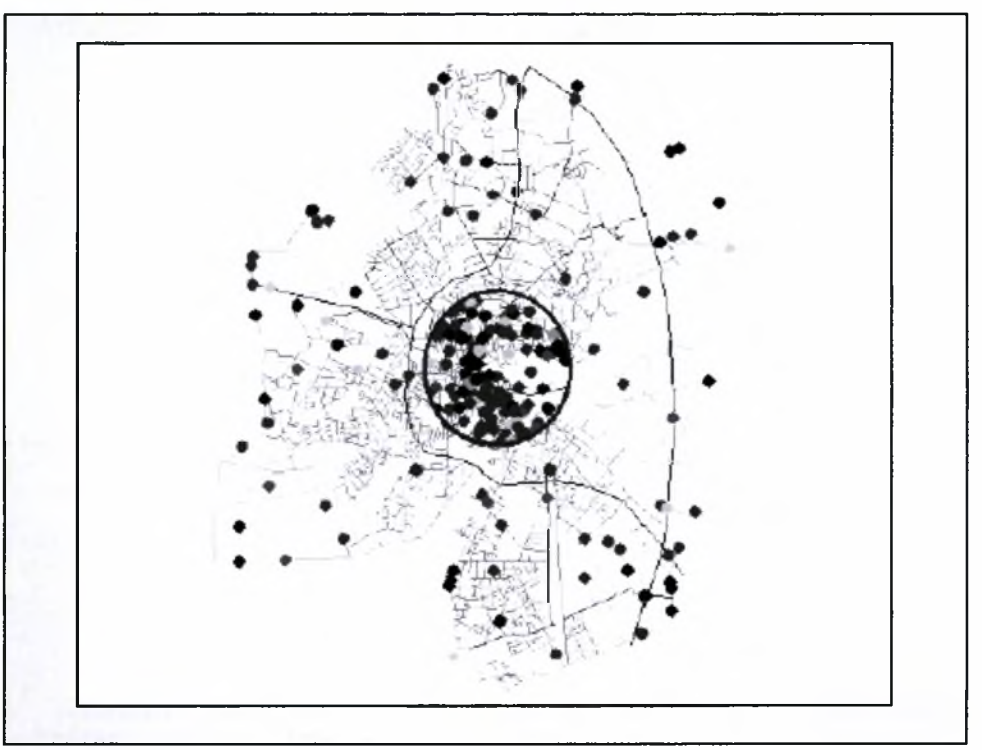
2. Μέθοδος που βασίζεται στην περιοχή (RB).

Μια μέθοδος που βελτιώνει την προηγούμενη προσέγγιση, η οποία βασίζονταν στον χώρο δεδομένων, είναι η μέθοδος που βασίζεται στην περιοχή .Με βάση αυτή την μέθοδο προσπαθούμε να προσαρμόσουμε την κατανομή των δεδομένων στο δίκτυο. Αυτό μπορεί να γίνει αν εισάγουμε κάποια επιπλέον στατιστικά στοιχεία. Έτσι κάθε περιοχή του δικτύου περιγράφεται από μια

πιθανότητα η οποία εκφράζει κατά πόσο είναι πιθανό αυτή η περιοχή να περιέχει τον κόμβο έναρξης .

Για παράδειγμα προκειμένου να προσδιορίσουμε αυτή την πιθανότητα, μπορούμε να βασιστούμε στην πυκνότητα του πληθυσμού, σε κάθε περιοχή που έχουμε ορίσει για το δικτυού μας. Όπως και στην προηγούμενη προσέγγιση αφού υπολογιστεί μια θέση **(x,y)**, η οποία είναι πιο πιθανό να βρίσκεται σε μια περιοχή του δικτύου που έχει υψηλή πυκνότητα πληθυσμού, ο κοντινότερος κόμβος καθορίζεται σαν κόμβος εκκίνησης.

Στο σχήμα 4.3 φαίνεται ένας χάρτης όπου στο κέντρο της πόλης, το οποίο αποτελεί μια περιοχή, η πιθανότητα να έχουμε έναν κόμβο εκκίνησης σε αυτή την περιοχή είναι πολύ μεγαλύτερη απ' ότι σε γύρω περιοχές .

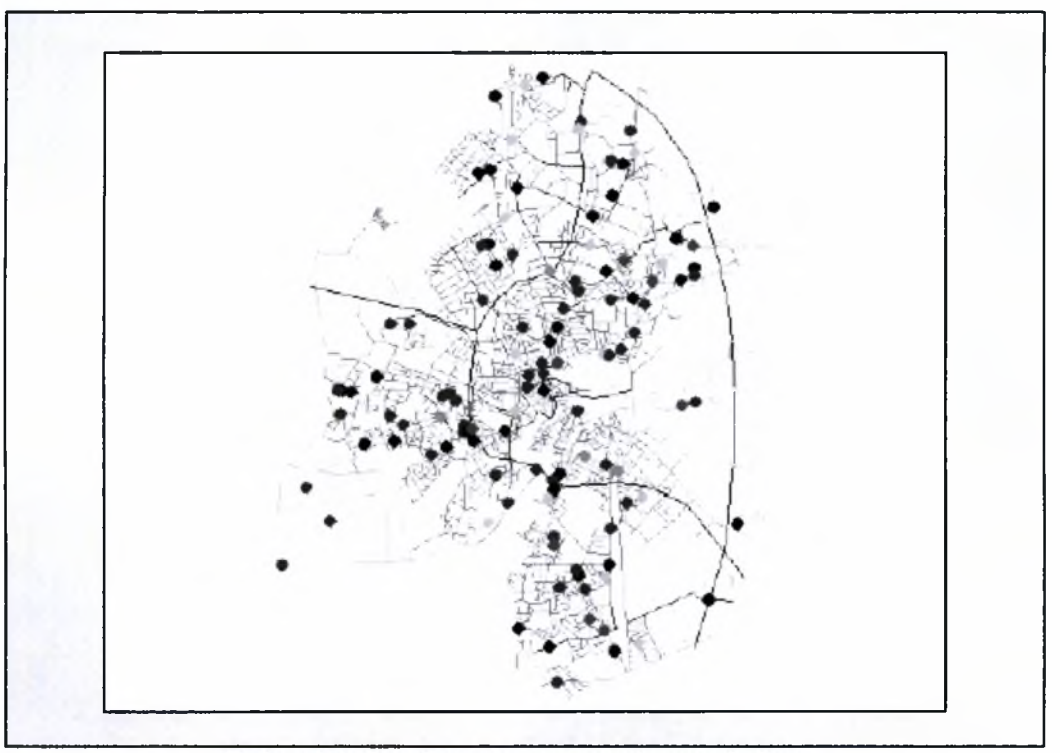


Σχήμα 4.3:Ο κύκλος αναπαριστά την περιοχή του κέντρου της πόλης.

3. Μέθοδος που βασίζεται στο δίκτυο (NB).

Με βάση την προσέγγιση που βασίζεται στο δίκτυο, επιλέγουμε τον κόμβο εκκίνησης θεωρώντας μια ομοιόμορφη κατανομή των αντικειμένων. Συνεπώς κάθε κόμβος του δικτύου έχει την ίδια πιθανότητα να επιλεγεί σαν κόμβος

εκκίνησης. Έτσι σε αντίθεση με την προηγούμενη προσέγγιση, η κατανομή των κόμβων εκκίνησης δεν εξαρτάται από την πυκνότητα της κάθε περιοχής ξεχωριστά, αλλά από την πυκνότητα του δικτύου συνολικά. Στο σχήμα 4.4 φαίνεται ο χάρτης μια πόλης και οι θέσεις εκκίνησης των κινούμενων αντικειμένων. Παρατηρούμε ότι οι κόμβοι έχουν κατανεμηθεί ομοιόμορφα στον χώρο.



Σχήμα 4.4: Ομοιόμορφη κατανομή των κόμβων.

4.3.2 Υπολογισμός της διαδρομής των χώρο-χρονικών αντικειμένων.

Ο υπολογισμός της διαδρομής γίνεται με τέτοιον τρόπο, ώστε κάθε ένα κινούμενο αντικείμενο να φτάσει στον τελικό προορισμό του, όσο το δυνατόν πιο γρήγορα. Η ιδέα που έχει προταθεί για να επιτευχθεί αυτός ο στόχος, βασίζεται στον υπολογισμό του συντομότερου μονοπατιού από την αφετηρία μέχρι τον τελικό προορισμό. Αυτό υλοποιείται με την εφαρμογή ενός αλγορίθμου δρομολόγησης, όπως είναι ο Dijkstra. Πρέπει να σημειωθεί ότι όπως όλοι οι αλγόριθμοι δρομολόγησης, έτσι και ο Dijkstra για να τρέξει απαιτεί το γράφημα να είναι συνεκτικό. Λέγοντας συνεκτικό εννοούμε να υπάρχει μονοπάτι μεταξύ δύο οποιοδήποτε κόμβων του γραφήματος.

Ένα ακόμα θέμα που προκύπτει είναι το γεγονός ότι ο αλγόριθμος δρομολόγησης τρέχει κατά την παραγωγή των κινούμενων αντικείμενων. Έτσι γίνεται μια υπόθεση που δεν είναι και τόσο ρεαλιστική. Δηλαδή θεωρείται ότι ένα αντικείμενο δεν αλλάζει ταχύτητα όσο βρίσκεται πάνω σε μια ακμή του δικτύου μας. Μία βελτίωση που θα μπορούσε να γίνει είναι να μην υπολογιζόταν από την αρχή το μονοπάτι προς τον τελικό προορισμό, αλλά να υπολογιζόταν εκ' νέου σε κάθε μια χρονική στιγμή. Δηλαδή αν την χρονική στιγμή t_i είχε υπολογιστεί το μονοπάτι **A** την χρονική στιγμή t_{i+1} υπολογίζεται ένα νέο μονοπάτι το οποίο μπορεί να συμπίπτει με το μονοπάτι **A,** αλλά μπορεί να είναι και ένα νέο, εντελώς διαφορετικό μονοπάτι.

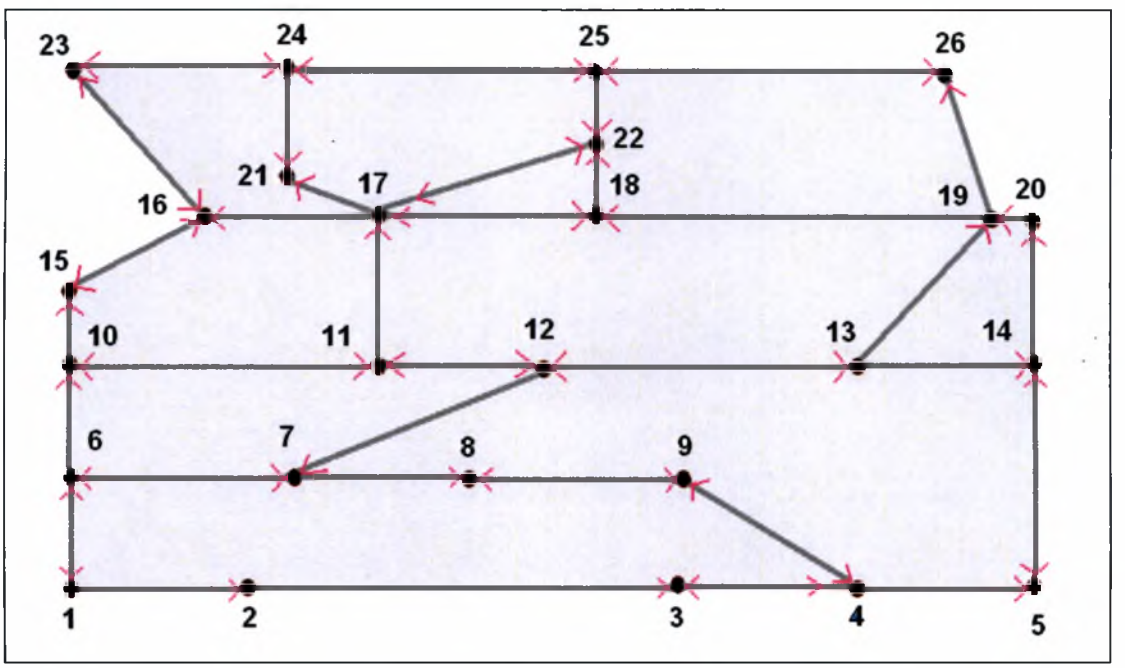
Κάτι τέτοιο δεν γίνεται στα πραγματικά συστήματα όπως για παράδειγμα στο GPS. Σε όλες τις πραγματικές εφαρμογές το μονοπάτι καθορίζεται εκ' των προτέρων, γιατί ο υπολογισμός νέου μονοπατιού σε κάθε μία διαφορετική χρονική στιγμή απαιτεί τεραστία υπολογιστική ισχύ. Ωστόσο η αλλαγή του αρχικού μονοπατιού που έχει υπολογιστεί μπορεί να γίνει, αλλά μόνο σε ειδικές περιπτώσεις. Ένα παράδειγμα είναι αν σε κάποιο κομμάτι του δρόμου έχει συμβεί ένα ατύχημα και ο δρόμος είναι κλειστός.

4.4 Κατασκευή του χωρικού δικτύου 'MYNET' και φόρτωση του στο πρόγραμμα του Thomas Brinkhoff.

Για τις ανάγκες τις παρούσας εργασίας, κατασκευάστηκε ένα χωρικό δίκτυο, το οποίο αναπαριστά ένα οδικό δίκτυο και γραφικά φαίνεται στο σχήμα 4.5. Αυτό το δίκτυο είναι ένα χωρικό δίκτυο το οποίο κατασκευάστηκε στην oracle με την χρήση της γλώσσας SQL και ονομάστηκε **'MYNET'**. Το δίκτυο αυτό δεν απεικονίζει μια πραγματική περιοχή, αλλά σχεδιάστηκε τυχαία. Σε αυτή την εργασία το δίκτυο αυτό, θα χρησιμοποιηθεί για να δοθούν ορισμένα παραδείγματα. Για την διεξαγωγή των πειραμάτων, θα χρησιμοποιηθούν έτοιμα δίκτυα που παρέχονται στο διαδίκτυο, τα οποία αναπαριστούν μια πραγματική περιοχή. Αυτό το κάνουμε για να παράγουμε όσο το δυνατόν πιο αξιόπιστα αποτελέσματα. Πιο συγκεκριμένα το δίκτυο που θα χρησιμοποιήσουμε ονομάζεται: **'oldenburg'** και εικονίζεται στο σχήμα 4.6. Ο σκοπός που παρουσιάσαμε αυτή την διαδικασία κατασκευής δικτύων με την χρήση της γλώσσας SQL είναι, για να δείξουμε ότι υπάρχει η δυνατότητα, να χαρτογραφηθεί οποιαδήποτε περιοχή. Για παράδειγμα θα μπορούσε κάποιος να χαρτογραφήσει την πόλη του Βόλου, με την χρήση της Oracle με τον τρόπο που περιγράψαμε πιο πάνω. Κάτι τέτοιο ξεφεύγει από τα όρια τις παρούσας εργασίας, και για αυτό τον λόγο χρησιμοποιούμε κάποιο έτοιμο δίκτυο.

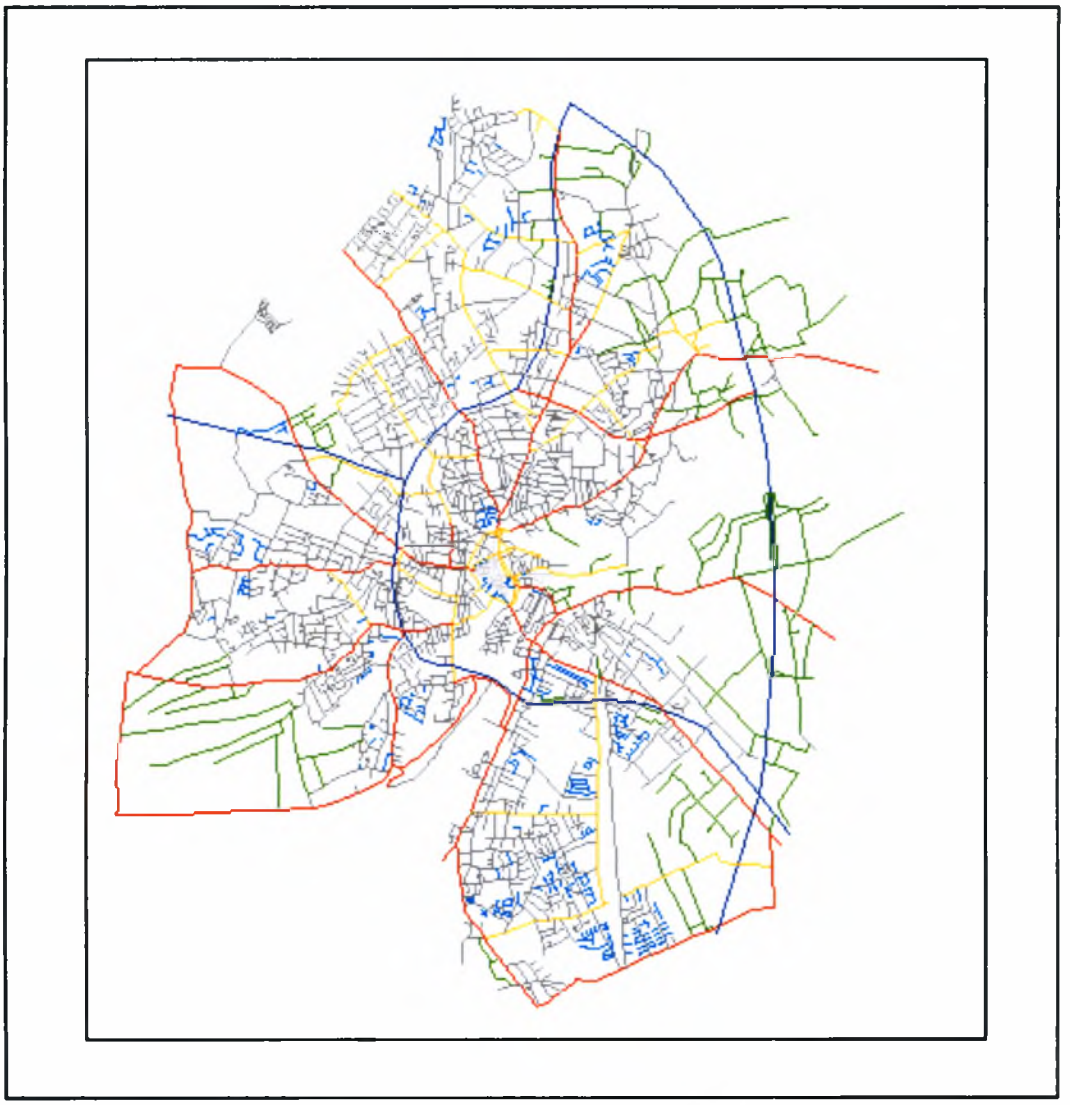
Το δίκτυο 'MYNET' είναι ένα χωρικό δίκτυο, το οποίο είναι αποθηκευμένο στην Oracle με την μορφή κατάλληλων πινάκων. Το δίκτυο αυτό λόγο της μορφής του, δεν μπορεί να φορτωθεί από το πρόγραμμα του Thomas Brinkhoff. Προκειμένου να φορτωθεί από το πρόγραμμα του Thomas Brinkhoff, χρειάστηκε να μετατραπεί σε μορφή συμβατή με τα δίκτυα που δέχεται το πρόγραμμα ως είσοδο. Πιο συγκεκριμένα το πρόγραμμα του Brinkhoff, δέχεται δίκτυα στην είσοδο του που είναι στην μορφή συμπιεσμένων αρχείων. Έτσι το πρόγραμμα για να παράγει χώρο-χρονικά αντικείμενα θα πρέπει να του δοθεί στην είσοδο ένα αρχείο 'nodes' για τους κόμβους του δικτύου, και ένα αρχείο 'edges' για τις ακμές του δικτύου. Αυτό που υλοποιήθηκε προκειμένου το χωρικό δίκτυο 'MYNET' να έρθει στην μορφή που προαναφέραμε είναι τρία java scripts τα οποία χρησιμοποιούν το πρωτόκολλο JDBC για την επικοινωνία του run time

environment της java με το ΣΔΒΔ της Oracle. Συνοπτικά αναφέρουμε ότι το πρώτο script διαβάζει το δίκτυο 'MYNET' από την Oracle με την χρήση JDBC και αποθηκεύει τις συντεταγμένες των κόμβων του δικτύου σε ένα αρχείο.txt. Στη συνέχεια το δεύτερο script διαβάζει τις συντεταγμένες από το txt αρχείο, τις μετατρέπει σε κατάλληλη μορφή και αφού προστεθούν και κάποιες άλλες πληροφορίες δημιουργείται το αρχείο 'nodes' το οποίο είναι συμβατό με το πρόγραμμα του Brinkhoff. Παρόμοια λειτουργία έχει και τρίτο script το οποίο δημιουργεί το αρχείο 'edges'.



Σχήμα 4.5: Δίκτυο 'MYNET'.

Στο πιο πάνω σχήμα μερικές ακμές επιλέξαμε να είναι μονής και άλλες διπλής κατεύθυνσης.



Σχήμα 4.6: Δίκτυο oldenburg.

ΚΕΦΑΛΑΙΟ 5

5. Διατύπωση του προβλήματος της ιδιωτικότητας χώρο-χρονικών δεδομένων και σχετικοί ορισμοί

Προκειμένου να διασφαλισθεί η ιδιωτικότητα των ευαίσθητων δεδομένων των χρηστών, οι οποίοι κάνουν αιτήσεις σε έναν έμπιστο εξυπηρετητή, πρέπει να υλοποιηθούν συγκεκριμένες τεχνικές. Αυτές τις τεχνικές, τις αναλύουμε στο επόμενο κεφάλαιο. Αυτό το κεφάλαιο πραγματεύεται τις βασικές έννοιες, η κατανόηση των οποίων είναι απαραίτητη προκειμένου να κατανοήσουμε πλήρως το πρόβλημα.

5.1 Προσδιοριστές Προστασίας Βασιζόμενοι στη Τοποθεσία (LBQIDs)

Στις βάσεις δεδομένων μερικές φορές αποθηκεύονται στοιχεία τα οποία προσδιορίζουν κάποια άτομα. Αυτά τα στοιχεία μπορεί να περιέχουν ευαίσθητα δεδομένα όπως: το όνομα, το επώνυμο, την διεύθυνση κατοικίας ενός ατόμου κτλ. Εκτός από αυτά τα ευαίσθητα δεδομένα, στη βάση είναι αποθηκεμένα και άλλα στοιχεία που δεν είναι ευαίσθητα και στα οποία επιτρέπεται η πρόσβαση. Αυτά τα δεδομένα από μόνα τους δε μπορούν να προσδιορίσουν την ταυτότητα ενός ατόμου, αν όμως συνδυαστούν με κάποια άλλα στοιχεία (π.χ με κάποια δημόσια έγγραφα, ή με στοιχεία από άλλες βάσεις δεδομένων), ίσως είναι δυνατή η αποκάλυψη της ταυτότητας των χρηστών. Η αποκάλυψη της ταυτότητας των χρηστών, σημαίνει ότι δεν διασφαλίζεται ικανοποιητικά η **ιδιωτικότητά** τους. Για την διασφάλιση της ιδιωτικότητας των χρηστών, θα πρέπει να μην υπάρχει η δυνατότητα αποκάλυψης της ταυτότητάς τους, με συνδυασμό γνωρισμάτων από την βάση δεδομένων. Για αυτό τον λόγο έχουν προταθεί οι προσδιοριστές προστασίας, ο ορισμός των οποίων ακολουθεί στην συνέχεια.

Ορισμός 5.1: Το σύνολο των γνωρισμάτων μίας βάσης δεδομένων, τα οποία σε συνδυασμό με πληροφορίες από εξωγενείς πηγές μπορούν να προσδιορίσουν την ταυτότητα συγκεκριμένων ατόμων, ονομάζεται ***προσδιοριστής προστασίας (QI)***.

Ένας προσδιοριστής προστασίας είναι ένα σύνολο γνωρισμάτων. Άρα μπορεί να αποτελείται από ένα και μόνο γνώρισμα(π.χ αριθμός ταυτότητας) το οποίο είναι ικανό να προσδιορίσει ένα άτομο μονοσήμαντα, ή να αποτελείται από περισσότερα γνωρίσματα τα οποία αν συνδυαστούν μεταξύ τους τότε προσδιορίζουν την ταυτότητα ενός ατόμου. Για παράδειγμα αν συνδυαστούν τα γνωρίσματα: τόπος γέννησης, ημερομηνία γέννησης, τόπος διαμονής και φύλο ίσως είναι δυνατόν να προσδιορισθεί η ταυτότητα ενός ατόμου.

Εκτός από τους παραδοσιακούς προσδιοριστές προστασίας, υπάρχουν και οι προσδιοριστές προστασίας που βασίζονται στην θέση *(LBQIDs)*. Αυτού του είδους οι προσδιοριστές προστασίας δεν αποτελούνται από συνδυασμό γνωρισμάτων της βάσης μας, αλλά αποτελούνται από τις διαδρομές που κάνουν συνήθως οι αιτούντες, δηλαδή από τις τροχιές που είναι πολύ συχνά-επαναλαμβανόμενες για κάθε χρήστη που κάνει μια αίτηση προς έναν παροχέα υπηρεσιών. Όπως οι παραδοσιακοί προσδιοριστές προστασίας έτσι και τα LBQIDs έχουν σαν στόχο να διασφαλίσουν την ιδιωτικότητα των προσωπικών δεδομένων των χρηστών. Η διαφορά είναι ότι τα LBQIDs χρησιμοποιούν αντί για ένα σύνολο γνωρισμάτων της βάσης, τις συχνές διαδρομές που κάνει ο χρήστης. Στην συνέχεια παρατίθεται το παράδειγμα 5.1 για να γίνει πλήρως κατανοητή η έννοια ενός LBQID.

Παράδειγμα 5.1

Έστω ένα άτομο που κάνει την διαδρομή από το σπίτι του στο νοσοκομείο. Αν αυτή η διαδρομή γίνεται καθημερινά τότε αυτή η διαδρομή, είναι ένα **LBQID.** Αντίθετα αν αυτή η διαδρομή γίνεται μια φορά τον χρόνο τότε δεν μπορεί η διαδρομή να θεωρηθεί **LBQID.** Βλέπουμε λοιπόν ότι για να αποφανθούμε αν μια διαδρομή είναι LBQID ή όχι, πρέπει να καθορίσουμε την **συχνότητα** με την οποία ένας χρήστης κάνει την συγκεκριμένη διαδρομή. Αυτό είναι απαραίτητο για να μπορούμε να ξέρουμε αν η συγκεκριμένη διαδρομή έγινε τυχαία μια φορά, ή αν γίνεται κατ' επανάληψη και άρα μπορεί να αποκαλύψει ευαίσθητα προσωπικά δεδομένα που αφορούν τον χρήστη. Στο παράδειγμά μας με την διαδρομή που κάνει ένας χρήστης από το σπίτι προς το νοσοκομείο, αν αυτή η διαδρομή γίνεται πολύ συχνά και άρα είναι LBQID, πρέπει να μην αποκαλύπτεται. Αν αυτή η πληροφορία κοινοποιείται, τότε μπορεί να βγει το συμπέρασμα ότι το δεδομένο άτομο έχει κάποιο σημαντικό πρόβλημα υγείας, και άρα κινδυνεύει η ιδιωτικότητα του. Αντίθετα αν αυτή η διαδρομή γίνεται μία φορά τον χρόνο, δεν θεωρείται LBQID και αυτή η πληροφορία δεν μπορεί να αποκαλύψει ευαίσθητα προσωπικά δεδομένα των χρηστών.

Με βάση όσα έχουμε αναφέρει για τα LBQIDs, μπορούμε να κατανοήσουμε ότι ένας χρήστης είναι δυνατόν να έχει πάνω από ένα LBQID. Για παράδειγμα ο χρήστης που κάνει καθημερινά την διαδρομή από το σπίτι του προς το νοσοκομείο, μπορεί να κάνει αρκετά συχνά(π.χ 3 φορές την εβδομάδα) και την διαδρομή από το σπίτι του προς το καζίνο. Και αυτή η διαδρομή θεωρείται LBQID διότι η αποκάλυψη της πληροφορίας αυτής(μέσω των αιτήσεων που κάνει ο χρήστης) δηλώνει ότι ο συγκεκριμένος χρήστης είτε εργάζεται στο καζίνο περιστασιακά, είτε ότι είναι εθισμένος στον τζόγο. Και στις δυο αυτές περιπτώσεις θίγεται η ιδιωτικότητα του συγκεκριμένου χρήστη.
Γενικότερα ο κάθε χρήστης μπορεί να κάνει απεριόριστες διαδρομές και ίσως να έχει απεριόριστο αριθμό από LBQID. Αυτές οι διαδρομές αποθηκεύονται σε

έναν έμπιστο εξυπηρετητή, ο οποίος με αυτό τον τρόπο κράτα ένα ιστορικό για τις διαδρομές που κάνει ο κάθε χρήστης, επίσης ανάλογα με το πόσο συχνά γίνεται η κάθε διαδρομή ο εξυπηρετητής καθορίζει αν μια διαδρομή αποτελεί η όχι LBQID. Συνεχίζοντας δίνουμε έναν τυπικό ορισμό του LBQID.

Ορισμός 5.2*: Ένα LBQID** είναι ένα μοτίβο κίνησης κάποιου χρήστη, το οποίο περιέχει μια ακολουθία από ***χώρο-χρονικά στοιχεία και ένα ***τύπο επανάληψης*** *(recurrence formula).* Τα χώρο-χρονικά στοιχεία αναπαριστούν κάποιες τοποθεσίες που ο χρήστης επισκέπτεται συχνά και έχουν την εξής μορφή: ***<θέση, Χρονόσημο>.*** Ο *τύπος επανάληψης* ενός LBQID, είναι εκείνο το στοιχείο το οποίο καθορίζει αν μια ακολουθία από χώρο-χρονικά στοιχεία αποτελεί LBQID.

Τα χώρο-χρονικά στοιχεία όπως αναφέρεται στον πιο πάνω ορισμό, είναι της μορφής: *<θέση, Χρονόσημο>*. Πιο συγκεκριμένα η θέση αναπαριστά ένα σημείο της περιοχής που έχουμε ορίσει ότι ο χρήστης μπορεί να κινείται, για παράδειγμα αν έχουμε ορίσει ότι ο χρήστης μπορεί να κινείται στα όρια μιας πόλης, η θέση μπορεί να είναι το νοσοκομείο. Αυτή η θέση (δηλαδή το νοσοκομείο) αποτελεί ένα χωρικό δεδομένο και αναπαρίσταται με ένα ζεύγος συντεταγμένων **(x,y).** Ωστόσο η θέση σε ένα LBQID παριστάνεται με δύο ζεύγη συντεταγμένων **[(x1,y1) και (x2,y2)]**. Αυτά τα δύο ζεύγη παριστάνουν ένα ορθογώνιο, το οποίο είναι μια ευρύτερη περιοχή που περιέχει την ακριβή θέση από όπου βρέθηκε ο χρήστης. Ειδικότερα το ζεύγος **(x1,y1)** αντιστοιχεί στις συντεταγμένες της κάτω αριστερής κορυφής της ορθογώνιας περιοχής και το ζεύγος **(x2,y2)** αντιστοιχεί στις συντεταγμένες της πάνω δεξιάς κορυφής της ορθογώνια περιοχής. Όσον αφορά το χρονόσημο, αυτό δηλώνει το χρονικό διάστημα μέσα στο οποίο μπορεί

ένας χρήστης να βρεθεί στη συγκεκριμένη περιοχή. Το χρονόσημο έχει την μορφή: **[t1,t2],** όπου τα **t1, t2** αναπαριστούν χρονικές στιγμές. Τα **t1, t2** μπορούμε να τα ορίσουμε σε ώρες, λεπτά, ακόμα και σε δευτερόλεπτα. Ένα χρονόσημο για παράδειγμα είναι το : [09:00 , 11:00], όπου τα 09:00 και 11:00 είναι οι ώρες μιας ημέρας. Συνολικά τώρα θα δώσουμε ένα παράδειγμα ενός *χώρο- χρονικού στοιχείου* και θα εξηγήσουμε πως αυτό ερμηνεύεται. Έστω λοιπόν το χώρο-χρονικό στοιχείο:

<(x1,y1) , (x2,y2) , [09:00,11:00]>. Αν οι συντεταγμένες (x1,y1) και (x2,y2) αντιστοιχούν σε μια ορθογώνια περιοχή, η οποία περιέχει την ακριβή θέση((x, y)) όπου βρίσκεται το νοσοκομείο, τότε αυτό το χώρο-χρονικό στοιχείο δηλώνει ότι ο χρήστης μπορεί να βρεθεί στο νοσοκομείο οποιαδήποτε χρονική στιγμή, που περιέχεται στο χρονικό διάστημα: [09:00 , 11:00].

Ο *τύπος επανάληψης* ενός LBQID, είναι εκείνο το στοιχείο το οποίο καθορίζει αν μια ακολουθία από χώρο-χρονικά στοιχεία (δηλαδή μια διαδρομή του χρήστη) αποτελεί LBQID. Πιο συγκεκριμένα ο τύπος επανάληψης καθορίζει πόσες φορές πρέπει να γίνει μια συγκεκριμένη διαδρομή έτσι ώστε να θεωρείται LBQID.
Ο τύπος επανάληψης ενός LBQID σχετίζεται με όλα τα χώρο-χρονικά στοιχεία αυτού του LBQID. Διαισθητικά, ο τύπος αυτός αποτελεί ένα είδος χρονικού περιορισμού. Αν ένας χρήστης εκτελεί μία συγκεκριμένη διαδρομή τόσες φορές όσες ορίζει ο τύπος επανάληψης, τότε αυτό σημαίνει ότι η συγκεκριμένη διαδρομή αποτελεί ένα LBQID. Η σύνταξη που έχει ο τύπος επανάληψης είναι η ακόλουθη:

$$r_1.G_1 * r_2.G_2 * \ldots\ldots * r_{i-1}.G_{i-1} * r_i.G_i \quad , \quad \text{με } i=0,1,2\ldots.n$$

Το G_i είναι μία διαβάθμιση του χρόνου **(granularity)**, και συμβολίζει ένα χρονικό διάστημα το οποίο μπορεί να έχει οριστεί σε μήνες, εβδομάδες, μέρες κτλ. Το r_i είναι ένας ακέραιος, ο οποίος δηλώνει πόσες φορές εμφανίζεται μια χώρο-χρονική ακολουθία **(<θέση, Χρονόσημο>)** σε κάθε ένα διάστημα G_i.
Η ερμηνεία του πιο πάνω τύπου είναι η εξής: Μια χώρο-χρονική ακολουθία, πρέπει να εμφανίζεται τουλάχιστον r_1 φορές εντός του ελάχιστου χρονικού διαστήματος δηλαδή εντός του G_1. Ύστερα αυτή η χώρο-χρονική ακολουθία πρέπει να εμφανίζεται τουλάχιστον r_2 φορές εντός του αμέσως μεγαλύτερου χρονικού διαστήματος, δηλαδή εντός του G_2. Με αυτή την λογική ο τύπος επεκτείνεται μέχρι το n, όπου G_n είναι το μέγιστο χρονικό διάστημα και r_n είναι οι φορές όπου η ακολουθία πρέπει να εμφανίζεται εντός του G_n. Συμπληρωματικά πρέπει να αναφέρουμε ότι η τιμή του κάθε r_i πρέπει να είναι μεγαλύτερη του ένα. Επίσης αν ο τύπος επανάληψης δεν έχει προσδιορισθεί, τότε υπονοείται ότι αρκεί η χώρο-χρονική ακολουθία μας, να εμφανίζεται μία φορά οποιαδήποτε χρονική στιγμή. Στην συνέχεια θα παραθέσουμε ένα παράδειγμα ενός LBQID.

Παράδειγμα 5.2

Έστω ότι έχουμε το άτομο του παραδείγματος 5.1 όπου κάνει την διαδρομή από το σπίτι του προς το νοσοκομείο. Θεωρούμε επίσης ότι αυτό το άτομο κάνει και την αντίθετη διαδρομή(δηλαδή από το νοσοκομείο προς το σπίτι του).
Το **LBQID** μπορεί να ορισθεί ως εξής:
< <Σπίτι , [09:00 , 11:00]> , <Νοσοκομείο , [11:00 , 13:00]>, <Νοσοκομείο , [15:00 , 17:00]> , <Σπίτι [17:00 , 19:00]> ,3.Ημέρες * 2.Εβδομάδες >

Στο πιο πάνω **LBQID** τα στοιχεία :
1- <Σπίτι , [09:00 , 11:00]>
2- <Νοσοκομείο , [11:00 , 13:00]>
3- <Νοσοκομείο , [15:00 , 17:00]>
4- <Σπίτι , [17:00 , 19:00]>

αποτελούν την χώρο-χρονική ακολουθία η οποία αποτελείται από 4 χώρο-χρονικά στοιχεία. Παρατηρούμε επίσης ότι ο τύπος επανάληψης είναι της μορφής: **4.Ημέρες * 3.Εβδομάδες.** Αντιπαραβάλλοντας αυτόν τον τύπο με τον γενικό τύπο που έχουμε δώσει, προκύπτει ότι το 3 αντιστοιχεί στο r_1 και το 2 στο r_2. Επίσης η πρώτη χρονική διαβάθμιση (G_1) είναι ορισμένη σε **Ημέρες** και η δεύτερη χρονική διαβάθμιση (G_2) είναι ορισμένη σε **Εβδομάδες.**
Και σε αυτό το παράδειγμα βλέπουμε ότι η πρώτη χρονική διαβάθμιση(G_1) είναι υποσύνολο της δεύτερης(G_2). Πιο συγκεκριμένα ο παραπάνω τύπος επανάληψης δηλώνει ότι για να θεωρηθεί η διαδρομή μας (Σπίτι->Νοσοκομείο ,Νοσοκομείο->Σπίτι) σαν LBQID θα πρέπει ο χρήστης να κάνει αυτή την διαδρομή τουλάχιστον 3 ημέρες την εβδομάδα και τουλάχιστον επί 2 εβδομάδες.
Προκειμένου να δώσουμε κάποιους ορισμούς χρειάζεται να έχουμε έναν πιο συνοπτικό τύπο για τα LBQID. Έτσι ένας εναλλακτικός τύπος είναι ο εξής:

LBQID: $<E_1,E_2,....E_n>, G_1 * r_2.G_2 * .....* r_{i-1}.G_{i-1} * r_i.G_i$

Με $\mathbf{E}_i$ συμβολίζουμε τα χώρο-χρονικά στοιχεία του **LBQID**, τα οποία όπως έχουμε πει έχουν την μορφή: <θέση, Χρονόσημο>.

Τα **LBQID** του κάθε χρήστη αποθηκεύονται σε έναν έμπιστο εξυπηρετητή. Όταν λοιπόν ένας χρήστης κάνει μια αίτηση από μια συγκεκριμένη θέση, πρέπει ο έμπιστος εξυπηρετητής να μπορεί να αποφανθεί αν αυτή η αίτηση ταιριάζει με τα **LBQID** που έχει υποθηκευμένα.

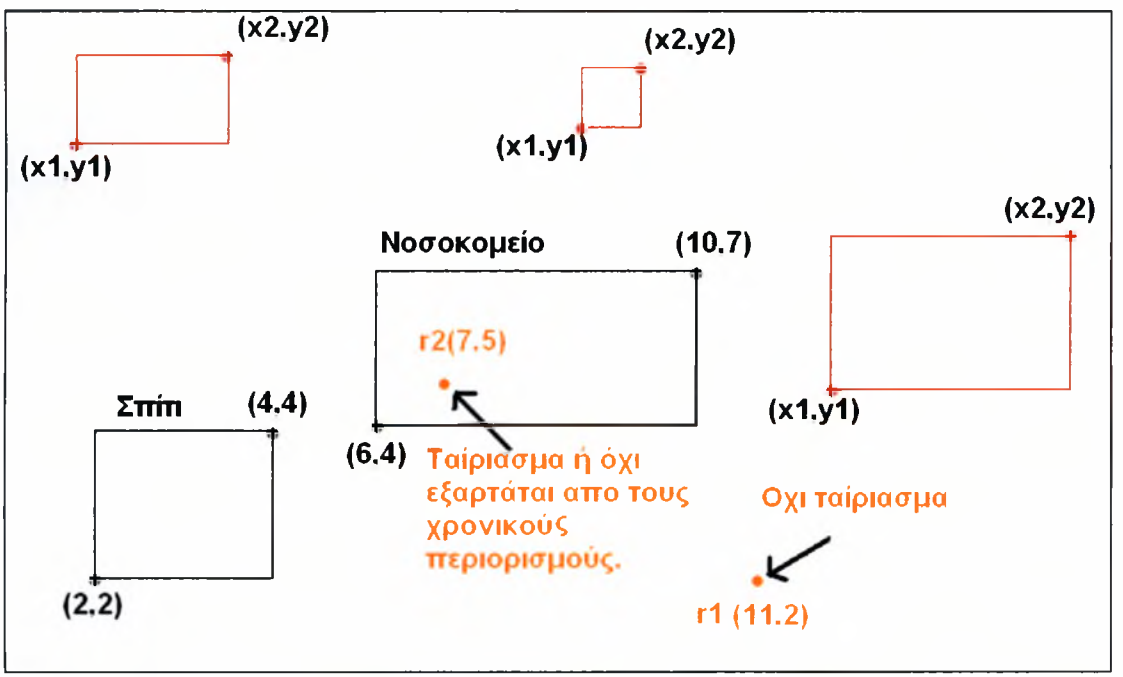
Ορισμός 4.2: Αν ένας χρήστης **(U_i)** κάνει μία αίτηση **r_i** από την θέση **<x_i, y_i>** την χρονική στιγμή **t_i**, τότε λέμε ότι η αίτηση **r_i** ταιριάζει με ένα χώρο-χρονικό στοιχείο **E_j** του LBQID, αν η θέση από όπου έγινε η αίτηση **(<xi, yi>)**, εμπεριέχεται στην ορθογώνια περιοχή που έχουμε ορίσει στο **Ej**, και επίσης η χρονική στιγμή που έγινε η αίτηση **(t_i)** περιέχεται στο διάστημα που έχουμε ορίσει στο **E_j**.

Με το ακόλουθο παράδειγμα (παράδειγμα 5.3), θα γίνει πλήρως κατανοητός ο ορισμός που μόλις διατυπώθηκε.

Παράδειγμα 5.3

Έστω ότι χρησιμοποιούμε τα δεδομένα που έχουμε από το παράδειγμα 5.2 αλλά προσδιορίζουμε πλήρως τις ορθογώνιες περιοχές για το σπίτι και για το νοσοκομείο. Έστω ότι έχουμε ορίσει ένα γεωγραφικό χώρο ο οποίος περιέχει κάποια χωρικά αντικείμενα. Οι πληροφορίες αυτές φαίνονται στο σχήμα 5.1 που ακολουθεί. Έστω λοιπόν ότι το σπίτι προσδιορίζεται από τα ζεύγη

συντεταγμένων: **((2,2) και (4,4))**, και το νοσοκομείο προσδιορίζεται από τα ζεύγη συντεταγμένων: **((6,4) και (10,7))**. Οι ορθογώνιες περιοχές που έχουν σχεδιαστεί με κόκκινο χρώμα αναπαριστούν ενδεχομένως άλλες τοποθεσίες οι οποίες σε αυτό το παράδειγμα δεν μας απασχολούν.(Οι συντεταγμένες (x1,y1) και (x2,y2) της κάθε περιοχής είναι διαφορετικές για κάθε περιοχή).



Σχήμα 5.1:Γεωγραφικός χώρος

Το LBQID είναι και πάλι:

< <Σπίτι , [09:00 , 11:00]> , <Νοσοκομείο , [11:00 , 13:00]>, <Νοσοκομείο , [15:00 , 17:00]> , <Σπίτι [17:00 , 19:00]> ,3.Ημέρες * 2.Εβδομάδες >

Όπου:

- <Σπίτι , [09:00 , 11:00]> = E1
- <Νοσοκομείο , [11:00 , 13:00]>=E2
- <Νοσοκομείο , [15:00 , 17:00]> =E3
- <Σπίτι , [17:00 , 19:00]> =E4

Με βάση τα δεδομένα που έχουμε ορίσει, θα δώσουμε υποθετικές αιτήσεις των χρηστών και θα εξηγήσουμε αν ταιριάζουν ή όχι με κάποιο $\mathbf{E}_j$ του LBQID .

Έστω ότι ένας χρήστης (με id=1) κάνει την αίτηση $\mathbf{r}_1$ από την θέση **(11,2)** την χρονική στιγμή $\mathbf{t}_1$**=09:30.** Αυτή η αίτηση αν και γίνεται την χρονική στιγμή $\mathbf{t}_1$**=09:30** η οποία περιέχεται στους χρονικούς περιορισμούς που έχουν ορισθεί για το $\mathbf{E}_1$([09:00 , 11:00]), παρόλα αυτά δεν ταιριάζει με το συγκεκριμένο LBQID διότι η θέση από όπου έγινε η αίτηση:**(11,2),** δεν περιέχεται ούτε στην ορθογώνια περιοχή που έχουμε ορίσει για το Σπίτι αλλά ούτε και σε αυτήν για το Νοσοκομείο. Συνεπώς οποιαδήποτε χρονική στιγμή και αν γίνει μια αίτηση από το συγκεκριμένο σημείο δεν θα ταιριάξει με κανένα $\mathbf{E}_j$ του LBQID.

Έστω τώρα ότι ένας χρήστης (με id=2) κάνει μια αίτηση $\mathbf{r}_2$ από την θέση **(7, 5)** την χρονική στιγμή $\mathbf{t}_2$**= 09:30**. Βλέπουμε ότι η θέση από όπου έγινε η αίτηση περιέχεται στην ορθογώνια περιοχή για το νοσοκομείο. Άρα κοιτάμε τα $\mathbf{E}_2$ **και** $\mathbf{E}_3$ για να δούμε αν η χρονική στιγμή που έγινε η αίτηση περιέχεται στους χρονικούς περιορισμούς. Βλέπουμε ότι η χρονική στιγμή $\mathbf{t}_2$ δεν περιέχεται ούτε στο [11:00 , 13:00] αλλά ούτε και στο [15:00 , 17:00] και άρα δεν ταιριάζει με κανένα $\mathbf{E}_j$. Από την άλλη πλευρά βλέπουμε ότι η χρονική στιγμή που έγινε η αίτηση($\mathbf{t}_2$**= 09:30**) περιέχεται στο [09:00 , 11:00] του $\mathbf{E}_1$ αλλά αυτό δεν έχει καμία σημασία, διότι το $\mathbf{E}_1$ αναφέρεται στην τοποθεσία Σπίτι.

Τέλος μερικές αιτήσεις που ταιριάζουν με το LBQID που έχουμε ορίσει είναι οι εξής:

- Από την θέση (3, 3) την χρονική στιγμή t=09:30 ταιριάζει με το E_1
- Από την θέση (3, 4) την χρονική στιγμή t=10:30 ταιριάζει με το E_1
- Από την θέση (4, 3) την χρονική στιγμή t=17:30 ταιριάζει με το E_4
- Από την θέση (7, 5) την χρονική στιγμή t=11:30 ταιριάζει με το E_2
- Από την θέση (9, 6) την χρονική στιγμή t=12:30 ταιριάζει με το E_2
- Από την θέση (8, 7) την χρονική στιγμή t=15:30 ταιριάζει με το E_3

Στον έμπιστο εξυπηρετητή, συχνά είναι ανάγκη να ελεγχθεί αν ένα σύνολο από αιτήσεις (και όχι μια μεμονωμένη αίτηση) ταιριάζουν με ένα LBQID. Η λογική είναι ακριβώς η ίδια, απλά κάνουμε όλους τους ελέγχους που αναφέραμε πιο πάνω για όλες τις αιτήσεις που συμπεριλαμβάνονται στο σύνολο των αιτήσεων.

Ορισμός 5.3: Ένα σύνολο αιτήσεων **R** λέμε ότι ταιριάζει με ένα χώρο-χρονικό στοιχείο $\mathbf{E}_j$ του **LBQID**, αν ισχύουν οι πιο κάτω συνθήκες :

1) Κάθε αίτηση $\mathbf{r}_i$ του συνόλου **R** ταιριάζει με κάποιο $\mathbf{E}_j$.

2) Οι χρονικές στιγμές που έγιναν όλες οι αιτήσεις του συνόλου **R**(δηλαδή κάθε αίτηση $\mathbf{r}_i$ έγινε την χρονική στιγμή $\mathbf{t}_i$), ικανοποιούν όλες τον τύπο επανάληψης.

5.2 Προσωπικό Ιστορικό Τοποθεσιών (PHL).

Το προσωπικό ιστορικό τοποθεσιών κάθε χρήστη, είναι μια ακολουθία από χώρο-χρονικά στοιχεία. Κάθε χρήστης λοιπόν έχει ένα δικό του PHL το οποίο δηλώνει ποια σημεία επισκέφτηκε ο χρήστης και ποια χρονική στιγμή επισκέφτηκε το κάθε σημείο. Το PHL κάθε χρήστη αποθηκεύεται στον έμπιστο εξυπηρετητή. Το PHL ενός χρήστη, δείχνει από ποια σημεία έχει περάσει ο χρήστης, αφού λοιπόν οι χρήστες κινούνται συνεχώς τα PHLs τους αλλάζουν. Προκύπτει έτσι η ανάγκη ο έμπιστος εξυπηρετητής να παρακολουθεί τις κινήσεις που κάνει ο κάθε χρήστης και να ενημερώνει κατάλληλα τα PHLs τους. Πιο κάτω δίνουμε τον πιο ακριβή ορισμό των PHLs.

Ορισμός 5.4: Το PHL ενός χρήστη είναι μια ακολουθία από χώρο-χρονικά στοιχεία, τα οποία έχουν την εξής μορφή: **(x,y,t)**. Τα **x, y** παριστάνουν τις ακριβείς συντεταγμένες του σημείου από όπου πέρασε ο χρήστης, ενώ το **t** είναι η ακριβής χρονική στιγμή όπου ο χρήστης βρέθηκε στο σημείο με συντεταγμένες **(x, y).**

Στον παραπάνω ορισμό, η χρονική στιγμή **t** θα είναι της μορφής: **<Ημερομηνία, Ώρα: Λεπτά: Δευτερόλεπτα>**, δηλαδή θα είναι μία λεπτομερής περιγραφή της χρονικής στιγμής όπου έγινε η καταγραφή. Επίσης πρέπει να τονίσουμε ότι ανεξάρτητα από το αν ένας χρήστης έκανε μία αίτηση από ένα δεδομένο σημείο (έστω (x,y)), την χρονική στιγμή όπου βρέθηκε εκεί(έστω t), η τριάδα (x,y,t) πρέπει να αποθηκευτεί στο PHL του συγκεκριμένου χρήστη.

Ορισμός 5.5: Έστω ότι έχουμε ένα σύνολο αιτήσεων **R=(r_1,r_2,r_3,....r_n)** και ένα PHL ενός χρήστη. Το PHL του χρήστη λέμε ότι είναι χώρο-χρονικά συνεπές με το σύνολο των αιτήσεων **R** αν ισχύει **μία** από τις ποιο κάτω περιπτώσεις :

1)Για κάθε αίτηση r_j **(x_j , y_j , t_j)** του συνόλου **R**, υπάρχει ένα στοιχείο **(x_i , y_i , t_i)** στο **PHL** του χρήστη τέτοιο ώστε: $x_j=x_i$ & $y_j=y_i$ & $t_j=t_i$.

Ή

2) Για κάθε αίτηση r_j **(x_j , y_j , t_j)** του συνόλου **R**, υπάρχουν δύο στοιχεία **(x_i , y_i , t_i)** και **(x_k , y_k , t_k)** στο **PHL** του χρήστη τέτοια ώστε:

$x_i \leq x_j \leq x_k$ & $y_i \leq y_j \leq y_k$ & $t_i \leq t_j \leq t_k$

5.3 κ-ανωνυμία για υπηρεσίες που βασίζονται στην θέση.

Πιο πάνω αναφερθήκαμε στον όρο κ-ανωνυμία χωρίς να έχουμε ορίσει με ακρίβεια αυτό τον όρο. Σε αυτή την ενότητα θα ορίσουμε τι είναι **κ-ανωνυμία** και θα παρουσιάσουμε μια τεχνική με την οποία μπορούμε να διασφαλίσουμε κ-ανωνυμία.

Η κ-ανωνυμία στα πλαίσια αυτής της εργασίας, σχετίζεται με υπηρεσίες που βασίζονται στην θέση. Αυτές οι υπηρεσίες παρέχονται από έναν παροχέα υπηρεσιών, ο οποίος πρέπει να έχει πληροφορίες σχετικά με την θέση και την χρονική στιγμή από όπου έγινε η αίτηση. Αυτές οι πληροφορίες είναι απαραίτητο να σταλούν στον παροχέα υπηρεσιών, αφότου **τροποποιηθούν κατάλληλα** , προκειμένου να μην μπορεί να προσδιοριστεί η ταυτότητα του χρήστη. Πιο συγκεκριμένα όταν ένας χρήστης στέλνει μια αίτηση, για να επιτευχθεί κ-ανωνυμία, θα πρέπει να υπάρχουν **τουλάχιστον κ-1 χρήστες** στην περιοχή από όπου ο αιτών έκανε την αίτηση. Έτσι αφού υπάρχουν κ-1 χρήστες, στη περιοχή από όπου έχει κάνει ο αιτών την αίτηση, δεν είναι δυνατόν να προσδιοριστεί ποιος από τους κ χρήστες, είναι αυτός που έκανε την συγκεκριμένη αίτηση. Όπως είναι προφανές όσο αυξάνεται η τιμή του κ τόσο πιο δύσκολο είναι να προσδιοριστεί ένας χρήστης. Για παράδειγμα αν κ=1 σημαίνει ότι στην περιοχή από όπου ο αιτών έκανε την αίτηση, περιέχεται μόνο ένας χρήστης. Σε αυτή την περίπτωση αν επιλεγεί τυχαία ένας χρήστης υπάρχει 50% πιθανότητα να είναι ο χρήστης όπου έκανε την αίτηση. Αντίθετα αν κ=100 και επιλεγεί τυχαία ένας από τους 100 χρήστες, η πιθανότητα να βρεθεί ο χρήστης που έκανε την συγκεκριμένη αίτηση είναι πολύ πιο μικρή(1%) και άρα η ανωνυμία του χρήστη διασφαλίζεται σε πολύ μεγαλύτερο βαθμό.

Για την διασφάλιση κ-ανωνυμίας προτείνουμε την ακόλουθη τεχνική. Έστω ότι ένας χρήστης κάνει μια αίτηση προς το παροχέα υπηρεσιών από την θέση: **(x,y)** και την χρονική στιγμή **t**. Με βάση την τεχνική, στον παροχέα υπηρεσιών, δεν στέλνονται οι ακριβείς χώρο-χρονικές συντεταγμένες της αίτησης, αλλά μια γενίκευση τους. Πιο συγκεκριμένα για τις συντεταγμένες της θέσης(x,y) παράγεται μία περιοχή-**Area** η οποία περιέχει την ακριβή θέση από όπου έγινε η αίτηση((x,y)). Αντίστοιχα η χρονική στιγμή που έγινε η αίτηση(t) αντικαθίσταται από ένα χρονικό διάστημα-**Time Interval** το οποίο περιέχει την χρονική στιγμή t.

Τέλος για να σταλεί στον παροχέα υπηρεσιών η αίτηση, θα πρέπει στην περιοχή-**Area** να έχουν βρεθεί τουλάχιστον κ-1 χρήστες εντός του χρονικού διαστήματος-**Time Interval.** Στην συνέχεια στέλνεται η αίτηση στον παροχέα υπηρεσιών, ο οποίος αν και ξέρει ποιοι χρήστες βρίσκονταν στην περιοχή Area εντός του χρονικού διαστήματος Time Interval, παρόλα αυτά δεν μπορεί να ξεχωρίσει ποιος από τους κ χρήστες έκανε την αίτηση. Πιο κάτω δίνουμε έναν πιο τυπικό ορισμό για το πότε ένας χρήστης εξασφαλίζει κ-ανωνυμία.

Ορισμός 5.6 :Αν σε έναν παροχέα υπηρεσιών, έχει σταλεί ένα σύνολο αιτήσεων **R** και από αυτό το σύνολο, το υποσύνολο $\mathbf{R}_i$ έχει σταλεί από τον χρήστη($\mathbf{U}_i$), λέμε ότι διασφαλίζεται κ-ανωνυμία για τον χρήστη($\mathbf{U}_i$), αν υπάρχουν κ-1 PHLs **από διαφορετικούς χρήστες**, τέτοια ώστε κάθε PHL να είναι χώρο-χρονικά συνεπές με το σύνολο αιτήσεων $\mathbf{R}_i$.

5.4 Διασύνδεση μεταξύ χρηστών- αιτήσεων.

Λέγοντας διασύνδεση μεταξύ χρήστη και αιτήσεων, εννοούμε την συσχέτιση που μπορεί να υπάρξει ανάμεσα σε έναν χρήστη και στις αιτήσεις που αυτός ενδέχεται να κάνει στο μέλλον. Αυτή η σύνδεση ενός χρήστη με τις μελλοντικές του αιτήσεις αποτελεί πρόβλημα, διότι αν είναι γνωστό ποιες αιτήσεις έχει κάνει ένας χρήστης, ίσως είναι δυνατόν να αποκαλυφθεί η ταυτότητα του. Για να μην υπάρχει αυτό το πρόβλημα διασύνδεσης **(χρήστη-μελλοντικών του αιτήσεων)**, έχουν προταθεί πολλές τεχνικές. Μια από αυτές τις τεχνικές είναι ο κάθε χρήστης να αλλάζει αναγνωριστικό ανά τακτά χρονικά διαστήματα. Αυτή η τεχνική έχει δύο μειονεκτήματα. Το πρώτο μειονέκτημα, είναι ότι υπάρχει περίπτωση να υπάρξει σύνδεση δύο ή και περισσοτέρων αιτήσεων που έκανε ένας χρήστης, αν το χρονικό διάστημα που μεσολαβεί μεταξύ της αλλαγής αναγνωριστικών είναι σχετικά μικρό. Το δεύτερο πρόβλημα που έχει αυτή η λύση είναι ότι κοστίζει πολύ. Πρέπει να πούμε ότι όταν ένας χρήστης αλλάζει id, πρέπει να αντικαταστήσουμε από την βάση δεδομένων με τα PHLs και τα LBQIDs, το

παλιό id του χρήστη με το καινούριο, και αυτή η διαδικασία απαιτεί μεγάλη υπολογιστική ισχύ.

Σε αυτή την εργασία θα υιοθετήσουμε μια τεχνική που ελέγχει κατά τακτά χρονικά διαστήματα την θέση του κάθε χρήστη και στη συνέχεια χρησιμοποιεί διάφορες πιθανοτικές συναρτήσεις και μοντέλα, για να προσδιορίσει την πιθανότητα διασύνδεσης μεταξύ μελλοντικών αιτήσεων του ίδιου χρήστη. Αν αυτή η πιθανότητα διασύνδεσης που επιστρέφει η τεχνική είναι μεγάλη, τότε προκειμένου να διασφαλιστεί η ιδιωτικότητα του χρήστη πρέπει να ληφθούν κατάλληλα μέτρα.

Ειδικότερα στο μοντέλο μας θεωρούμε ότι ο έμπιστος εξυπηρετητής σχήμα 6.1 διαθέτει ένα σύνολο από συναρτήσεις τις οποίες συνολικά τις ονομάζουμε συναρτήσεις **LINK()**. Αυτές οι LINK συναρτήσεις επιστρέφουν την πιθανότητα να υπάρξει διασύνδεση μεταξύ ενός χρήστη και των μελλοντικών του αιτήσεων. Αν υποθέσουμε ότι έχουμε δύο αιτήσεις **(r_i και r_j)** οι οποίες μπορούν να είναι παράμετροι της συνάρτησης LINK, τότε για την συνάρτηση LINK ισχύουν οι παρακάτω ιδιότητες:

1)LINK(r_i, r_j) = LINK(r_i, r_j). Δηλαδή η συνάρτηση LINK είναι **συμμετρική.**

2)LINK(r_i, r_i)=1 & **LINK(r_j, r_i)=1.** Δηλαδή η συνάρτηση LINK είναι **ανασταλτική.**

Όπως προαναφέραμε, η τιμή που επιστρέφει η συνάρτηση LINK αντιπροσωπεύει πιθανότητα, άρα η τιμή αυτή θα είναι μια τιμή που θα ανήκει στο διάστημα [0,1]. Αν λοιπόν η συνάρτηση LINK επιστρέψει την τιμή '1' (LINK(r_i, r_j)=LINK(r_j, r_i)=1), τότε είναι σίγουρο ότι οι αιτήσεις r_j, r_i έχουν σταλεί από τον ίδιο χρήστη, συνεπώς πρέπει να ληφθούν κατάλληλα μέτρα για την διαφύλαξη της ταυτότητας του. Αντίθετα αν η συνάρτηση LINK επιστρέψει την τιμή '0' (LINK(r_i, r_j)=LINK(r_j, r_i)=0), τότε είναι σίγουρο ότι οι αιτήσεις είναι από διαφορετικούς χρήστες, άρα δεν χρειάζονται επιπλέον μέτρα να ληφθούν.

Ορισμός 5.7: Αν **R=(r_1,r_2,r_3,....., r_n)** είναι το σύνολο των αιτήσεων που έχουν σταλεί στον έμπιστο εξυπηρετητή, τότε η τιμή που επιστρέφει η συνάρτηση **LINK(R)** αντιπροσωπεύει την πιθανότητα δύο διαφορετικές αιτήσεις r_i, r_j(του συνόλου αιτήσεων **R**) να έχουν σταλεί από τον ίδιο χρήστη.

Ορισμός 5.8: Αν **R** είναι το σύνολο αιτήσεων που έχουν σταλεί στον έμπιστο εξυπηρετητή και **R'** είναι ένα υποσύνολο του **R**. Τότε λέμε ότι το σύνολο **R'** διασυνδέεται με πιθανότητα **Θ**, αν για κάθε ζεύγος αιτήσεων**(r_i, r_j) € R'** , υπάρχει ένα σύνολο αιτήσεων r_{i1}, r_{i2}, r_{i3},.....,r_{ik} **€ R'** όπου r_{i1}=r_i και r_{ik}=r_j, τέτοια ώστε **LINK(r_{il} , r_{il+1})≥ Θ** για όλα τα **l**=1,2,3....k-1.

Με βάση τον προηγούμενο ορισμό, θα λέμε ότι όλες οι αιτήσεις ενός συνόλου **R'** € R (όπου R όλες οι αιτήσεις που έχουν σταλεί σε κάποιο παροχέα υπηρεσιών), έχουν σταλεί από τον ίδιο χρήστη, αν και μόνο αν το R' *διασυνδέεται* με πιθανότητα **Θ=1**.

ΚΕΦΑΛΑΙΟ 6

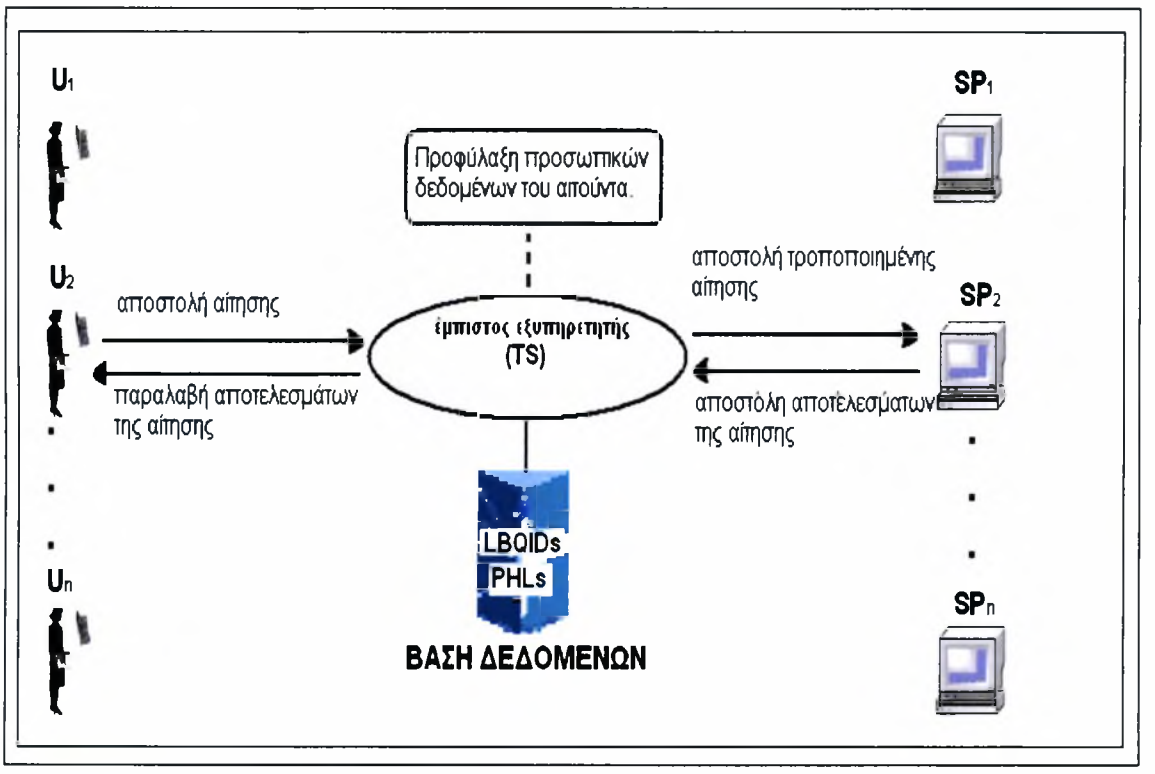
6. Το μοντέλο της κ-Ανωνυμίας.

Σε αυτό το κεφάλαιο θα παρουσιάσουμε την λύση στο πρόβλημα που έχουμε ορίσει. Όπως έχουμε πει όταν ένας χρήστης κάνει μια αίτηση προς έναν παροχέα υπηρεσιών, θα πρέπει να στέλνει και πληροφορίες που αφορούν την θέση και την χρονική στιγμή όπου έκανε την αίτηση. Αυτές οι πληροφορίες είπαμε ότι πρέπει να τροποποιούνται κατάλληλα έτσι ώστε οι χρήστες να διατηρούν την ανωνυμία τους.

Η λύση που προτείνουμε βασίζεται σε δύο αλγορίθμους, τον αλγόριθμο της **γενίκευσης** και τον αλγόριθμο **αποσύνδεσης**. Σε αυτό το κεφάλαιο θα περιγράψουμε αυτούς τους δύο αλγορίθμους, αλλά θα δώσουμε και τον ψευδοκώδικα για κάθε έναν. Πριν από αυτό όμως θα παρουσιάσουμε το μοντέλο στο οποίο βασίζεται η λύση μας, δηλαδή ένα μοντέλο ανωνυμίας που βασίζεται σε υπηρεσίες που σχετίζονται με την τοποθεσία.

6.1 Μοντέλο ανωνυμίας που βασίζεται σε υπηρεσίες τοποθεσίας.

Στο σχήμα 6.1 παρουσιάζεται η δομή του μοντέλου που έχει υιοθετηθεί για τις ανάγκες αυτής της εργασίας. Σε αυτό το μοντέλο βασίζονται και τα περισσότερα συστήματα που υπάρχουν σήμερα και παρέχουν υπηρεσίες που βασίζονται στην τοποθεσία. Στο σχήμα φαίνεται η περίπτωση όπου ο **Χρήστης** (U_2) στέλνει μια αίτηση προς εξυπηρέτηση, την οποία θα επεξεργαστεί ο **παροχέας υπηρεσιών** (SP_2). Οποιοσδήποτε χρήστης μπορεί να στείλει μια αίτηση και οποιοσδήποτε παροχέας υπηρεσιών(αρκεί να μπορεί να παρέχει την δεδομένη υπηρεσία που ζητά ο αιτών) μπορεί να αναλάβει την επεξεργασία της αίτησης αυτής. Στο σχήμα απλά παρατίθεται ένα στιγμιότυπο.



Σχήμα 6.1:Μοντέλο ανωνυμίας.

Από το πιο πάνω σχήμα, βλέπουμε ότι οι χρήστες μέσω μιας φορητής συσκευής μπορούν να κάνουν αιτήσεις προς κάποιους παροχείς υπηρεσιών. Οι αιτήσεις όμως, όπως φαίνεται και στο σχήμα, δεν στέλνονται άμεσα στον **παροχέα υπηρεσιών** αλλά πρώτα στέλνονται σε έναν **έμπιστο εξυπηρετητή**.

Οι χρήστες, μέσω της φορητής συσκευής που έχουν για την πρόσβαση στις απομακρυσμένες υπηρεσίες, έχουν την δυνατότητα να ενεργοποιήσουν ή να απενεργοποιήσουν ένα σύστημα προστασίας. Επίσης μπορούν να ρυθμίσουν το επίπεδο προστασίας που επιθυμούν, με βάση κάποιες διαβαθμίσεις(π.χ low, medium, high). Με αυτό τον τρόπο προστατεύεται και η ιδιωτικότητα των χρηστών, στον βαθμό που αυτοί επιθυμούν. Για παράδειγμα αν ένας χρήστης έχει ρυθμίσει την φορητή συσκευή που διαθέτει, στο επίπεδο προστασίας High, τότε τα μέτρα που λαμβάνονται προκειμένου να διαφυλαχτεί η ιδιωτικότητά του

είναι πολύ πιο αυστηρά, απ' ότι θα ήταν αν είχε ρυθμίσει το επίπεδο προστασίας στο low.

Όσον αφορά την ιδιωτικότητα των χρηστών, πρέπει να τονίσουμε ότι στις ευαίσθητες πληροφορίες του κάθε χρήστη, συμπεριλαμβάνονται η θέση αλλά και η χρονική στιγμή από όπου έκανε την αίτηση. Αυτές οι πληροφορίες αποθηκεύονται στον έμπιστο εξυπηρετητή. Πιο συγκεκριμένα στον έμπιστο εξυπηρετητή υπάρχει μια βάση δεδομένων, η οποία έχει την δυνατότητα να αποθηκεύει χώρο-χρονικά δεδομένα. Σε αυτή την βάση αποθηκεύονται τα **PHLs** κάθε χρήστη αλλά και τα **LBQIDs** του. Σκοπός του έμπιστου εξυπηρετητή είναι να προστατεύει την ιδιωτικότητα των χρηστών. Αυτό γίνεται με το να τροποποιεί τις πληροφορίες που στέλνει κάθε χρήστης μαζί με μια αίτηση που κάνει και στην συνέχεια να προωθεί την αίτηση στον παροχέα υπηρεσιών σε μια τροποποιημένη μορφή η οποία δεν μπορεί να αποκαλύψει προσωπικά δεδομένα του κάθε χρήστη. Στο μοντέλο που έχουμε υιοθετήσει, ο χρήστης στέλνει στον έμπιστο εξυπηρετητή τις εξής πληροφορίες: **1)**Την διεύθυνση δικτύου της φορητής συσκευής, μέσω της οποίας έκανε την αίτηση, **2)**Κάποια στοιχεία που προσδιορίζουν την ταυτότητα του χρήστη, **3)**Την ακριβή θέση από όπου έκανε την αίτηση, **4)**Την ακριβή χρονική στιγμή που έκανε την αίτηση, **5)**Ένα σύνολο δεδομένων που αφορούν την αίτηση. Όλα τα πιο πάνω στοιχεία δεν στέλνονται στον παροχέα υπηρεσιών ως έχουν, αλλά τροποποιούνται σύμφωνα με τον τρόπο που περιγράφεται στην συνέχεια.

Τελικά ο παροχέας υπηρεσιών δέχεται μια αίτηση(από τον έμπιστο εξυπηρετητή) η οποία έχει την εξής μορφή: **(msgID, User Pseudonym, Area, Time Interval, Data)**. Το πεδίο **msgID** χρησιμοποιείται προκειμένου να υποκρυφθεί η διεύθυνση δικτύου της συσκευής που χρησιμοποιεί ο κάθε χρήστης. Στην συνέχεια το **msgID** θα το χρησιμοποιήσει ο έμπιστος εξυπηρετητής για να στείλει την απάντηση στην αίτηση που έκανε ο χρήστης. Το πεδίο **User Pseudonym** χρησιμοποιείται για να αποκρύψει την ταυτότητα του χρήστη. Αυτό το ψευδώνυμο που δίνεται στους χρήστες, πρέπει να είναι διαφορετικό για κάθε χρήστη, έτσι ώστε να μπορεί ο παροχέας υπηρεσιών να κάνει **αυθεντικοποίηση** του χρήστη. Με τον όρο αυθεντικοποίηση εννοούμε την διαδικασία σύμφωνα με την οποία ο κάθε χρήστης μπορεί να αναγνωριστεί μονοσήμαντα από το σύστημα και να του δοθούν τα κατάλληλα δικαιώματα.

Τα πεδία **Area** και **Time Interval,** χρησιμοποιούνται για να αποκρύψουν την ακριβή θέση και χρονική στιγμή που έγινε η αίτηση.

Η ακριβής θέση αλλά και η χρονική στιγμή που έγινε μια αίτηση δεν πρέπει να αποκαλύπτονται διότι μπορεί να κινδυνεύσει η ιδιωτικότητα των χρηστών. Έτσι η ακριβής θέση (x , y) από όπου έγινε η αίτηση, αντικαθίσταται στον έμπιστο εξυπηρετητή από μια ευρύτερη περιοχή-**Area** η οποία περιέχει την ακριβή θέση (x , y). Όμοια η ακριβής χρονική στιγμή που έγινε η αίτηση(t) αντικαθίσταται από ένα χρονικό διάστημα **Time Interval** το οποίο περιέχει την ακριβή χρονική στιγμή (t). Αυτή η διαδικασία ονομάζεται γενίκευση του χώρου και του χρόνου αντίστοιχα και επιτυγχάνεται με την εφαρμογή κατάλληλων αλγορίθμων στον έμπιστο εξυπηρετητή. Αυτοί οι αλγόριθμοι αποτελούν το κυρίαρχο θέμα στην παρούσα εργασία και θα τους αναλύσουμε στην συνέχεια διεξοδικά. Τέλος στην αίτηση που φτάνει στον παροχέα υπηρεσιών υπάρχει και ένα πεδίο **Data**, το οποίο περιέχει δεδομένα που αφορούν την αίτηση.

6.2 Συνοπτική περιγραφή των αλγορίθμων γενίκευσης και αποσύνδεσης για την επίτευξη κ-ανωνυμίας.

Σε αυτή την ενότητα θα περιγράψουμε με συντομία τους δυο αλγορίθμους που έχουμε υλοποιήσει προκειμένου να επιτευχθεί κ-ανωνυμία ενός χρήστη που κάνει μια αίτηση. Ο πρώτος αλγόριθμος ονομάζεται αλγόριθμος γενίκευσης και ο δεύτερος αλγόριθμος αποσύνδεσης. Αυτοί οι δυο αλγόριθμοι εφαρμόζονται στον έμπιστο εξυπηρετητή κάθε φορά που ένας χρήστης αποστέλλει μια αίτηση προς εξυπηρέτηση. Στόχος τους είναι να διασφαλισθεί **κ-ανωνυμία**, με άλλα λόγια οι αλγόριθμοι αυτοί πρέπει να μετασχηματίζουν κατάλληλα τις συντεταγμένες του τρισδιάστατου χώρου, που αποστέλλει ο χρήστης μαζί με την αίτηση, έτσι ώστε να μην είναι δυνατόν να αποκαλυφθούν ευαίσθητες πληροφορίες του αιτούντα.
Μια σειρά βημάτων είναι απαραίτητη, προκειμένου να αποφασισθεί αν τα χώρο-χρονικά δεδομένα που αποστέλλει ο χρήστης μαζί με την αίτηση, είναι δυνατόν να θέσουν τα προσωπικά του στοιχεία σε κίνδυνο. Αν κάτι τέτοιο διαπιστωθεί τότε πρέπει να γίνουν αλλαγές στα χώρο-χρονικά δεδομένα έτσι ώστε να διασφαλισθεί η ιδιωτικότητα του αιτούντα.
Αρχικά εκτελείται ο **αλγόριθμος γενίκευσης** ο οποίος ελέγχει αν η αίτηση που έγινε από κάποιον χρήστη, ταιριάζει με κάποιο στοιχείο από τα **LBQIDs** του. Αν η αίτηση ταιριάζει με κάποιο στοιχείο $\mathbf{E}_j$ από τα **LBQIDs** του, αυτό σημαίνει ότι ο χρήστης έκανε την συγκεκριμένη αίτηση από μια θέση όπου συνηθίζει να επισκέπτεται (π.χ ένας χρήστης έκανε μια αίτηση από την δουλεία του). Σε αυτή την περίπτωση τα χώρο-χρονικά στοιχεία που στέλνονται μαζί με την αίτηση, θα πρέπει να τροποποιηθούν κατάλληλα και να σταλούν στον **παροχέα υπηρεσιών τροποποιημένα**, έτσι ώστε να μην υπάρχει κίνδυνος της ιδιωτικότητας του αιτούντα. Στην αντίθετη περίπτωση όπου τα χώρο-χρονικά δεδομένα της αίτησης, δεν ταιριάζουν με κανένα στοιχείο $\mathbf{E}_j$ από τα **LBQIDs** του χρήστη, τα χώρο-χρονικά στοιχεία της αίτησης δεν τροποποιούνται και στέλνονται ως έχουν στον παροχέα υπηρεσιών.
Η τροποποίηση των χώρο-χρονικών δεδομένων της αίτησης μερικές φορές δεν εκτελείται με επιτυχία. Σε αυτή τη περίπτωση, λέμε ότι ο αλγόριθμος γενίκευσης αποτυγχάνει. Όταν ο αλγόριθμος γενίκευσης αποτύχει, στην συνέχεια εκτελείται ο **αλγόριθμος αποσύνδεσης**. Ο αλγόριθμος αποσύνδεσης έχει σαν στόχο να μην είναι δυνατόν οι αιτήσεις που έχει κάνει ο αιτών στο παρελθόν, να διασυνδεθούν

με τις **μελλοντικές** του αιτήσεις. Για να επιτευχθεί αυτός ο στόχος, ο αλγόριθμος αποσύνδεσης δίνει στον χρήστη που έκανε την αίτηση ένα **νέο** αναγνωριστικό(id).

6.3 Αλγόριθμος Γενίκευσης.

Σε αυτή την ενότητα θα περιγράψουμε αναλυτικά τα βήματα του αλγορίθμου γενίκευσης, ο οποίος σαν στόχο έχει να διασφαλισθεί κ-ανωνυμία. Ο αλγόριθμος γενίκευσης προσπαθεί να γενικεύσει τα χώρο-χρονικά δεδομένα που επισυνάπτονται σε μια αίτηση που κάνει ένας χρήστης. Αυτή η γενίκευση γίνεται στον έμπιστο εξυπηρετητή, πριν σταλούν τα 'γενικευμένα' χώρο-χρονικά δεδομένα στον παροχέα υπηρεσιών. Αν ένας χρήστης κάνει μια αίτηση από την θέση: (x, y) την χρονική στιγμή t, τότε η τρισδιάστατη χώρο-χρονική πληροφορία (x,y,t) επισυνάπτεται στη αίτηση και στέλνεται στον έμπιστο εξυπηρετητή, ο οποίος γενικεύει την τρισδιάστατη πληροφορία(x,y,t). Πιο συγκεκριμένα η θέση (x, y) αντικαθίσταται από μια περιοχή-Area η οποία περιέχει την θέση (x, y). Η περιοχή έχει τη εξής μορφή:
Area=[(x1,y1),(x2,y2)], όπου το ζεύγος συντεταγμένων (x1,y1) αντιστοιχεί στην κάτω αριστερή κορυφή της περιοχής και το ζεύγος συντεταγμένων (x2,y2) αντιστοιχεί στην πάνω δεξιά κορυφή της περιοχής. Από την άλλη πλευρά η χρονική στιγμή t που έγινε η αίτηση αντικαθίσταται από ένα χρονικό διάστημα: [t1,t2] με t € [t1,t2] .
Η διαδικασία που περιγράψαμε ονομάζεται γενίκευση, καθώς η θέση από όπου έγινε η αίτηση γενικεύεται σε μια ευρύτερη περιοχή, όμοια η χρονική στιγμή που έγινε η αίτηση γενικεύεται σε ένα ευρύτερο χρονικό διάστημα. Πρέπει να πούμε ότι η γενίκευση του χρόνου και του χώρου σταματά, με βάση κάποιους χωρικούς αλλά και χρονικούς περιορισμούς που έχουμε ορίσει.
Ο αλγόριθμος της γενίκευσης δέχεται σαν είσοδο τα στοιχεία της αίτησης που έκανε ο χρήστης, δηλαδή την ακριβή θέση και χρονική στιγμή από όπου έγινε η αίτηση **(x,y,t)**. Επίσης σαν είσοδο δέχεται μια τιμή **'κ'** η οποία δηλώνει ότι από το σημείο που έγινε η αίτηση πρέπει να έχουν περάσει, τουλάχιστον 'κ' διαφορετικοί

χρήστες(συμπεριλαμβανομένου του αιτούντα) εντός των χρονικών περιορισμών. Άρα ο αλγόριθμος της γενίκευσης, γενικεύει τον χρόνο και τον χώρο(όσο το επιτρέπουν οι περιορισμοί που έχουμε ορίσει), μέχρι να βρει 'κ-1' διαφορετικούς κοντινότερους γείτονες.

Ο αλγόριθμος της γενίκευσης αποτελείται από δύο βασικά βήματα. Στο πρώτο βήμα ελέγχεται αν η αίτηση ταιριάζει με το πρώτο στοιχείο του LBQID του αιτούντα. Αν κάτι τέτοιο ισχύει, σημαίνει ότι η αίτηση έγινε από ένα σημείο το οποίο ανήκει στα μοτίβα μετακίνησης του αιτούντα και άρα πρέπει να γενικευθεί ο χώρος και ο χρόνος, πριν η αίτηση σταλεί στον παροχέα υπηρεσιών. Πιο συγκεκριμένα ο αλγόριθμος γενικεύει τον χώρο και τον χρόνο(όσο του επιτρέπουν οι χώρο-χρονικοί περιορισμοί) μέχρι να βρει 'κ-1' κοντινότερους γείτονες. Αν βρεθούν 'κ-1' διαφορετικοί γείτονες τότε λέμε ότι ο αλγόριθμος επιτυγχάνει κ-ανωνυμία και στην έξοδο του αλγορίθμου εμφανίζονται τα αναγνωριστικά των γειτόνων.

Το δεύτερο βήμα του αλγορίθμου είναι να ελέγξει αν η αίτηση που έκανε ο χρήστης ταιριάζει με ένα ενδιάμεσο στοιχείο από το LBQID του. Και σε αυτή την περίπτωση η αίτηση έχει γίνει από ένα σημείο που ανήκει στα μοτίβα μετακίνησης του χρήστη και πρέπει να ακολουθήσει γενίκευση. Σε αυτή την περίπτωση η διαφορά είναι ότι σαν είσοδο δεν δίνουμε την σταθερά 'κ' αλλά τα αναγνωριστικά των 'κ-1' κοντινότερων γειτόνων. Στην συνέχεια αφού έχουμε τα αναγνωριστικά των γειτόνων, ο αλγόριθμος ελέγχει για κάθε έναν γείτονα τα PHLs του. Αν και τα κ-1 PHLs είναι χώρο-χρονικά συνεπή με την αίτηση που έχει γίνει από τον αιτούντα τότε ο αλγόριθμος επιτυγχάνει κ-ανωνυμία.

6.3.1 Ψευδοκώδικας Αλγορίθμου γενίκευσης.

Ο αλγόριθμος γενίκευσης αποτελείται από δύο βασικά βήματα (γραμμή 2 και γραμμή 8). Αν η αίτηση που έχει κάνει ένας χρήστης ταιριάζει με το πρώτο στοιχείο από ένα LBQID τότε εκτελούνται τα βήματα 6 και 7. Σε αυτά τα βήματα επιλέγονται από τα PHLs όλων των χρηστών, κ-1 στιγμιότυπα(το κάθε στιγμιότυπο από **διαφορετικό** χρήστη). Τα στιγμιότυπα που επιλέγονται είναι αυτά που έχουν την μικρότερη απόσταση από την αίτηση.
Αν η αίτηση που κάνει ένας χρήστης ταιριάζει με ένα ενδιάμεσο στοιχείο από ένα LBQID τότε εκτελούνται τα βήματα 3 και 4. Σε αυτά τα βήματα πρέπει να πούμε ότι χρησιμοποιούνται τα PHLs των κ-χρηστών που έχουν επιλεγεί όταν η αίτηση είχε ταιριάξει με το πρώτο στοιχείο από το LBQID.
Το δεύτερο βασικό βήμα του αλγορίθμου είναι το βήμα 8. Σε αυτό το βήμα ελέγχεται αν ο τρισδιάστατος χώρος που έχει παραχθεί από την γενίκευση του χρόνου και του χώρου, ικανοποιεί τους περιορισμούς που έχουμε ορίσει στο βήμα 1. Αν οι περιορισμοί ικανοποιούνται τότε ο αλγόριθμος επιτυγχάνει κ-ανωνυμία. Αν δεν ικανοποιούνται οι περιορισμοί, τότε ο αλγόριθμος αποτυγχάνει. Σε αυτή περίπτωση δεν επιτυγχάνεται κ-ανωνυμία και πρέπει να ενημερώσουμε σχετικά τον χρήστη ότι η ιδιωτικότητα του είναι σε κίνδυνο.

Αλγόριθμος γενίκευσης

Είσοδος:

- Η ακριβής θέση και χρονική στιγμή της αίτησης (x,y,t).
- Η σταθερά '**κ**'.
- Τα αναγνωριστικά των ***κ-1*** κοντινότερων γειτόνων(αν η αίτηση ταιριάζει με ένα ενδιάμεσο στοιχείο του LBQID).

Έξοδος:

- Η γενικευμένη περιοχή- '**Area**' που περιέχει την ακριβή θέση από όπου έγινε η αίτηση, και τις θέσεις των ***κ-1*** γειτόνων.
- Το γενικευμένο χρονικό διάστημα-**Time interval**.
- Την τιμή μιας Boolean μεταβλητής: '**κ-ανωνυμία**' που δηλώνει την επιτυχία ή την αποτυχία του αλγορίθμου.
- Τα αναγνωριστικά των ***κ-1*** γειτόνων(μόνο στην περίπτωση ταιριάσματος με το 1° στοιχείο του LBQID).

Αλγόριθμος:

1. Όρισε τους χρονικούς και χωρικούς περιορισμούς.
2. Αν είσοδος είναι τα **κ-1** αναγνωριστικά των γειτόνων(ταίριασμα με ενδιάμεσο στοιχείο του LBQID),πήγαινε στο επόμενο βήμα, αλλιώς πήγαινε στο βήμα **5**.
3. Για κάθε ένα από τα **κ-1** αναγνωριστικά, ψάξε στο **PHL** του αντίστοιχου γείτονα και βρες εκείνο το χώρο-χρονικό στοιχείο που είναι πλησιέστερα στην αίτηση.
4. Γενίκευσε τον χρόνο και τον χώρο, έτσι ώστε η γενικευμένη τρισδιάστατη περιοχή(**Area, Time interval**), να περιέχει όλα τα **κ** χώρο-χρονικά στοιχεία(**κ-1** από τους γείτονες + **1** από τον αιτούντα).Πήγαινε στο βήμα **8**.
5. Αν είσοδος είναι η σταθερά '**κ**'(ταίριασμα με το πρώτο στοιχείο του LBQID του χρήστη), πήγαινε στο επόμενο βήμα, αλλιώς πήγαινε στο βήμα **11**.
6. Γενίκευσε τον χρόνο και τον χώρο, έτσι ώστε η γενικευμένη τρισδιάστατη περιοχή **(Area, Time interval)** να περιέχει τη αίτηση, καθώς και **κ-1** στιγμιότυπα από τις τροχιές των χρηστών. **Οι χρήστες και τα στιγμιότυπα των χρηστών επιλέγονται έτσι ώστε η τρισδιάστατη περιοχή να είναι η μικρότερη δυνατή**.
7. Αποθήκευσε τα αναγνωριστικά των χρηστών που βρέθηκαν στο προηγούμενο βήμα.
8. Αν η γενικευμένη περιοχή (**Area, Time interval**) ικανοποιεί τους χρονικούς και χωρικούς περιορισμούς θέσε : '**κ-ανωνυμία ==true**'.
9. Αλλιώς θέσε '**κ-ανωνυμία = =False**'.
10. Αν σαν είσοδος έχει δοθεί η σταθερά '**κ**' και '**κ-ανωνυμία ==true**', επέστρεψε τα **κ-1** αναγνωριστικά των γειτόνων.
11. Η αίτηση δεν έχει ταιριάξει με κανένα στοιχείο από τα LBQID και δεν χρειάζεται γενίκευση.

Αλγόριθμος γενίκευσης

6.3.2 Παράδειγμα εκτέλεσης Αλγορίθμου γενίκευσης

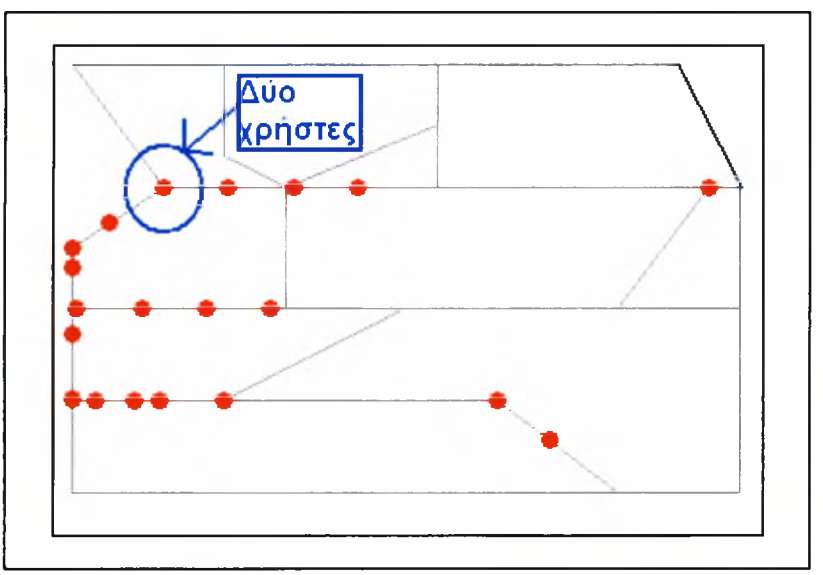
Σε αυτή την ενότητα, με την χρήση ενός παραδείγματος, θα παρουσιαστούν τα βήματα του αλγορίθμου γενίκευσης. Έστω λοιπόν ότι ο γεωγραφικός χώρος που μπορούν οι χρήστες του συστήματος μας να κινούνται και να κάνουν αιτήσεις, είναι το δίκτυο 'MYNET' που έχει κατασκευαστεί στη Oracle με την χρήση της γλώσσας SQL. Το δίκτυο 'MYNET' και απεικονίζεται στο σχήμα 4.5 .

Το πρώτο βήμα είναι να φορτωθεί το δίκτυο 'MYNET', στον πρόγραμμα παραγωγής χώρο-χρονικών αντικειμένων(generator) του Thomas Brinkhoff προκειμένου να παραχθούν οι κινούμενοι χρήστες. Για να φορτωθεί το δίκτυο 'MYNET' στο πρόγραμμα, έχουν πραγματοποιηθεί οι απαραίτητες αλλαγές στο format του δικτύου, έτσι ώστε να είναι συμβατό με το πρόγραμμα (generator), οι αλλαγές αυτές περιγράφονται στο κεφάλαιο 4.4. Αφού το δίκτυο φορτωθεί στο πρόγραμμα, το επόμενο βήμα είναι η παραμετροποίηση του προγράμματος. Η παραμετροποίηση έχει γίνει με τέτοιο τρόπο που να διευκολύνει την παρουσίαση του παραδείγματος. Ειδικότερα η παραμετροποίηση έχει γίνει έτσι ώστε το πρόγραμμα να παράγει **6** χρήστες, οι οποίοι κινούνται εντός του χρονικού διαστήματος **[0,5]**. Αυτό το χρονικό διάστημα είναι εκφρασμένο σε χρονικές μονάδες. Η παραμετροποίηση όπως βλέπουμε έχει γίνει έτσι ώστε το παράδειγμα να είναι όσο το δυνατόν πιο απλό. Στο κεφάλαιο 7 όπου παρουσιάζονται τα πειράματα, η παραμετροποίηση είναι διαφορετική και έτσι προσεγγίζονται πολύ καλύτερα πραγματικές συνθήκες.

Στο σχήμα 6.2 εικονίζεται η παραγωγή των χρηστών με την χρήση του προγράμματος παραγωγής χώρο-χρονικών δεδομένων του Thomas Brinkhoff. Στο σχήμα 6.2 έχουν παραχθεί τα στιγμιότυπα των **6** χρηστών, τα οποία απεικονίζονται όλα με κόκκινες κουκίδες. Επίσης στο σημείο που έχει σημειωθεί με **'μπλε κύκλο', υπάρχουν δύο χρήστες**, οι οποίοι έτυχε να έχουν τις ίδιες ακριβώς συντεταγμένες.

Όλα τα δεδομένα που έχει παράγει ο generator έχουν αποθηκευτεί στην βάση δεδομένων, σε έναν πίνακα που έχει δημιουργηθεί για αυτόν τον σκοπό. Αυτός ο πίνακας έχει ονομαστεί **'moving objects'** ,και περιέχει το ID του κάθε χρήστη καθώς και ένα σύνολο στοιχείων της μορφής (x,y,t) , που δηλώνουν τις θέσεις

που έχει βρεθεί ο κάθε χρήστης σε κάποιες χρονικές στιγμές. Στο σχήμα 6.3 φαίνονται τα στοιχεία που έχουν αποθηκευτεί στο πίνακα 'moving objects'. Τα στοιχεία αυτά, για ευκολία παρουσιάζονται με αύξοντα αριθμό ID (order by id). Από το σχήμα 6.3 βλέπουμε ότι ο κάθε χρήστης έχει διαφορετικό αριθμό από στιγμιότυπα(χώρο-χρονικά στοιχεία). Έτσι ο χρήστης με id=0 έχει 6 στιγμιότυπα, ο χρήστης με id=1 έχει 5 στιγμιότυπα, ο χρήστης με id=2 έχει 4 στιγμιότυπα κτλ. Επίσης διαπιστώνεται και η παρατήρηση ότι δύο χρήστες έχουν τις ίδιες ακριβώς συντεταγμένες. Οι χρήστες αυτοί σημειώνονται με κίτρινο χρώμα στο σχήμα 6.3.



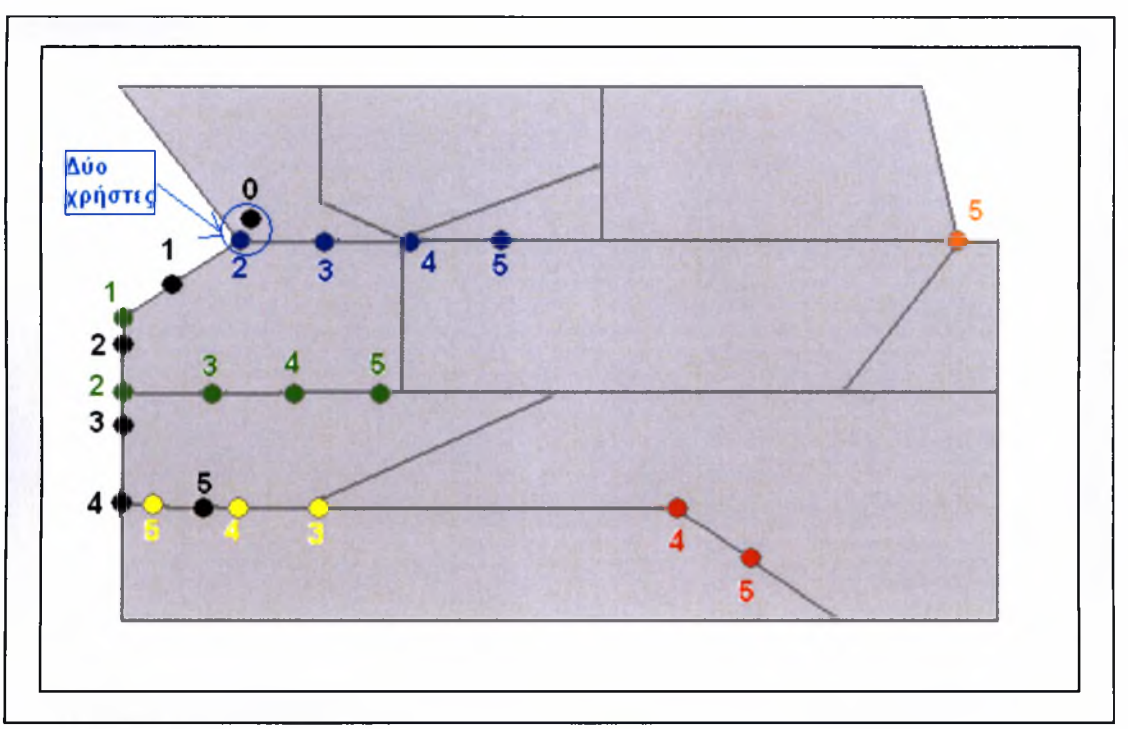
Σχήμα 6.2: παραγωγή κινούμενων χρηστών με Brinkhoff

ID	TIME	X	Y
0	0	4000	11000
0	1	2225	9816
0	5	3059	4000
0	4	1000	4073
0	3	1000	6206
0	2	1000	8339
1	3	3266	7000
1	1	1000	9000
1	2	1133	7000
1	5	7532	7000
1	4	5399	7000
2	5	10399	11000
2	2	4000	11000
2	4	8266	11000
2	3	6133	11000
3	5	1734	4000
3	4	3867	4000
3	3	6000	4000
4	4	15000	4000
4	5	16706	2720
5	5	22000	11000

Σχήμα 6.3: Δεδομένα του πίνακα 'moving objects'

Για λόγους καλύτερης παρουσίασης του παραδείγματος, με την βοήθεια του πίνακα που εικονίζεται στο σχήμα 6.3, αλλά και του δικτύου των χρηστών, το οποίο εικονίζεται στο σχήμα 6.2, κατασκευάστηκε το σχήμα 6.4. Αυτό το σχήμα περιέχει όλες τις απαραίτητες πληροφορίες συγκεντρωμένες. Πιο συγκεκριμένα:

- Οι μαύροι κόμβοι αντιστοιχούν στον Χρήστη με ID=0
- Οι πράσινοι κόμβοι αντιστοιχούν στον Χρήστη με ID=1
- Οι μπλε κόμβοι αντιστοιχούν στον Χρήστη με ID=2
- Οι κίτρινοι κόμβοι αντιστοιχούν στον Χρήστη με ID=3
- Οι κόκκινοι κόμβοι αντιστοιχούν στον Χρήστη με ID=4
- Ο πορτοκαλί κόμβος αντιστοιχεί στον Χρήστη με ID=5
- Δίπλα σε κάθε στιγμιότυπο κάθε χρήστη, σημειώνεται η χρονική στιγμή την οποία βρέθηκε στο αντίστοιχο σημείο. Για κάθε χρήστη οι χρονικές στιγμές εικονίζονται στο αντίστοιχο χρώμα.
- Τέλος στο σημείο που εικονίζεται με μπλε κύκλο υπάρχουν δύο χρήστες(μαύρος->ID=0 και μπλε->ID=2).



Σχήμα 6.4: Προτεινόμενη αναπαράσταση κινούμενων αντικειμένων

Επίσης πρέπει να οριστούν τα **PHLs** των χρηστών. Τα PHLs των χρηστών είναι της μορφής (x,y,t) και δηλώνουν ότι ο συγκεκριμένος χρήστης βρέθηκε στην θέση με συντεταγμένες(x,y) την χρονική στιγμή t. Από αυτή την θέση μπορεί να πραγματοποίησε κάποια αίτηση αλλά μπορεί και όχι. Αυτά τα PHLs των χρηστών προσδιορίζονται από τον έμπιστο εξυπηρετητή και αποθηκεύονται σε μια βάση δεδομένων που διαθέτει. Στο παρόν παράδειγμα, τα PHLs των χρηστών καθορίζονται από τα περιεχόμενα του **πίνακα 'moving objects'**(σχήμα 6.3). Τελικά τα PHLs των χρηστών είναι αυτά που εικονίζονται στο σχήμα 6.5.

PHL Χρήστη με ID=0

t	(x , y)
0	(4000 , 11000)
1	(2225 , 9816)
2	(1000 , 6206)
3	(1000 , 8339)
4	(1000 , 4073)
5	(3059 , 4000)

PHL Χρήστη με ID=1

t	(x , y)
1	(1000 , 9000)
2	(1133 , 7000)
3	(3266 , 7000)
4	(5399 , 7000)
5	(7532 , 7000)

PHL Χρήστη με ID=2

t	(x , y)
2	(4000 , 11000)
3	(6133 , 11000)
4	(8266 , 11000)
5	(10399 , 11000)

PHL Χρήστη με ID=3

t	(x , y)
3	(6000 , 4000)
4	(3867 , 4000)
5	(1734 , 4000)

PHL Χρήστη με ID=4

t	(x , y)
4	(15000 , 4000)
5	(16706 , 2720)

PHL Χρήστη με ID=5

t	(x , y)
5	(22000 , 11000)

Σχήμα 6.5: PHLs των 6 χρηστών

Το τελευταίο στοιχείο που πρέπει να ορισθεί προκειμένου να εφαρμοσθεί ο αλγόριθμος της αποσύνδεσης είναι τα **LBQIDs** των χρηστών. Για τις ανάγκες του παραδείγματος θεωρούμε ότι μόνο ο χρήστης με ID=0 θα πραγματοποιήσει

αιτήσεις, άρα αρκεί να ορίσουμε το LBQID του χρήστη με ID=0. Έστω λοιπόν ότι ο χρήστης με ID=0 έχει το πιο κάτω LBQID, το οποίο περιέχει 4 χώρο-χρονικά στοιχεία:

LBQID χρήστη με ID=0:

E_1={ (500 , 4000) , (1500 , 5000), [4,5] **}**
E_2={ (500 , 500) , (1500 , 1500), [1,2] **}**
E_3={ (11000 , 6000) , (12000 , 7000), [3,4] **}**
E_4={ (2500 , 2500) , (3500 , 3500), [2,3] **}**

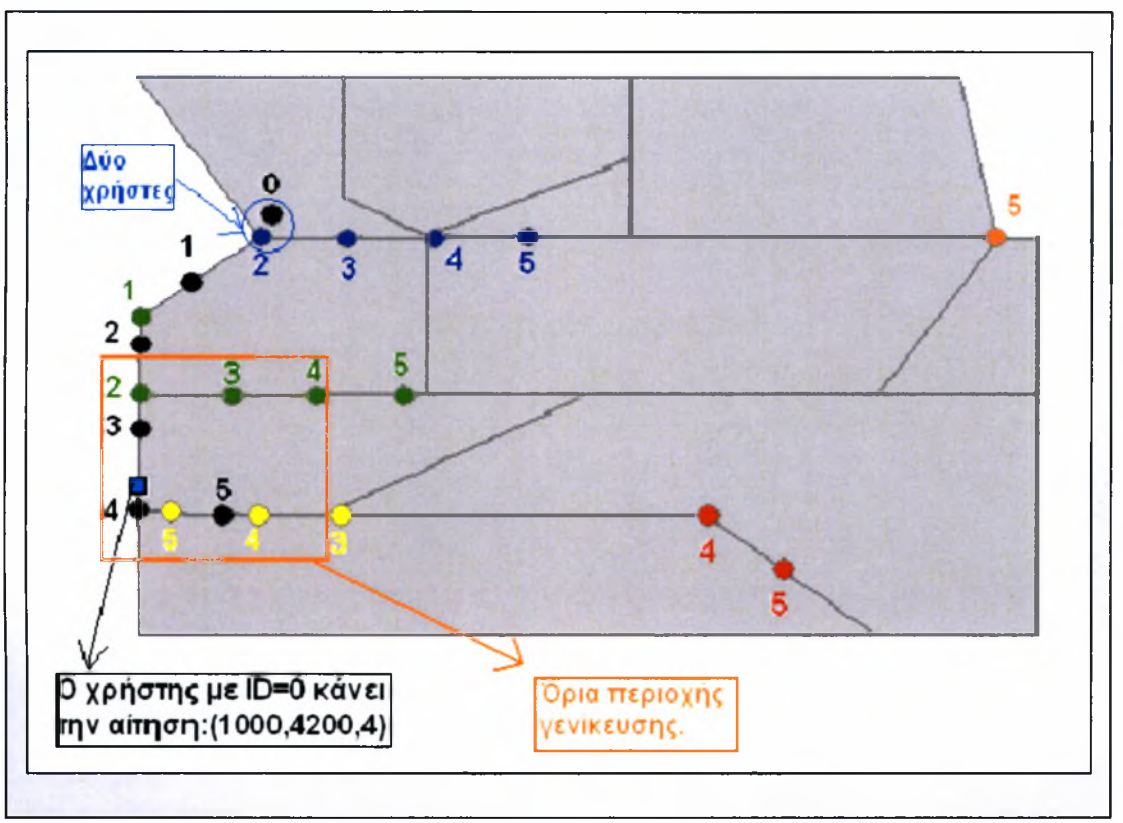
Έχοντας ορίσει όλα τα απαραίτητα στοιχεία, τώρα μπορούμε να τρέξουμε τον αλγόριθμο γενίκευσης. Έστω λοιπόν ότι δίνουμε σαν είσοδο στον αλγόριθμο γενίκευσης τα πιο κάτω στοιχεία:

- Τιμή **κ=3**
- Χρήστης που κάνει αίτηση= Χρήστης με ID=0
- Θέση από όπου έγινε η αίτηση=(1000,4200).
- Χρονική στιγμή που έγινε η αίτηση =4.
- Περιορισμοί γενίκευσης χώρου =τετράγωνη περιοχή πλευράς 4000 μέτρων.
- Περιορισμοί γενίκευσης χρόνου=2 χρονικές μονάδες.

Σε αυτή την περίπτωση, έχουμε ταίριασμα της αίτησης που έκανε ο χρήστης(με ID=0), με το πρώτο στοιχείο του LBQID του. Πιο συγκεκριμένα η θέση από όπου έγινε η αίτηση --> (1000,4200), είναι εντός της ορθογώνιας περιοχής που έχει ορισθεί για το **E_1** --> **{** (500 , 4000),(1500 , 5000) **}**. Όμοια η χρονική στιγμή που έγινε η αίτηση-->4 είναι εντός του χρονικού διαστήματος που έχει ορισθεί για το **E_1**-->[4,5].

Πρέπει να πούμε ότι για να επιτευχθεί κ-ανωνυμία θα πρέπει να βρεθούν κ-1 χρήστες που να ικανοποιούν τους χρονικούς αλλά και τους χωρικούς περιορισμούς. Άρα σε αυτό το παράδειγμα, πρέπει να βρεθούν 3-1=2 χρήστες που να ικανοποιούν τους περιορισμούς. **Έτσι αν βρεθούν 2 χρήστες που να ικανοποιούν τους περιορισμούς επιτυγχάνουμε 3-ανωνυμία.**

Στην συνέχεια ο αλγόριθμος συνεχίζει με την γενίκευση του χρόνου(2 χρονικές μονάδες) και του χώρου(τετράγωνη περιοχή πλευράς 4.000 μέτρα). Η διαδικασία της γενίκευσης παρουσιάζεται στο σχήμα 6.6.



Σχήμα 6.6:Γενικευμένη περιοχή που παράγεται από τον αλγόριθμο γενίκευσης.

Από το προηγούμενο σχήμα βλέπουμε ότι ο χρήστης με ID=0 κάνει μια αίτηση από την θέση: (1000,4200), τη χρονική στιγμή: 4. **Επίσης βλέπουμε ότι η γενικευμένη περιοχή είναι το 'πορτοκαλί τετράγωνο πλαίσιο' και περιέχει στιγμιότυπα του μαύρου(ID=0 & αιτών),του πράσινου(ID=1) και του κίτρινου(ID=3) χρήστη**. Συνεπώς οι χωρικοί περιορισμοί ικανοποιούνται για 2

χρήστες(IDs 1 & 3), εκτός του αιτούντα. Αυτή η πληροφορία στην πραγματική εκτέλεση του αλγορίθμου εξάγεται από τα PHLs των χρηστών (σχήμα 6.5.). Σε αυτό το παράδειγμα για απλότητα ενσωματώσαμε τις πληροφορίες των PHLs στο σχήμα 6.6.

Στην συνέχεια ο αλγόριθμος ελέγχει αν ικανοποιούνται οι χρονικοί περιορισμοί. Η αίτηση έγινε την χρονική στιγμή: 4, και οι χρονικοί περιορισμοί είναι 2 χρονικές μονάδες, άρα εντός της γενικευμένης χωρικής περιοχής πρέπει να βρεθούν **τουλάχιστον 2** στιγμιότυπα **διαφορετικών** χρηστών(με βάση τα PHLs τους) που να **μην** απέχουν χρονικά από την αίτηση **περισσότερο από 2 χρονικές μονάδες**. Με βάση το σχήμα 6.6 βλέπουμε ότι ο κίτρινος χρήστης(ID=3) βρίσκεται εντός των χωρικών ορίων, τις χρονικές στιγμές: 4 & 5, άρα απέχει χρονικά από την αίτηση 4-4=0 χρονικές μονάδες και 5-4= 1 χρονική μονάδα αντίστοιχα. Συνεπώς η ελάχιστη χρονική απόσταση από την αίτηση, για αυτόν τον χρήστη είναι 0 χρονικές μονάδες. Αυτή η χρονική απόσταση είναι μικρότερη από τους χρονικούς περιορισμούς που έχουμε ορίσει.

Με όμοιο σκεπτικό ο πράσινος χρήστης(ID=1), εντός των χωρικών ορίων της γενίκευσης, έχει βρεθεί τις χρονικές στιγμές: 2, 3 & 4 και οι χρονικές αποστάσεις από την αίτηση είναι 4-2=2, 4-3=1 και 4-4=0 χρονικές μονάδες αντίστοιχα. Η ελάχιστη χρονική απόσταση είναι '0 χρονικές μονάδες' < '2 χρονικές μονάδες' (Χρονικοί περιορισμοί). Τελικά αφού βρέθηκαν **2 διαφορετικοί** χρήστες που να ικανοποιούν τους χώρο-χρονικούς περιορισμούς, ο αλγόριθμος γενίκευσης επιτυγχάνει **3-ανωνυμία**. Στην έξοδο ο αλγόριθμος εκτός από την γενικευμένη περιοχή και τον γενικευμένο χρόνο, επιστρέφει και τα κ-1 αναγνωριστικά των κοντινότερων γειτόνων που βρήκε, αφού έχουμε ταίριασμα με το πρώτο στοιχείο του LBQID. Σε αυτό το παράδειγμα δηλαδή θα επιστραφούν τα 3-1=2 αναγνωριστικά, των γειτόνων(δηλαδή τα IDs 1 και 3).

Αν σαν είσοδο αντί για την τιμή του 'κ' δίναμε τα αναγνωριστικά των κ-1 γειτόνων(τα οποία προέρχονται από προηγούμενη εκτέλεση του αλγορίθμου) τότε ο αλγόριθμος θα έλεγχε αν η αίτηση ταιριάζει με ένα ενδιάμεσο στοιχείο του LBQID. Σε αυτό το παράδειγμα δηλαδή αν δίναμε σαν είσοδο τα αναγνωριστικά 1 και 3(έχουν βρεθεί από την προηγούμενη εκτέλεση ταιριάσματος με το πρώτο στοιχείο του LBQID), ο αλγόριθμος θα έψαχνε στα PHLs των γειτόνων(1 και 3) προκειμένου να βρει τα πιο κοντινά (στην αίτηση), χώρο-χρονικά στιγμιότυπα

τους. Στην συνέχεια θα ακολουθούσε γενίκευση του χρόνου και του χώρου και αν η τρισδιάστατη περιοχή ικανοποιούσε τους χωρικούς αλλά και τους χρονικούς περιορισμούς, τότε ο αλγόριθμος θα επιτύγχανε 3-ανωνυμία. Σε αυτή την περίπτωση ο αλγόριθμος παράγει στην έξοδο την τρισδιάστατη χώρο-χρονική περιοχή αλλά όχι τα αναγνωριστικά των γειτόνων.

6.4 Αλγόριθμος αποσύνδεσης με την χρήση μικτών ζωνών.

Ο αλγόριθμος της γενίκευσης μερικές φορές, υπάρχει περίπτωση να μην καταφέρει να εξασφαλίσει κ-ανωνυμία. Στην περίπτωση λοιπόν που ο αλγόριθμος γενίκευσης αποτύχει, στη συνέχεια εκτελείται ο αλγόριθμος αποσύνδεσης. Βασικός στόχος του αλγορίθμου αποσύνδεσης είναι η αλλαγή του αναγνωριστικού του αιτούντα. Η αλλαγή του αναγνωριστικού γίνεται για να μην υπάρχει ο κίνδυνος της διασύνδεσης των **προηγούμενων** αιτήσεων που έχει κάνει ο αιτών, με αυτές τις αιτήσεις που ενδεχομένως θα κάνει στο **μέλλον**. Όσον αφορά την αλλαγή του αναγνωριστικού δε γίνεται τυχαία οποιαδήποτε χρονική στιγμή αλλά πρέπει να ικανοποιούνται κάποιες προϋποθέσεις. Ειδικότερα η αλλαγή του αναγνωριστικού γίνεται μόνο αν ο αιτών βρίσκεται εντός των χώρο-χρονικών ορίων μια περιοχής, η οποία ονομάζεται **μικτή ζώνη** (mix-zone). Οι μικτές ζώνες είναι τοποθεσίες μέσα στις οποίες οι χρήστες δεν μπορούν να έχουν πρόσβαση σε καμία από τις υπηρεσίες που παρέχουν οι παροχείς υπηρεσιών.

Όταν ένας χρήστης κάνει μια αίτηση και ο αλγόριθμος γενίκευσης αποτύχει, στην συνέχεια εκτελείται ο αλγόριθμος αποσύνδεσης. Για να τρέξει αλγόριθμος αποσύνδεσης πρέπει να έχει οριστεί μια μικτή ζώνη. Η στρατηγική που ακολουθήθηκε είναι, οι μικτές ζώνες να μην ορίζονται στατικά από την αρχή, αλλά να δημιουργούνται κάθε φορά εκ' νέου, κάθε φορά που τρέχει ο αλγόριθμος

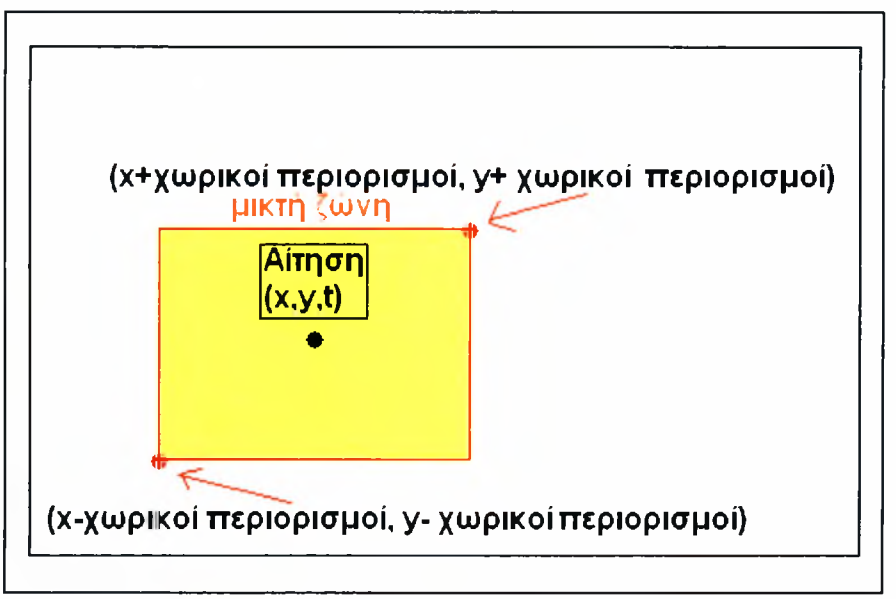
αποσύνδεσης. Αφού οριστεί η μικτή ζώνη, το επόμενο βήμα είναι να ελεγχθεί αν στα χώρο-χρονικά όρια που έχουν ορισθεί για την μικτή ζώνη, υπάρχουν **'αρκετοί'** χρήστες έτσι ώστε η αλλαγή του αναγνωριστικού του αιτούντα, να μπορεί να διασφαλίσει την ιδιωτικότητα του. Αν δεν υπάρχουν αρκετοί χρήστες εντός των ορίων της μικτής ζώνης, ακόμα και να γίνει αλλαγή του αναγνωριστικού του αιτούντα, ο παροχέας υπηρεσιών ίσως να μπορέσει να συνδέσει παλιές αιτήσεις του αιτούντα, με καινούριες αιτήσεις που θα πραγματοποιήσει στο μέλλον και άρα θα θιγεί η ιδιωτικότητα του αιτούντα.

Τέλος αν βρεθούν αρκετοί χρήστες εντός των χώρο-χρονικών ορίων της μικτής ζώνης, τότε ανατίθεται στον αιτούντα ένα νέο μοναδικό αναγνωριστικό, το οποίο δεν έχει χρησιμοποιηθεί στο παρελθόν από κανέναν άλλον χρήστη. Επίσης σβήνονται από την βάση δεδομένων στον έμπιστο εξυπηρετητή, όλα τα στοιχεία του PHL και του LBQID του αιτούντα. Έτσι ο χρήστης για το σύστημα είναι σαν ένας εντελώς καινούριος χρήστης. Αντίθετα αν δεν βρεθούν αρκετοί χρήστες εντός των χώρο-χρονικών ορίων της μικτής ζώνης τότε ο αλγόριθμος αποσύνδεσης αποτυγχάνει και στέλνεται στον χρήστη κατάλληλο μήνυμα που τον προειδοποιεί ότι η ιδιωτικότητα του κινδυνεύει.

Πρέπει να πούμε ότι ο αλγόριθμος αποσύνδεσης αποτυγχάνει είτε αν δεν βρεθούν αρκετοί χρήστες εντός της μικτής ζώνης, είτε αν η αίτηση που έχει πραγματοποιηθεί, έχει ταιριάξει με το τελευταίο στοιχείο του LBQID του αιτούντα. Στην περίπτωση που έχουμε ταίριασμα με το τελευταίο στοιχείο του LBQID του αιτούντα, έχουμε αποτυχία, διότι υπάρχει πιθανότητα τα προηγούμενα στοιχεία του LBQID να έχουν ταιριάξει με προηγούμενες αιτήσεις που έχει κάνει ο αιτών με το παλιό του αναγνωριστικό. Έτσι ακόμα και αν δοθεί στον χρήστη ένα νέο αναγνωριστικό, ο παροχέας υπηρεσιών ίσως έχει την δυνατότητα να διασυνδέσει τις παλιές αιτήσεις του αιτούντα με της μελλοντικές του αιτήσεις, κάτι τέτοιο όμως έχει σαν αποτέλεσμα να μην διασφαλίζεται η ιδιωτικότητα του χρήστη.

Όπως προαναφέραμε οι μικτές ζώνες ορίζονται δυναμικά κάθε φορά που καλείται ο αλγόριθμος αποσύνδεσης. Για να κατανοηθεί καλύτερα η έννοια της μικτής ζώνης δίνεται το ακόλουθο παράδειγμα :

Έστω ότι ένας χρήστης κάνει μια αίτηση από τη θέση **(x,y)** την χρονική στιγμή **t**. Επίσης στην είσοδο του αλγορίθμου γενίκευσης έχουν δοθεί οι χωρικοί και οι χρονικοί περιορισμοί, σχετικά με τα όρια της γενίκευσης. Σαν μικτή ζώνη θεωρείται η ορθογώνια περιοχή **με κέντρο τις συντεταγμένες (x,y)**, και με άκρα τα ζεύγη συντεταγμένων της κάτω αριστερής και της πάνω δεξιάς κορυφής. Ειδικότερα η κάτω αριστερή κορυφή της μικτής ζώνης έχει συντεταγμένες: **(x - 'χωρικοί περιορισμοί' , y - 'χωρικοί περιορισμοί'),** και αντίστοιχα η πάνω δεξιά κορυφή της μικτής ζώνης έχει συντεταγμένες:
(x + 'χωρικοί περιορισμοί' , y + 'χωρικοί περιορισμοί'). Όσον αφορά τα χρονικά όρια τη μικτής ζώνης ορίζονται με όμοιο τρόπο από το χρονικό διάστημα: **[t - 'χρονικοί περιορισμοί' , t + 'χρονικοί περιορισμοί']**. Ένα παράδειγμα μιας μικτής ζώνης φαίνεται στο σχήμα 6.7, όπου με κίτρινο χρώμα εικονίζονται τα χωρικά όρια της μικτής ζώνης.



Σχήμα 6.7: Μικτή ζώνη.

6.4.1 Ψευδοκώδικας Αλγορίθμου αποσύνδεσης

Σε αυτή την ενότητα θα περιγραφεί ο τρόπος με τον οποίο έχουν υλοποιηθεί τα βασικά βήματα, του αλγορίθμου αποσύνδεσης. Αρχικά σαν είσοδο, ο αλγόριθμος αποσύνδεσης, δέχεται όλα τα στοιχεία που του είναι απαραίτητα από τον αλγόριθμο γενίκευσης, ο οποίος έχει ήδη εκτελεστεί και έχει αποτύχει. Στα βήματα **1** & **2** παράγεται η μικτή ζώνη για τον χρήστη που έκανε την αίτηση. Κάτι που είναι σημαντικό να αναφερθεί, είναι ότι η μικτή ζώνη προσδιορίζεται πάντα από το κέντρο της, που είναι το σημείο από όπου έγινε η αίτηση (x,y), αλλά δεν έχει πάντα τα ίδια χωρικά και χρονικά όρια. Αυτό συμβαίνει διότι ο τρόπος που υπολογίζονται τα χώρο-χρονικά όρια της μικτής ζώνης, μπορεί να παράγει μια μικτή ζώνη, που να είναι εκτός των ορίων του γεωγραφικού χώρου που έχει οριστεί ότι επιτρέπεται να κινούνται και να κάνουν αιτήσεις οι χρήστες.
Προκειμένου λοιπόν να μην παραχθεί μια μικτή ζώνη που να είναι εκτός των ορίων του γεωγραφικού χώρου που έχει ορισθεί, εφαρμόζεται η ακόλουθη στρατηγική: Αρχικά ορίζουμε τα χωρικά όρια της μικτής ζώνης, σαν ένα ορθογώνιο, οι συντεταγμένες της κάτω αριστερής κορυφής του ορθογωνίου είναι: **(x - 'χωρικοί περιορισμοί' , y - 'χωρικοί περιορισμοί'),** και οι συντεταγμένες της πάνω δεξιάς κορυφής του ορθογωνίου είναι: **(x + 'χωρικοί περιορισμοί' , y + 'χωρικοί περιορισμοί').** Αν κάποια από τις συντεταγμένες που παρήχθησαν, είναι εκτός του γεωγραφικού χώρου που έχει ορισθεί, τότε **όλες** οι συντεταγμένες αρχίζουν και μειώνονται **σταδιακά** και **ομοιόμορφα**. Οι συντεταγμένες μειώνονται σταδιακά για να παραχθεί όσο το δυνατόν πιο μεγάλη σε έκταση μικτή ζώνη και ομοιόμορφα για να είναι η τελική μικτή ζώνη μια ορθογώνια περιοχή και όχι ένα ανομοιόμορφο σχήμα. Το ίδιο ακριβώς σκεπτικό, εφαρμόζεται για τα χρονικά όρια που ορίζονται για μια μικτή ζώνη, αν αυτά ξεπεράσουν τα προκαθορισμένα όρια.
Στην συνέχεια στο βήμα **3** ελέγχεται αν η αίτηση έχει ταιριάξει με το τελευταίο στοιχείο του LBQID του αιτούντα, αν κάτι τέτοιο συμβαίνει, τότε ο αλγόριθμος αποσύνδεσης αποτυγχάνει. Στο βήμα **4** υπολογίζεται ο αριθμός των χρηστών που έχουν βρεθεί εντός των χωρικών αλλά και των χρονικών ορίων της μικτής ζώνης. Αν αυτός ο αριθμός είναι 'ικανοποιητικός' ελέγχεται στο βήμα **5**, το οποίο είναι και το πιο σημαντικό βήμα του αλγορίθμου.

Για την κατανόηση του βήματος **5** του αλγορίθμου αποσύνδεσης πρέπει να γίνουν κάποιες διευκρινίσεις. Πρώτα απ' όλα πρέπει να πούμε ότι η τιμή **'κ'** που δίνεται σαν είσοδος στον αλγόριθμο γενίκευσης έχει σχέση με την τιμή **'l'** η οποία επιστρέφεται από τον αλγόριθμο αποσύνδεσης αλλά οι δύο αυτές τιμές **δεν ταυτίζονται**. Πιο συγκεκριμένα ο αριθμός **'l'** που παράγεται στην έξοδο του αλγορίθμου αποσύνδεσης, έχει επιλεγεί να είναι μεγαλύτερος από τον αριθμό **'κ'** που έχει δοθεί σαν είσοδος στον αλγόριθμο γενίκευσης **(δηλαδή l>κ υποχρεωτικά)**. Η λογική που κρύβεται πίσω από αυτή την προσέγγιση είναι η εξής: Όπως έχουμε προαναφέρει, για να εκτελεστεί ο αλγόριθμος της αποσύνδεσης, πρέπει να έχει εκτελεστεί πρώτα ο αλγόριθμος της γενίκευσης και να έχει αποτύχει. Για να αποτύχει όμως ο αλγόριθμος γενίκευσης σημαίνει ότι δεν κατάφερε να βρει 'κ-1' γείτονες οι οποίοι να ικανοποιούν τους χώρο-χρονικούς περιορισμούς της γενίκευσης. Στην συνέχεια, αφού έχει αποτύχει ο αλγόριθμος της γενίκευσης, ο αλγόριθμος αποσύνδεσης αρχίζει την εκτέλεση του και παράγει την μικτή ζώνη η οποία είναι μια περιοχή αρκετά μεγαλύτερη από την περιοχή που ορίζουν οι χώρο-χρονικοί περιορισμοί της γενίκευσης. Αν λοιπόν απαιτούσαμε από τον αλγόριθμο της αποσύνδεσης να βρει εντός της μικτής ζώνης 'κ' (κ από αλγόριθμο γενίκευσης) χρήστες, αφού η μικτή ζώνη είναι μια αρκετά μεγάλη περιοχή, τις περισσότερες φορές ο αλγόριθμος αποσύνδεσης θα είχε επιτυχία και 'υποτίθεται' ότι θα διασφάλιζε την ιδιωτικότητα του συγκεκριμένου χρήστη. Η πιο πάνω όμως θεώρηση είναι λανθασμένη, διότι για την ίδια τιμή 'κ' ο αλγόριθμος γενίκευσης είχε αποφανθεί ότι η ιδιωτικότητα του χρήστη είναι σε κίνδυνο. Για την λύση αυτού του προβλήματος θεωρήθηκε ότι η τιμή **'l' (αλγόριθμος αποσύνδεσης)** αρχικά έχει την τετραπλάσια τιμή από την τιμή **'κ' (αλγόριθμος γενίκευσης)** και ότι σταδιακά μειώνεται, χωρίς όμως να γίνει μικρότερη από την τιμή **'κ'** του αλγορίθμου γενίκευσης.

Στο βήμα 5 λοιπόν ελέγχεται αν έχουν βρεθεί 'k-1' διαφορετικοί χρήστες. Αν δεν έχουν βρεθεί, τότε ο αλγόριθμος αποτυγχάνει, αντίθετα αν έχουν βρεθεί 'k-1' διαφορετικοί χρήστες τότε ο αλγόριθμος επιτυγχάνει και συνεχίζει στα βήματα 8 και 9 στα οποία ανατίθεται ένα νέο αναγνωριστικό στον αιτούντα και σβήνονται όλα τα στοιχεί από το παλιό του LBQID, καθώς και από το PHL του.

Αλγόριθμος αποσύνδεσης

Είσοδος :

- Το αναγνωριστικό του αιτούντα.
- Τα χώρο-χρονικά στοιχεία της αίτησης (x,y,t).
- Η τιμή κ(από τον αλγόριθμο γενίκευσης).
- Οι χωρικοί και χρονικοί περιορισμοί της γενίκευσης.
- **Όλα τα πιο πάνω στοιχεία έχουν δοθεί σαν είσοδο στον αλγόριθμο γενίκευσης και δεν χρειάζεται να δοθούν ξανά.**

Έξοδος:

- Η τιμή της Boolean μεταβλητής **l**-αποσύνδεση, η οποία δηλώνει την επιτυχία ή την αποτυχία του αλγορίθμου.
- Η τιμή **'l'** , η οποία δηλώνει τον βαθμό της αποσύνδεση που επιτεύχθηκε(δηλαδή πόσοι χρήστες βρέθηκαν εντός της μικτής ζώνης).

Αλγόριθμος :

1) Όρισε τα χωρικά όρια της μικτής ζώνης, με βάση τα χωρικά στοιχεία της αίτηση και τους χωρικούς περιορισμούς της γενίκευσης.
2) Όρισε τα χρονικά όρια της μικτής ζώνης, με βάση τα χρονικά στοιχεία της αίτηση και τους χρονικούς περιορισμούς της γενίκευσης.
3) Έλεγξε αν η αίτηση έχει ταιριάξει με το **τελευταίο** στοιχείο του LBQID του αιτούντα αν ναι θέσε **l**-αποσύνδεση=false, αλλιώς συνέχισε στο επόμενο βήμα.
4) Ψάξε να βρεις πόσοι διαφορετικοί χρήστες έχουν βρεθεί εντός των χωρικών αλλά και τον χρονικών ορίων της μικτής ζώνης.
5) Αν ο αριθμός των χρηστών είναι τουλάχιστο **'l-1'** (+ 1 ο αιτών) πήγαινε στο επόμενο βήμα αλλιώς θέσε **l**-αποσύνδεση =false.
6) Θέση **l**- αποσύνδεση=true.
7) Επέστεψε την τιμή **'l'** .
8) Δώσε στον αιτών ένα νέο μοναδικό αναγνωριστικό.
9) Σβήσε όλα τα στοιχεία από το PHL και το LBQID του αιτών.

Σχήμα 6.8:Αλγόριθμος αποσύνδεσης

6.4.2 Παράδειγμα εκτέλεσης Αλγορίθμου αποσύνδεσης

Για τις ανάγκες αυτού του παραδείγματος, θεωρούμε τα ίδια δεδομένα που έχουν προσδιοριστεί στο παράδειγμα του αλγορίθμου γενίκευσης. Δηλαδή θεωρούμε ότι με το πρόγραμμα παραγωγής χώρο-χρονικών δεδομένων του Brinkhoff, έχουν παραχθεί οι χρήστες που φαίνονται στο σχήμα 6.2, επιπλέον τα δεδομένα για αυτούς τους χρήστες είναι αποθηκευμένα στην βάση δεδομένων με την μορφή που εικονίζεται στο σχήμα 6.3, και τα PHLs των χρηστών είναι αυτά που φαίνονται στο σχήμα 6.5.

Έστω ότι στον αλγόριθμο γενίκευσης έχουν δοθεί στην είσοδο τα πιο κάτω δεδομένα:

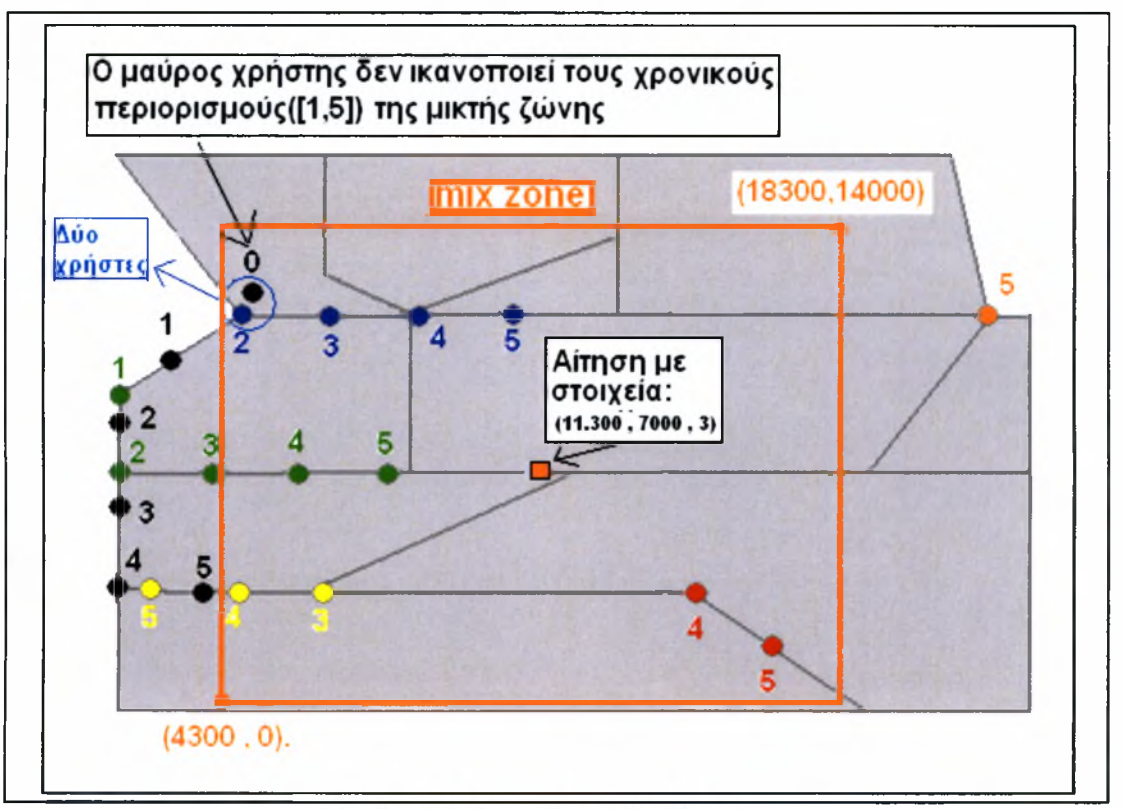
- Τιμή κ=3
- Χρήστης που κάνει αίτηση= Χρήστης με ID=0
- Θέση από όπου έγινε η αίτηση=(11300,7000).
- Χρονική στιγμή που έγινε η αίτηση =3.
- Περιορισμοί γενίκευσης χώρου =τετράγωνη περιοχή πλευράς 7000 μέτρων.
- Περιορισμοί γενίκευσης χρόνου=2 χρονικές μονάδες.

Ο αλγόριθμος γενίκευσης σε αυτή τη περίπτωση αποτυγχάνει, άρα θα εκτελεστεί ο αλγόριθμος αποσύνδεσης. Από την εφαρμογή του αλγορίθμου γενίκευσης ξέρουμε ότι η συγκεκριμένη αίτηση έχει ταιριάξει με ένα ενδιάμεσο στοιχείο του LBQID του χρήστη με ID=0. Πιο συγκεκριμένα έχει ταιριάξει με το στοιχείο $\mathbf{E}_3$ (βλέπε LBQID χρήστη με ID=0). Αν η αίτηση είχε ταιριάξει με το τελευταίο στοιχείο του LBQID, δηλαδή με το $\mathbf{E}_4$ τότε ο αλγόριθμος αποσύνδεσης θα αποτύγχανε.

Ο αλγόριθμος αποσύνδεσης αρχικά παράγει την μικτή ζώνη (τα χώρο-χρονικά τις όρια). Έτσι πρώτα παράγονται οι συντεταγμένες της πάνω δεξιάς κορυφής της μικτής ζώνης ως εξής :

(11300+7000, 7000+7000)=**(18300,14000)**. Στην συνέχεια παράγονται οι συντεταγμένες της κάτω αριστερής κορυφής:(11300-7000, 7000-7000)=**(4300,0).** Τέλος παράγονται τα χρονικά όρια της μικτής ζώνης:[3-2,3+2]=**[1,5].**

Όλα αυτά τα στοιχεία της μικτής ζώνης εικονίζονται στο σχήμα 6.9. Από το σχήμα αυτό, βλέπουμε ότι στα **χωρικά** όρια της μικτής ζώνης υπάρχουν **5** χρήστες μαζί με τον αιτούντα. Οι χρήστες αυτοί είναι οι εξής: μαύρος χρήστης (ID=0), πράσινος χρήστης (ID=1), μπλε χρήστης (ID=2), κίτρινος χρήστης (ID=3), κόκκινος χρήστης (ID=4). Από αυτούς τους 5 χρήστες εντός των **χρονικών** ορίων της μικτής ζώνης ([1,5]) βρίσκονται **4** χρήστες. Ο μαύρος χρήστης όπως βλέπουμε από το σχήμα 6.9. έχει βρεθεί στα χωρικά όρια της μικτής ζώνης αλλά βρέθηκε σε αυτή τη θέση, την χρονική στιγμή **'0'** που **δεν ανήκει στο διάστημα [1,5]**, και άρα δεν ικανοποιεί τους χρονικούς περιορισμούς της μικτής ζώνης.



Σχήμα 6.9:Στοιχεία μικτής ζώνης

Αρχικά ο αλγόριθμος προσπαθεί να βρει **'l-1'** διαφορετικούς χρήστες εντός την μικτής ζώνης. Το **l** του αλγορίθμου της αποσύνδεσης όπως έχει προαναφερθεί έχει αρχικά τιμή: **l=4*κ**, δηλαδή στα πλαίσια αυτού του παραδείγματος έχει τιμή: **l=4*3=12**. Αρχικά ο αλγόριθμος αποσύνδεσης προσπαθεί να βρει **'l-1'=12-1=11**

διαφορετικούς χρήστες οι οποίοι να είναι εντός των χωρικών ορίων της μικτής ζώνης και παράλληλα να ικανοποιούνται και οι χρονικοί περιορισμοί της μικτής ζώνης. Το σύνολο των χρηστών μας σε αυτό το παράδειγμα είναι **6**, άρα δεν είναι δυνατόν να βρεθούν 11 χρήστες. Στην συνέχεια σταδιακά ο αλγόριθμος μειώνει το **l**, Το οποίο παίρνει διαδοχικά τις τιμές: 12, 11, 10, 9, 8, 7, 6 και 5 . Όταν φτάσει στην τιμή 5, αρκεί να βρεθούν 5-1=4 χρήστες που να ικανοποιούν τους χώρο-χρονικούς περιορισμούς της μικτής ζώνης. Όντως υπάρχουν 4 τέτοιοι χρήστες και άρα ο αλγόριθμος επιτυγχάνει. Στην συνέχεια δίνεται στον μαύρο χρήστη(παλιό ID=0) το ID=9, το οποίο δεν έχει χρησιμοποιηθεί από κανέναν άλλον χρήστη. Επίσης σβήνονται όλα τα στοιχεία από το LBQID και το PHL του αιτούντα.

Αυτό που είναι πολύ σημαντικό και πρέπει να το τονιστεί, διότι δεν είναι άμεσα προφανές από την συγκεκριμένη εκτέλεση του αλγορίθμου είναι το εξής: Στην προηγούμενη εκτέλεση του αλγορίθμου η τιμή του **l** πήρε αρχικά την τιμή 12 και στη συνέχεια μειωνόταν σταδιακά κατά 1 μονάδα. Πιο συγκεκριμένα πήρε τις τιμές:12, 11, 10, 9, 8, 7, 6, και 5, και τελικά βρέθηκαν 5-1=4 χρήστες που να ικανοποιούν τα χώρο-χρονικά όρια της μικτής ζώνης. Αν κάτι τέτοιο δεν ήταν δυνατό να συμβεί, τότε ο αριθμός **l δεν θα μειωνόταν συνεχώς** . Αντίθετα θα μπορούσε να μειωθεί, μέχρι να φτάσει την τιμή του αριθμού **κ**, που είχε δοθεί στην είσοδο του αλγορίθμου της γενίκευσης. Αν το **l** έφτανε μέχρι την τιμή **κ** (από γενίκευση) τότε ο αλγόριθμος αποσύνδεση θα αποτύγχανε.

ΚΕΦΑΛΑΙΟ 7

7. Πειραματικά Δεδομένα

Σε αυτή την ενότητα θα παρουσιαστούν πειραματικά δεδομένα, που έχουν προκύψει από την εφαρμογή των αλγορίθμων που έχουν προταθεί στο κεφάλαιο 6. Με αυτόν τον τρόπο θα αξιολογηθεί η στρατηγική που έχει προταθεί.

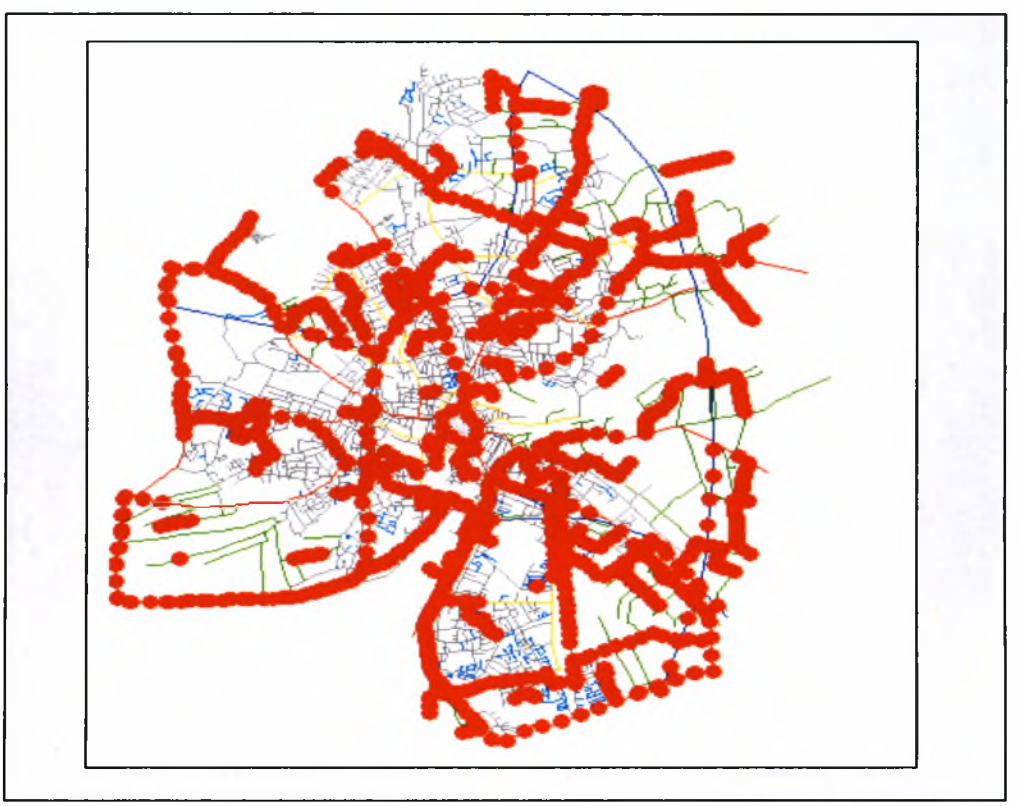
Για την διεξαγωγή των πειραμάτων προτιμήθηκε η χρήση του δικτύου oldenburg αντί του δικτύου MYNET. Επίσης θεωρήσαμε ότι οι χρήστες μπορούν να κινούνται εντός των χωρικών ορίων του δικτύου **'oldenburg'**. Το δίκτυο αυτό έχει κατασκευαστεί έτσι ώστε να περιλαμβάνεται εντός μίας τετράγωνης περιοχής. Αυτή η περιοχή προσδιορίζεται από τις συντεταγμένες της κάτω αριστερής κορυφής: **(0,0)** και τις συντεταγμένες της πάνω δεξιάς κορυφής: **(30.000, 30.000)**. Οι τιμές των συντεταγμένων αυτών είναι εκφρασμένες σε μέτρα. Επίσης έχει θεωρηθεί ότι οι κινούμενοι χρήστες μπορούν να αποστέλλουν αιτήσεις, εντός του χρονικού διαστήματος: **[0,23].** Επίσης η ταχύτητα μετακίνησης των χρηστών, έχει την τιμή **50**. Στο πρόγραμμα του Brinkhoff η ταχύτητα των χρηστών, δεν ορίζεται με την εισαγωγή μιας τιμής που αναπαριστά την τιμή της ταχύτητας. Για παράδειγμα δεν γίνεται να ορισθεί συγκεκριμένα, ότι οι χρήστες κινούνται με 30χμ/ώρα. Αντίθετα, για να προσδιορισθεί η ταχύτητα των κινούμενων χρηστών, εισάγουμε έναν αριθμό, ο οποίος πρέπει να ανήκει στο διάστημα [10,250]. Αυτός ο αριθμός καθορίζει την ταχύτητα μετακίνησης των χρηστών. Ειδικότερα για την σημασία της τιμής αυτής, ισχύουν οι εξής αντιστοιχίες: (10=fast , 50=medium , 250 slow).

Όσον αφορά τον αριθμό των χρηστών έχουμε θεωρήσει τις εξής περιπτώσεις: 100, 500 και 1000 χρήστες. Στο σχήμα 7.1 παραθέτουμε ενδεικτικά την παραγωγή 100 χρηστών, πάνω στο δίκτυο 'oldenburg' με την χρήση του προγράμματος παραγωγής χώρο-χρονικών δεδομένων του Brinkhoff.

Όσον αφορά τα δεδομένα που αποθηκεύονται στη βάση δεδομένων, για τους κινούμενους χρήστες έχουν παρόμοια μορφή με αυτή του σχήματος 6.3, και τα PHLs των χρηστών έχουν παρόμοια μορφή με αυτή του σχήματος 6.5. Η ακριβής απεικόνιση των δεδομένων που έχουν αποθηκευτεί στην βάση καθώς και των PHLs των χρηστών δεν είναι δυνατή καθώς τα πειράματα έχουν

διεξαχθεί για 100, 500, και 1000 χρήστες και άρα το σύνολο δεδομένων είναι πολύ μεγάλο. Επίσης στα πλαίσια των πειραμάτων θεωρήσαμε τυχαία, ότι ο χρήστης με ID=11 κάνει τις αιτήσεις οι οποίες είναι στο σύνολό τους 100. Έτσι σε κάθε περίπτωση θα παρουσιαστεί το ποσοστό επιτυχίας της στρατηγικής εκφρασμένο επί τοις εκατό.

Πρέπει επίσης να πούμε ότι τα πειράματα πραγματοποιήθηκαν για τις ακόλουθες τιμές του **'κ': {2,5,10,20,50,70,100}**. Αυτές οι τιμές καθορίζουν το επίπεδο ανωνυμίας που επιθυμούμε να επιτευχθεί κάθε φορά για τον αιτούντα. Τέλος, οι χώρο-χρονικοί περιορισμοί που δόθηκαν στην είσοδο ήταν οι εξής: Για τον χώρο δόθηκαν οι τιμές: **{2.000,4.000,8.000,15.000}**, οι οποίες είναι εκφρασμένες σε μέτρα, και για τον χρόνο οι τιμές**: {2,3,5,7},** οι οποίες είναι εκφρασμένες σε χρονικές μονάδες. Όλα αυτά τα δεδομένα, τα οποία χρησιμοποιηθήκαν για την διεξαγωγή των πειραμάτων, εικονίζονται στον πίνακα του σχήματος 7.2.



Σχήμα 7.1: Παραγωγή 100 χρηστών στο δίκτυο 'oldenburg'.

ΠΕΔΙΟ	***ΤΙΜΗ***
Δίκτυο εισόδου	'oldenburg '
Χωρικά όρια δικτύου	Τετράγωνη περιοχή:(κάτω αριστερή κορυφή**(0,0)**, πάνω δεξιά κορυφή **(30.000,30.000)**)
Χρονικά όρια	[0,23] , σε χρονικές μονάδες
Ταχύτητα κινουμένων χρηστών	50=medium
Αριθμός χρηστών	{100,500,1000}
Αιτών	Χρήστης με ID=11
Επίπεδο ανωνυμίας (τιμές του κ)	{2,5,10,20,50,70,100}
Χωρικοί περιορισμοί	{2000,4000,8000,15000} , σε μέτρα
Χρονικοί περιορισμοί	{2,3,5,7} , σε χρονικές μονάδες

Σχήμα 7.2: Δεδομένα πειραμάτων.

Στο σχήμα 7.3, εικονίζονται τα αποτελέσματα που έχουν προκύψει, με την εφαρμογή της στρατηγικής για 100 χρήστες. Σε αυτή την περίπτωση που έχουμε 100 χρήστες, επιλέχθηκε να εφαρμοστεί η τεχνική που υλοποιήθηκε, για τις εξής τιμές του **κ: {2,5,10}.** Το ότι δεν επιλεχθήκαν μεγαλύτερες τιμές για την μεταβλητή '**κ**', οφείλεται στο γεγονός ότι για 100 χρήστες το επίπεδο ανωνυμίας που μπορεί να διασφαλισθεί (με βάση τους **πιο αυστηρούς** χώρο-χρονικούς περιορισμούς) είναι 'περίπου' μέχρι 10-ανωνυμία. Παρατηρώντας το σχήμα 7.3, διαπιστώνουμε ότι για τους πιο αυστηρούς χώρο-χρονικούς περιορισμούς (2.000 μέτρα και 2 χρονικές μονάδες) το ποσοστό επιτυχίας είναι 6%. Αν η τιμή του '**κ**' πάρει μεγαλύτερη τιμή, για παράδειγμα αν **κ**=20, τότε για τους αυστηρότερους χώρο χρονικούς περιορισμούς, δεν θα είναι δυνατόν να επιτευχθεί κ-ανωνυμία. Για αυτό τον λόγο και επιλέξαμε οι τιμές του **κ** να φτάνουν μέχρι 10.

Φυσικά θα μπορούσε να διασφαλίζεται μεγαλύτερο επίπεδο ανωνυμίας, αλλά θα έπρεπε να διευρύνουμε περισσότερο τους χωρικούς και χρονικούς περιορισμούς. Κάτι τέτοιο όμως δεν έχει νόημα, δεδομένου ότι στόχος μας είναι να αξιολογηθεί η στρατηγική πάνω σε πραγματικά δεδομένα και όχι να γίνεται γενίκευση του χώρου και του χρόνου σε τόσο μεγάλο βαθμό, που να μην είναι τα αποτελέσματα ρεαλιστικά. Γενικότερα οι χωρικοί και χρονικοί περιορισμοί, καθώς και η τιμή του '**κ**', σε κάθε μια περίπτωση(100,500,1000 χρήστες), έχουν επιλεγεί με τρόπο τέτοιο ώστε να επιδεικνύεται το επίπεδο ανωνυμίας που παρέχεται σε κάθε μια περίπτωση, αλλά παράλληλα, κάθε φορά να φαίνεται και μέχρι ποιον βαθμό γενίκευσης μπορούμε να φτάσουμε. Με άλλα λόγια οι παράμετροι έχουν οριστεί με τέτοιο τρόπο έτσι ώστε να φαίνεται πότε η στρατηγική που υλοποιήσαμε, φτάνει στα όρια της και δεν είναι δυνατόν να επιτύχει κ-ανωνυμία.

Πιο συγκεκριμένα από το σχήμα 7.3 παρατηρούμε ότι για κ=2 όσο αυξάνουμε τους χώρο-χρονικούς περιορισμούς τόσο αυξάνεται και το ποσοστό επιτυχίας. Ειδικότερα όταν γενικεύουμε τον χώρο 15.000 μέτρα, βλέπουμε ότι έχουμε 100% επιτυχία. Αυτό είναι απόλυτα λογικό αφού έχοντα 100 χρήστες και γενικεύοντας 15.000 μέτρα, (δηλαδή η γενικευμένη περιοχή καλύπτει το 50% όλης της χωρικής περιοχής που έχει οριστεί), είναι πολύ εύκολα να βρεθούν 2 γείτονες.
Από την άλλη πλευρά παρατηρούμε ότι όταν κ=10 και γενικεύουμε τον χώρο 2000 μέτρα και τον χρόνο 2 χρονικές μονάδες, η τεχνική έχει μόλις 6% επιτυχία. Αυτό αν αναλογιστούμε ότι έχουμε 100 χρήστες, είναι κάπως δύσκολο να κατανοήσουμε γιατί έγινε. Οι λόγοι για τους οποίους συνέβη αυτό είναι δύο. Ο πρώτος είναι ότι τα όρια γενίκευσης του χώρου αλλά και του χρόνου είναι σχετικά μικρά(2000 μέτρα και 2 χρονικές μονάδες αντίστοιχα). Αυτό και μόνο όμως και πάλι δεν εξηγεί αυτό το αποτέλεσμα, διότι από την στιγμή που έχουμε 100 χρήστες, θα αναμέναμε να υπάρχει η δυνατότητα να βρεθούν 10 γείτονες, και να επιτευχθεί 10-ανωνυμία. Ο κύριος λόγος που η τεχνική δεν κατάφερε να επιτύχει 10-ανωνυμία(πάνω από 6% των περιπτώσεων), αν και υπάρχουν 100 χρήστες στο σύστημα, κρύβεται στην παραμετροποίηση που έχει γίνει στο πρόγραμμα (Generator) του Brinkhoff. Πιο συγκεκριμένα από τον πίνακα που εικονίζεται στο σχήμα 7.2 βλέπουμε ότι έχει οριστεί η ταχύτητα των κινούμενων χρηστών να έχει την τιμή 50(δηλαδή medium). Αφού λοιπόν οι χρήστες κινούνται με σχετικά

μικρή ταχύτητα, δεν έχουν την δυνατότητα να καλύψουν μεγάλες αποστάσεις. Αυτό σημαίνει ότι τα στιγμιότυπά τους δεν είναι διεσπαρμένα σε όλη την έκταση του χάρτη, αλλά αντίθετα περιορίζονται σε μια σχετικά μικρή περιοχή. Για αυτό τον λόγο είναι αρκετά δύσκολο να βρεθούν 10 γείτονες.

Αν είχε οριστεί ότι οι χρήστες κινούνται με μεγάλη ταχύτητα(πχ 10=fast), τότε τα στιγμιότυπα των χρηστών θα ήταν και πάλι ίδια σε αριθμό, αλλά θα ήταν το ένα πιο μακριά από το άλλο. Έτσι ο κάθε χρήστης θα είχε στιγμιότυπα σε μια πολύ ευρύτερη περιοχή και θα ήταν πιο εύκολο να επιτευχθεί 10-ανωνυμία, με τους ίδιους χωρικούς και χρονικούς περιορισμούς. Όπως προαναφέραμε, στόχος μας είναι να εφαρμόσουμε την προτεινόμενη τεχνική **και σε 'οριακά σημεία'**, έτσι ώστε να κατανοήσουμε πότε δεν είναι δυνατόν να διασφαλισθεί ανωνυμία των χρηστών.

Αριθμός χρηστών	Επίπεδο ανωνυμίας (Τιμή κ)	Χωρικοί περιορισμοί (σε μέτρα)	Χρονικοί περιορισμοί (σε χρονικες μονάδες)	Ποσοστό επιτυχίας (%)
100	2	2000	2	47
100	2	4000	3	64
100	2	8000	5	94
100	2	15000	7	100
100	5	2000	2	27
100	5	4000	3	57
100	5	8000	5	88
100	5	15000	7	100
100	10	2000	2	6
100	10	4000	3	50
100	10	8000	5	70
100	10	15000	7	100

Σχήμα 7.3: Αποτελέσματα για 100 χρήστες.

Στο σχήμα 7.4, εικονίζονται τα αποτελέσματα που έχουν προκύψει, με την εφαρμογή της στρατηγικής για 500 χρήστες. Σε αυτή την περίπτωση θεωρούμε ότι οι τιμές του **'κ'** είναι**: {10,20,50}**. Πάλι παρατηρούμε ότι το μεγαλύτερο ποσοστό επιτυχιών (100%), επιτυγχάνεται όταν έχουμε τους μεγαλύτερους χώρο-χρονικούς περιορισμούς. Επίσης παρατηρούμε ότι για κ=50 έχουμε ποσοστό επιτυχίας 5%, αυτό όπως προαναφέραμε οφείλεται στην σχετικά μικρή ταχύτητα μετακίνησης των χρηστών, καθώς και στους αυστηρούς χώρο-χρονικούς περιορισμούς που έχουμε επιβάλει, προκειμένου να γίνει φανερό πότε η τεχνική φτάνει στα όρια της.

Αριθμός χρηστών	Επίπεδο ανωνυμίας (Τιμή κ)	Χωρικοί περιορισμοί (σε μέτρα)	Χρονικοί περιορισμοί (σε χρονικές μονάδες)	Ποσοστό επιτυχίας (%)
500	10	2000	2	34
500	10	4000	3	63
500	10	8000	5	94
500	10	15000	7	100
500	20	2000	2	25
500	20	4000	3	58
500	20	8000	5	87
500	20	15000	7	100
500	50	2000	2	5
500	50	4000	3	39
500	50	8000	5	67
500	50	15000	7	100

Σχήμα 7.4: Αποτελέσματα για 500 χρήστες.

Τέλος στο σχήμα 7.5, εικονίζονται τα αποτελέσματα που έχουν προκύψει, με την εφαρμογή της στρατηγικής για 1000 χρήστες. Η ερμηνεία των αποτελεσμάτων για την εφαρμογή της στρατηγικής σε 1000 χρήστες, είναι παρόμοια με τις περιπτώσεις που είχαμε 100 και 500 χρήστες. Και πάλι με 'πράσινο φόντο', εικονίζονται το μέγιστο ποσοστό επιτυχίας της στρατηγικής. Ενώ με 'κόκκινο' εικονίζεται το ελάχιστο ποσοστό επιτυχίας.

Αριθμός χρηστών	Επίπεδο ανωνυμίας (Τιμή κ)	Χωρικοί περιορισμοί (σε μέτρα)	Χρονικοί περιορισμοί (σε χρονικές μονάδες)	Ποσοστό επιτυχίας (%)
1000	50	2000	2	29
1000	50	4000	3	58
1000	50	8000	5	82
1000	50	15000	7	100
1000	70	2000	2	18
1000	70	4000	3	51
1000	70	8000	5	71
1000	70	15000	7	98
1000	100	2000	2	10
1000	100	4000	3	44
1000	100	8000	5	63
1000	100	15000	7	97

Σχήμα 7.5: Αποτελέσματα για 1000 χρήστες

8. Επίλογος

Συνοψίζοντας τα όσα έχουν παρουσιαστεί στα πλαίσια αυτής της διπλωματικής εργασίας, έχουμε τα εξής: Πρώτα απ' όλα παρουσιάστηκε μια περιγραφή των χωρικών δεδομένων, καθώς και πως τέτοιου είδους δεδομένα αποθηκεύονται σε μια βάση χωρικών δεδομένων. Ύστερα παρουσιάστηκε μια τεχνική κατασκευής δικτύων, η οποία βασίζεται στις ιδιότητες των χωρικών δεδομένων. Στην συνέχεια έγινε μία περιγραφή των χώρο-χρονικών δεδομένων, καθώς και των τρόπων με τους οποίους μπορούν να παραχθούν και να αποθηκευτούν τέτοιου είδους δεδομένα, σε μία χώρο-χρονική βάση δεδομένων. Τέλος προτάθηκε μια τεχνική, με βάση την οποία μπορεί να διασφαλισθεί η ιδιωτικότητα των χώρο-χρονικών δεδομένων, τα οποία αποστέλλονται όταν κάποιοι χρήστες κάνουν μια αίτηση προς έναν παροχέα υπηρεσιών, προκειμένου να ζητήσουν μια υπηρεσία. Πιο συγκεκριμένα η τεχνική αυτή έχει σαν στόχο, την επίτευξη κ-ανωνυμίας για κάθε χρήστη που κάνει μια αίτηση προς έναν παροχέα υπηρεσιών. Δηλαδή η τεχνική αυτή πρέπει να διασφαλίσει ότι από την περιοχή που ο αιτών έκανε την αίτηση, έχουν περάσει τουλάχιστο κ-1 διαφορετικοί χρήστες. Έτσι ο παροχέας υπηρεσιών δε θα έχει την δυνατότητα να ξεχωρίσει ποιος από τους κ χρήστες έκανε την συγκεκριμένη αίτηση.

Το πρώτο βήμα της τεχνικής που παρουσιάστηκε είναι η εφαρμογή του αλγορίθμου γενίκευσης και το δεύτερο βήμα είναι η εφαρμογή του αλγορίθμου αποσύνδεσης, ο οποίος εκτελείται αν και μόνο αν ο αλγόριθμος γενίκευσης έχει αποτύχει. Ο αλγόριθμος γενίκευσης, γενικεύει τον χρόνο και τον χώρο από όπου έγινε η αίτηση, μέχρι να βρει κ-1 γείτονες. Αυτή η γενίκευση δε γίνεται επ' αόριστο αλλά σταματά με βάση τους χώρο-χρονικούς περιορισμούς που έχουν ορισθεί. Αν μετά το πέρας του αλγορίθμου γενίκευσης δεν έχουν βρεθεί κ-1 γείτονες τότε εφαρμόζεται ο αλγόριθμος αποσύνδεσης. Αυτός ο αλγόριθμος έχει ως στόχο να αλλάξει το αναγνωριστικό του αιτούντα και έτσι ο παροχέας υπηρεσιών να μην έχει την δυνατότητα να διασυνδέσει παλιές αιτήσεις του αιτούντα με μελλοντικές του αιτήσεις. Με αυτόν τον τρόπο διασφαλίζεται πλήρως η ιδιωτικότητα του

αιτούντα, καθώς ο χρήστης με το νέο αναγνωριστικό αντιμετωπίζεται σαν ένας εντελώς καινούριος χρήστης.

Τέλος, παρουσιάστηκαν πειραματικά δεδομένα τα οποία παρουσιάζουν το επίπεδο ανωνυμίας που παρέχει η τεχνική που προτείναμε. Ειδικότερα τα πειραματικά δεδομένα επιβεβαιώνουν ότι η τεχνική που υλοποιήθηκε, έχει την επιθυμητή λειτουργικότητα, και παρέχει τα επιθυμητά επίπεδα ασφάλειας της ιδιωτικότητας των χρηστών, αν ικανοποιούνται κάποιες βασικές προϋποθέσεις. Στην περίπτωση που αυτές οι προϋποθέσεις δεν ικανοποιούνται, για παράδειγμα αν οι χώρο-χρονικοί περιορισμοί της γενίκευσης είναι πολύ μικροί ή αν υπάρχει μικρός αριθμός χρηστών, η στρατηγική αποτυγχάνει και οφείλει να ενημερώσει τον χρήστη ότι υπό αυτές τις συνθήκες, η ιδιωτικότητα του είναι σε κίνδυνο.

9. Βιβλιογραφία

[1] L. Sweeney, P. Samarati. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. IEEE Security and Privacy, 1998.

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

[3] L. Sweeney. Achieving *k*-anonymity privacy protection using generalization and suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems,* 10 (5), 2002; 571-588.

[4] P. Samarati. Protecting respondents' identities in microdata release. IEEE Transactions on Knowledge and Data Engineering. 13(6), November/December 2001.

[5] Agrawal and R. Srikant. Privacy-preserving data mining. In *Proceedings of the ACM SIGMOD. International Conference on Management of Data* May 2000.

[6] Privacy Beyond *k*-Anonymity Ashwin Machanavajjhala Daniel Kifer. Department of Computer Science, Cornell University. December 2001

[7] M.Gruteser and D.Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. First International Conference on Mobile Systems, Applications, and Services (MobiSys'03). pages 31-42. May 2003.

[8] B.Gedik and L.Liu. A Customizable k-Anonymity Model for Protecting Location Privacy. The 25 International Conference on Distributed Computing Systems. IEEE ICDCS 2005.

[9] C. Bettini, X.S. Wang, S. Jajodia. Protecting Privacy against Location-Based Personal Identification. In Proceedings of 2nd VLDB Workshop on Secure Data Management (SDM). 2005.

[10] C. Bettini, S. Jajodia, X.S. Wang. Time Granularities in Databases, Data Mining and Temporal Reasoning. Springer. 2000.

[11] SUMO - Simulation of Urban Mobility - User Documentation
Daniel Krajzewicz & Christian Rossel.

[12] *Oporto Generator* Release notes Version 1.1 for Ms-DOS based systems José Moreira (Universidade Portucalense) & Jean-Marc Saglio (ENST – Paris) September, 1999.

[13] Thomas Brinkhoff. A Framework for Generating Network-Based Moving Objects. Institute of Applied Photogrammetry and Geoinformatics (IAPG), Fachhochschule Oldenburg/Ostfriesland/Wilhelmshaven (University of Applied Sciences), Published in: GeoInformatica, Vol. 6, No. 2, 2002

[14] Oracle Spatial User's Guide and Reference 10g Rel.2 (10.2)

[15] Oracle Spatial Topology and Network Data Models 10g Rel.2

[16] Oracle® Database Express Edition 2 Day Plus Locator Developer Guide 10g Release 2 (10.2). B28004-01. February 2006.Provides a quick start to storing and querying spatial(location-based) data, using the Oracle Locator feature of Oracle Database Express Edition.

[17] Oracle Database Express Edition 2 Day Plus Java Developer Guide, 10g Release 2 (10.2) B25320-01 Copyright © 2006, Oracle.

[18] JDBC TM RowSet Implementations Tutorial.Sun Microsystems Inc. 4150 Network Circle Santa Clara, CA 95054 USA.Revision 1.0