ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Προστασία Ιδιωτικότητας Θέσης σε IoT περιβάλλοντα βάσει Τεχνητής Νοημοσύνης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΚΩΝΣΤΑΝΤΙΝΟΣ ΔΗΜΗΤΡΙΟΥ

Επιβλέπουσα: Ιωάννα Ρουσσάκη
Επικ. Καθηγήτρια Ε.Μ.Π.

Αθήνα 2019

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Προστασία Ιδιωτικότητας Θέσης σε IoT περιβάλλοντα βάσει Τεχνητής Νοημοσύνης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΚΩΝΣΤΑΝΤΙΝΟΣ ΔΗΜΗΤΡΙΟΥ

Επιβλέπουσα: Ιωάννα Ρουσσάκη
Επικ.Καθηγήτρια Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 13 Μαρτίου 2019.

.....	.....	.....
Ιωάννα Ρουσσάκη	Μιλτιάδης Αναγνώστου	Συμεών Παπαβασιλείου
Επικ.Καθηγήτρια Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα 2019

.....

ΔΗΜΗΤΡΙΟΥ ΚΩΝΣΤΑΝΤΙΝΟΣ
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών

Περίληψη

Σκοπός της διπλωματικής είναι η ανάπτυξη αλγορίθμων σε data-centric sensor networks (DCSN), που αποτελούν υποκατηγορία των IoT, με σκοπό την ασφαλή μετάδοση δεδομένων τα οποία σχετίζονται με την παρακολούθηση της τοποθεσίας κινητών αντικειμένων εντός δικτύου. Ο στόχος είναι τα ασύρματα δίκτυα αυτού του τύπου να γίνουν περισσότερο ασφαλή και λιγότερο ευάλωτα απέναντι σε εισβολείς που βασικό σκοπό έχουν την παραβίαση ευαίσθητων δεδομένων που αφορούν τα κινητά αντικείμενα.

Πιο συγκεκριμένα η διπλωματική βασίζεται σε δύο κεντρικούς άξονες. Ο πρώτος είναι η στατική συσταδοποίηση, η αλλιώς clustering, των δεδομένων χωρίς να μας ενδιαφέρει η τροχιά του κινητού αντικείμενου. Εφαρμόζονται διαφορετικοί αλγόριθμοι συσταδοποίησης ακόμα και συνδυασμός τους προκειμένου να έχουμε την όσο τον δυνατόν καλύτερη στατική συσταδοποίηση πριν καν ακόμα εισέλθει κινητό αντικείμενο εντός δικτύου. Ο δεύτερος άξονας είναι η ανάπτυξη αλγορίθμων για δυναμική συσταδοποίηση των δεδομένων που εκτελείται τη στιγμή που το κινητό βρίσκεται σε κίνηση. Η αξιολόγηση των αλγορίθμων τόσο για την στατική όσο και την δυναμική ομαδοποίηση καθώς και γενικότερα όλα τα πειράματα πραγματοποιήθηκαν πάνω σε φυσικές τοπολογίες και φυσικά συστήματα που αντλήθηκαν από πλατφόρμες που προσφέρουν τοπολογίες και δυνατότητες δικτύων internet of things (fit-iot lab).

Τα δίκτυα internet of things αποτελούν ένα καινοτόμο ερευνητικό πεδίο στο οποίο τα τελευταία χρόνια έχει σημειωθεί σημαντική τεχνολογική εξέλιξη και έχουν μεγάλη απήχηση. Τα δίκτυα αυτού του τύπου μπαίνουν ολοένα και περισσότερο στην καθημερινή μας ζωή τραβώντας ολοένα και περισσότερο το ενδιαφέρον της κοινότητας ατόμων που ασχολούνται με διαδικτυακές επιθέσεις, οι λεγόμενοι χάκερς. Πολλές είναι οι τεχνικές που έχουν προταθεί για την προστασία των δικτύων iot όμως όλο και περισσότερες και πιο εξελιγμένες είναι οι επιθέσεις που δέχονται. Η μέθοδος και οι τεχνικές που παρουσιάζονται στην παρούσα διπλωματική δεν έχουν να κάνουν με την παρεμπόδιση των εισβολέων και την αποφυγή της επίθεσης αλλά με το πόσο ασφαλή είναι τα ευαίσθητα δεδομένα των χρηστών του δικτύου μετά την παραβίαση των κόμβων του. Προτείνεται έτσι μια νέα μέθοδος και ενθαρρύνεται η περαιτέρω έρευνα προς αυτήν την κατεύθυνση.

Λέξεις-Κλειδιά: Διαδίκτυο των Αντικειμένων (IoT), Δεδομενοκεντρικά Δίκτυα Αισθητήρων (DCSN), ασύρματα δίκτυα, Συσταδοποίηση, μηχανική μάθηση, τεχνητή νοημοσύνη

Abstract

The purpose of this diploma thesis constitutes the creation of an innovative algorithm in DCSN (Data Centric Sensor Networks), subcategory of IoT environments, in order to ensure the secure data dissemination among its nodes. Special emphasis is given to those networks that monitor and extract the location of mobile objects. The objective is, this type of networks, to become more secure and less vulnerable to intruders whose aim is to breach sensitive data related to those mobile objects.

In specific, this diploma thesis focuses on two different directions. The former is the static analysis of the network without considering the trajectory of mobile object. Miscellaneous algorithms, are implemented or even combination of them, to deduce the best possible static clustering before the mobile object entering into the network. The latter is the design and implementation of an innovative algorithm so as to achieve the best possible dynamic clustering. This algorithm is running when the mobile object is within the range of the network. The assessment for both static and dynamic clustering was conducted on topologies and physical devices of Fit-IoT lab which provides such IoT environments.

IoT constitutes a pioneering scientific field in which humanity has achieved many breakthroughs the latest decades and attracts constantly the attention of scientific and industrial community. IoT is part of our everyday life which has led the intruders, or hackers, to design new cyber attacks in order to compromise such networks. Many techniques have been proposed the latest years to protect IoT devices yet the cyber attacks are constantly becoming more and more complex. The purpose of the methods and the techniques that are introduced in this diploma thesis, does not constitute the prevention of the attack but how secure are the sensitive data after the nodes have been compromised. An innovative technique is introduced and opens up the way for further research concerning the privacy of such networks.

Keywords: Internet of things (IoT), Data-Centric Sensor networks (DCSN), wireless networks, clustering, machine learning, artificial intelligence

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερα την Επικ.Καθηγήτρια κ. Ιωάννα Ρουσσάκη που με την ανάθεση της συγκεκριμένης διπλωματικής μου έδωσε την ευκαιρία να ασχοληθώ με ενδιαφέροντες τομείς όπως το Διαδίκτυο των Αντικειμένων και τη μηχανική μάθηση. Επιπλέον, με ενέπνευσε να ασχοληθώ επαγγελματικά και επιστημονικά με τους συγκεκριμένους τομείς και να προσπαθώ συνεχώς να εξελίσσομαι και να αναπτύσσω τις γνώσεις μου. Επιπλέον, θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Μιλτιάδη Αναγνώστου και τον Καθηγητή κ. Συμεών Παπαβασιλείου για την τιμή που μου έκαναν να συμμετάσχουν στην τριμελή εξεταστική επιτροπή της εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω τον ερευνητή κ. Νικόλαο Καλατζή για τις πολύτιμες συμβουλές, υποδείξεις και σχόλια του καθ' όλη την διάρκεια διεξαγωγής της συγκεκριμένης διπλωματικής. Η βοήθεια του ήταν πολύτιμη για την ολοκλήρωση της εργασίας.

Αφιερώνω τη διπλωματική εργασία στους γονείς μου, Μενέλαο Δημητρίου και Ροδοθέα Βαζιργιαντζίκη καθώς και την αδερφή μου, Μαρία Δημητρίου, για την υποστήριξη και την αγάπη τους όλα αυτά τα χρόνο. Επίσης, θα ήθελα να ευχαριστήσω τους φίλους μου για την συνεχή υποστήριξη και όλα αυτά τα υπέροχα φοιτητικά χρόνια.

Πίνακας περιεχομένων

ΚΕΦΑΛΑΙΟ 1 -Εισαγωγή

1.1 Διαδίκτυο των Αντικειμένων (IoT)

Το Διαδίκτυο των Αντικειμένων (Internet of Things - IoT) [InKy15] αποτελεί ένα όραμα των ανθρώπων της βιομηχανίας να δημιουργήσουν ένα παγκόσμιο δίκτυο στο οποίο μηχανές και συσκευές θα είχαν τη δυνατότητα να επικοινωνήσουν μεταξύ τους. Θεωρείται ένας πολύ σημαντικός τεχνολογικός κλάδος ο οποίος εξελίσσεται ραγδαία τα τελευταία χρόνια και προορίζεται να χρησιμοποιηθεί σε πληθώρα εφαρμογών όπως εφαρμογές εξυπηρέτησης πελατών, ανάλυσης επιχειρησιακών δεδομένων και ούτω καθεξής. Αναμένεται το πλήθος συσκευών IoT να αυξηθεί ταχέως τα επόμενα χρόνια και από 0.9 δις που υπήρχαν το 2009 να ξεπεράσουμε τις 26 δις μέχρι το πέρας του 2020. Πολλές εταιρείες τα επόμενα χρόνια θα επενδύσουν σε δίκτυα IoT προκειμένου να μειώσουν το κόστος κατανομής, να βελτιώσουν την παρακολούθηση των υλικών και πρώτων υλών τους και γενικότερα να διευκολύνουν τη γραμμή παραγωγής των προϊόντων τους. Ήδη έχουν αρχίσει να χρησιμοποιούνται από κάποιες εταιρείες προκειμένου να αυξήσουν τα έσοδά τους. Διάφορες συσκευές χειρός, POS συσκευές, IP κάμερες είναι μόνο μερικά από τα παραδείγματα που κυκλοφορούν στην αφορά και έχουν βελτιώσει την εμπειρία χρήσης καθώς και γενικότερα το βιοτικό επίπεδο ζωής.

Παρακάτω παρουσιάζονται 5 αναγκαίες IoT τεχνολογίες που χρησιμοποιούνται σε προϊόντα και υπηρεσίες. Οι τεχνολογίες αυτές έχουν ως εξής:

1. **Radio frequency identification (RFID)**: Επιτρέπουν την αυτόματη αναγνώριση και συλλογή δεδομένων χρησιμοποιώντας ραδιοκύματα, ετικέτες (tags) και έναν αναγνώστη, μια συσκευή που συλλέγει δεδομένα από τις ετικέτες. Τα tags περιέχουν πληροφορία κωδικοποιημένη σε μορφή Electronic Product Code (EPC) που είναι ένας παγκόσμιος RFID κώδικας αναγνώρισης. Υπάρχουν 3 ειδών tags, τα παθητικά RFID tags (passive RFID), τα ημιπαθητικά RFID tags (semi-passive RFID) και τα ενεργητικά (active RFID). Στην πρώτη κατηγορία δεν υπάρχει μπαταρία παρά μόνο μεταφέρεται ενέργεια από τον αναγνώστη στο tag για να ενεργοποιηθεί. Στην δεύτερη κατηγορία χρησιμοποιείται μπαταρία μόνο για το microchip και για την επικοινωνία είναι και πάλι αναγκαία η ενέργεια από τον αναγνώστη. Στην τρίτη κατηγορία, υπάρχει μπαταρία και μπορούν οι ετικέτες μόνες τους να διεγείρουν την επικοινωνία με τον αναγνώστη.

2. **Wireless sensor networks (WSN)**: Τα ασύρματα δίκτυα ανίχνευσης αποτελούνται από συσκευές που είναι εξοπλισμένες με αισθητήρες ανίχνευσης και είναι διάσπαρτα κατανεμημένες στο χώρο. Χρησιμοποιούνται για να παρακολουθούν διάφορες περιβαλλοντικές συνθήκες και συνεργάζονται πολλές φορές με τα RFID για καλύτερη παρακολούθηση τοποθεσίας, κίνησης και θερμοκρασίας. Επιτρέπουν την παράταξη μια μεγάλης πληθώρας τοπολογιών και την επικοινωνία ανάμεσα από πολλούς κόμβους.

3. **Middleware**: Το middleware αποτελεί ένα επίπεδο λογισμικού που παρεμβάλλεται σε διάφορες εφαρμογές και διευκολύνει συγκεκριμένες λειτουργίες όπως η επικοινωνία και η είσοδος/έξοδος δεδομένων. Το μεγάλο του πλεονέκτημα είναι ότι αποκρύπτει τις λεπτομέρειες υλοποίησης των διάφορων τεχνολογιών και υπηρεσιών. Η χρησιμοποίηση του διευκόλυνε την ανάπτυξη υπηρεσιών σε κατανεμημένα περιβάλλοντα. Ειδικότερα σε δίκτυα IoT που αποτελούνται από μια πληθώρα ετερογενών συσκευών, οι οποίες δρουν με

κατανεμημένο και παράλληλο τρόπο, η απλοποίηση διάφορων υπηρεσιών και εφαρμογών ήταν αναγκαία και επιτεύχθηκε με την δημιουργία του middleware.

4. **Cloud computing**: Το cloud computing αποτελεί ένα μοντέλο το οποίο επιτρέπει την πρόσβαση χρηστών σε διαμοιραζόμενα δεδομένα όπως υπηρεσίες, λογισμικό, δίκτυα και servers. Η ραγδαία αύξηση των IoT συσκευών, που είναι συνδεδεμένες στο Διαδίκτυο, οδήγησε σε έναν τεράστιο όγκο δεδομένων που είναι δύσκολο να αποθηκευθεί και διαχειριστεί. Πολλές νέες IoT εφαρμογές έχουν ανάγκη από γρήγορη ταχύτητα επεξεργασίας δεδομένων και μαζική αποθήκευσή τους. Γι αυτόν τον λόγο το cloud computing αποτέλεσε το ιδανικό μοντέλο το οποίο μπόρεσε να ικανοποιήσει τις απαιτήσεις τέτοιου είδους εφαρμογών.

5. **IoT application software**: Οι IoT υπηρεσίες και τεχνολογίες, τα τελευταία χρόνια, χρησιμοποιούνται σε πληθώρα εφαρμογών. Το μεγάλο τους πλεονέκτημα είναι ότι επιτρέπουν την απευθείας αλληλεπίδραση ανθρώπου-μηχανής αλλά και μηχανής-μηχανής με αρκετά αξιόπιστο τρόπο. Επιπλέον, διασφαλίζουν την λήψη μηνυμάτων σε πραγματικό χρόνο, ειδικότερα σε εφαρμογές που την απαιτούν όπως την παρακολούθηση των αγαθών σε εφαρμογές logistics και μεταφορών. Οι νέες IoT εφαρμογές που δημιουργούνται τα τελευταία χρόνια εμπεριέχουν αναπαράσταση δεδομένων με τέτοιο τρόπο ώστε να είναι φιλική προς το χρήστη και επιτρέπουν την αλληλεπίδρασή του ανθρώπου με αυτές.

1.2 Συνεισφορά

- Έγινε εφαρμογή ήδη υπαρχόντων αλγορίθμων καθώς και ανάπτυξη νέων με σκοπό την όσο καλύτερη στατική και δυναμική συσταδοποίηση δεδομένων. Ουσιαστικά εισάγουμε την έννοια της μηχανικής μάθησης (machine learning) και ειδικότερα του clustering στον χώρο των δικτύων IoT. Σκοπός είναι να διαχωρίσουμε τους κόμβους του δικτύου σε ομάδες με κριτήριο και γνώμονα την ασφάλεια των ευαίσθητων δεδομένων που ανταλλάσσονται μεταξύ των κόμβων που κάνουν ανίχνευση και εκείνων που αποθηκεύουν τα δεδομένα. Έγινε ανάλυση των αλγορίθμων και εκπαίδευση τους πάνω στις διάφορες τοπολογίες προκειμένου να επιλέξουμε εκείνους που ταιριάζουν καλύτερα στα συγκεκριμένα δίκτυα. Επιπλέον προτείνεται ένας νέος αλγόριθμος για την περαιτέρω βελτιστοποίηση της ασφάλειας του δικτύου.

- Υλοποιήθηκαν και προγραμματίστηκαν απλές συσκευές που χρησιμοποιούνται σαν κόμβοι ανίχνευσης (sensor nodes). Οι κόμβοι αυτοί είναι σε θέση να γνωρίζουν μέσω κάποιον αλγορίθμων ασύρματης τοπολογίας τους γειτονικούς τους κόμβους στο δίκτυο. Έχουν συγκεκριμένη ακτίνα ανίχνευσης και συγκεκριμένη ακτίνα επικοινωνίας με τους υπόλοιπους κόμβους. Είναι σε θέση να ανιχνεύουν αντικείμενα που περνούν σε απόσταση μικρότερη από την ακτίνα ανίχνευσής τους.

- Υλοποιήθηκαν και προγραμματίστηκαν συσκευές που έχουν την δυνατότητα να αποθηκεύουν τα δεδομένα που λαμβάνουν από τις αντίστοιχες συσκευές ανίχνευσης (storage nodes). Επιπλέον έχουν την ικανότητα να επεξεργάζονται και να συνδυάζουν τα δεδομένα που έρχονται από τους κόμβους ανίχνευσης που εμπίπτουν στην ομάδα που είναι αφιερωμένη στην εκάστοτε συσκευή. Τέλος, σχεδιάστηκε μία συσκευή με παρόμοια χαρακτηριστικά που όμως λαμβάνει μηνύματα από όλους τους κόμβους του δικτύου και

έτσι έχει γνώση όλου του δικτύου και πιο ακριβή εντοπισμό του κινητού αντικειμένου (sink node).

- Τέλος έγινε η ανάλυση των αλγορίθμων σε διαφορετικές τοπολογίες οι οποίες επιδεικνύονται στην παρούσα διπλωματική καθώς και σύγκριση της απόδοσης μεταξύ των αλγορίθμων. Μέσα από την αξιολόγηση διαφόρων γραφημάτων και μετρικών γίνεται σαφές το πόσο βελτιώνεται το επίπεδο της ασφάλειας καθώς και η ελαχιστοποίηση των ευαίσθητων δεδομένων που μπορούν να υποκλαπούν από επίδοξους εισβολείς.

1.3 Οργάνωση Κειμένου

Η οργάνωση του κειμένου γίνεται ως εξής. Στο κεφάλαιο 2 γίνεται μία εισαγωγή στις βασικές αρχές της συσταδοποίησης καθώς και διάφορων αλγορίθμων του αλλά και μια εισαγωγή στις θεμελιώδη αξίες του constrained clustering, που είναι υποκατηγορία του semi supervised clustering, καθώς και ανάλυση ορισμένων βασικών αλγορίθμων που ανήκουν στην οικογένεια αυτή. Στο κεφάλαιο 3 θα γίνει μια παρουσίαση της πιο σύγχρονης σχετικής βιβλιογραφίας που είναι σχετική με το συγκεκριμένο θέμα καθώς και επεξήγηση του θεωρητικού υποβάθρου του προβλήματος. Στο κεφάλαιο 4 αναλύεται το βασικό πρόβλημα που μας οδήγησε στο να διεξάγουμε αυτή τη μελέτη. Στο κεφάλαιο 5 παρουσιάζεται η προτεινόμενη μέθοδος τόσο για την στατική ομαδοποίηση των κόμβων όσο και για την δυναμική. Στο κεφάλαιο 6 επιδεικνύονται οι βασικές μετρικές, τα πειράματα, οι αξιολογήσεις και όλη η σχετική ανάλυση που έγινε πάνω στους αλγορίθμους και τις τοπολογίες. Τέλος στο κεφάλαιο 7 διατυπώνονται τα συμπεράσματα και κάποιες σκέψεις/προτάσεις για περαιτέρω έρευνα στο συγκεκριμένο αντικείμενο.

ΚΕΦΑΛΑΙΟ 2 -Βασικές αρχές clustering και constrained clustering

2.1 Εισαγωγή

Η ανάλυση των δεδομένων και ειδικότερα η συσταδοποίηση τους (clustering) αποτελεί ένα πολύ βασικό επιστημονικό πεδίο το οποίο βρίσκει εφαρμογή σε πολλούς τομείς όπως η ιατρική, η διαφήμιση, η προώθηση προϊόντων και άλλους πολλούς. Τα τελευταία χρόνια με την εκθετική αύξηση των δεδομένων σε ιδιωτικές επιχειρήσεις και δημόσιους οργανισμούς οδηγηθήκαμε σε έναν νέο επιστημονικό κλάδο των big data. Με αυτόν τον τρόπο έγινε ακόμα πιο επιτακτική η ανάγκη για ακριβή ανάλυση, αξιολόγηση και ομαδοποίηση των δεδομένων. Μια υποκατηγορία της συσταδοποίησης που έχει μεγάλη άνθιση την τελευταία δεκαετία είναι το λεγόμενο constrained clustering (συσταδοποίηση με περιορισμούς) που αποτελεί μορφή του semi-supervised clustering (ήμι-επιβλεπόμενη μάθηση). Αυτού του είδους η μηχανική μάθηση δίνει τη δυνατότητα στους χρήστες να λάβουν υπ' όψιν τους διάφορους περιορισμούς που υπάρχουν μεταξύ των δεδομένων κάτι που η απλή συσταδοποίηση δεν λαμβάνει . Με αυτόν τον τρόπο οι χρήστες έχουν μεγαλύτερη ευελιξία και καλύτερο έλεγχο των δεδομένων και μπορούν να διαχωρίζουν τα δεδομένα τους με πιο ακριβή τρόπο.

2.2 Συσταδοποίηση Δεδομένων (Clustering)

Η συσταδοποίηση [EdBr13,AgRe14] είναι η πιο δημοφιλής μορφή της μη επιβλεπομένης μάθησης. Σκοπός της είναι η ανάλυση των δεδομένων και η ο χωρισμός τους σε ομάδες με τέτοιο τρόπο ώστε οι ομάδες αυτές να έχουν νόημα για την εκάστοτε εφαρμογή και τον χρήστη. Η κατηγοριοποίηση αυτή γίνεται με τέτοιο τρόπο ώστε η ομοιότητα μεταξύ των δεδομένων να είναι όσο τον δυνατόν πιο υψηλή ενώ η ομοιότητα μεταξύ των διαφορετικών συστάδων (clusters) να είναι όσο τον δυνατόν μικρότερη. Η συσταδοποίηση διαφέρει σημαντικά από την επιβλεπόμενη μάθηση (supervised clustering) και ειδικότερα το classification. Στη συσταδοποίηση χρησιμοποιούνται αποκλειστικά τα ίδια τα δεδομένα η χαρακτηριστικά των δεδομένων σύμφωνα με τα οποία γίνεται η αναπαράστασή τους και έτσι δημιουργούνται οι ομάδες. Αντιθέτως στο classification οι ομάδες είναι γνωστές, η ομάδα στην οποία ανήκει το κάθε ένα υπάρχον στοιχείο είναι γνωστό και στόχος είναι να ενσωματωθούν και να ομαδοποιηθούν νέα δεδομένα που το σύστημα δεν έχει δει ποτέ. Η συσταδοποίηση διαχωρίζεται σε δύο μεγάλες κατηγορίες:

1) Διαχωριστική συσταδοποίηση (partitional clustering)

Σε αυτήν την κατηγορία τα δεδομένα διαχωρίζονται σε ομάδες ανάλογα με τα χαρακτηριστικά τους. Γίνεται λοιπόν ένας διαχωρισμός του χώρου αναπαράστασης των δεδομένων δίχως να υπάρχει κάποια σαφή σχέση ή ιεραρχία. Χαρακτηριστικός αλγόριθμος αυτής της κατηγορίας αποτελεί ο k-means [St06].

2) Ιεραρχική συσταδοποίηση (hierarchical clustering)

Σε αυτήν την κατηγορία τα δεδομένα διαχωρίζονται σύμφωνα με κάποια ιεραρχία και ομαδοποιούνται με συστάδες (clusters) που έχουν όμοια χαρακτηριστικά με αυτά. Η συναθροιστική ή συσσωρευτική συσταδοποίηση (agglomerative clustering) [GaMa07, XuWu08] είναι η πιο γνωστή οικογένεια αλγορίθμων που εμπίπτει σε αυτήν την κατηγορία.

2.2.1 K-means clustering

Ο αλγόριθμος k-means [Ba13, St06] αποτελεί έναν από τους πιο γνωστούς αλγορίθμους της διαχωριστικής συσταδοποίησης (partitional clustering). Χρησιμοποιείται με δεδομένα των οποίων τα χαρακτηριστικά είναι ποσοτικοποιημένα και συνήθως με μετρική την ευκλείδεια απόσταση. Η ευκλείδεια απόσταση ορίζεται ακολούθως:

$$d(x_i, x_{i'}) = \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \qquad \text{Σχέση 2.α}$$

Τα x_i , x_i' αποτελούν παρατηρήσεις των δεδομένων με p το πλήθος χαρακτηριστικά και το x_{ij} αποτελεί τη μέτρηση του j-οστού χαρακτηριστικού της παρατήρησης x_i. Σκοπός του συγκεκριμένου αλγορίθμου είναι τα δεδομένα να κατηγοριοποιηθούν με τέτοιο τρόπο ώστε να ελαχιστοποιηθεί η παρακάτω αντικειμενική συνάρτηση:

$$\sum_{k=1}^{K} \sum_{C_i=k} \sum_{C_{i'}=k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \qquad \text{Σχέση 2.β}$$

Το K είναι ο αριθμός των συστάδων (clusters) ενώ το C_i είναι η συστάδα στην οποία ανατίθεται η παρατήρηση i όπου φυσικά ισχύει $1 \leq C_i \leq K$. Πολλές είναι οι παραλλαγές αυτού του αλγορίθμου που έχουν προταθεί αλλά όλες ακολουθούν κάποιες βασικές αρχές που συνοψίζονται στα επόμενα βήματα:

1. Τυχαία η κάθε παρατήρηση ανατίθεται σε ένα αρχικό cluster.

2. Για κάθε χαρακτηριστικό j και συστάδα k, υπολόγισε τον μέσο του χαρακτηριστικού j στην συστάδα k.

3. Ανάθεσε την παρατήρηση i σε μια νέα συστάδα C_i σύμφωνα με την ακόλουθη σχέση:

$$C_i = \arg\min_k \sum_{j=1}^{p} (x_{ij} - \bar{x}_{kj})^2 \qquad \text{Σχέση 2.γ}$$

4. Επανάλαβε τα βήματα 2 και 3 μέχρι ο αλγόριθμος να συγκλίνει.

2.2.2 Ιεραρχική Συσταδοποίηση (Hierarchical clustering)

Στην ιεραρχική συσταδοποίηση [Ba13, Co12] τα δεδομένα χωρίζονται σε ομάδες ακολουθώντας μια ιεραρχία. Ο διαχωρισμός αυτός θυμίζει μια μορφή δέντρου. Στο κατώτερο μέρος του δέντρου τα δεδομένα αποτελούν το κάθε ένα μία διαφορετική συστάδα (cluster). Όσο τα επίπεδα αυξάνονται συνενώνονται τα clusters του αμέσως προηγούμενου επιπέδου δημιουργώντας όλο και μεγαλύτερες συστάδες. Στην κορυφή του δέντρου έχουμε ένα και μοναδικό cluster το οποίο προέρχεται από τη συνένωση όλων των προηγούμενων στα πιο κάτω επίπεδα και περιέχει όλα τα δεδομένα. Το δέντρο μπορεί να παρασταθεί γραφικά μέσω του δενδρογράμματος όπου το ύψος κάθε κόμβου στο δέντρο αντιστοιχεί με την ομοιότητα των δύο clusters που συνενώθηκαν ώστε να δημιουργηθεί το ίδιο.

Μία από τις πιο γνωστές τεχνικές ιεραρχικής συσταδοποίησης είναι η συναθροιστική συσταδοποίηση κατά την οποία το κάθε δεδομένο ξεκινά ως ένα ξεχωριστό cluster και σε κάθε βήμα συνενώνονται τα πιο "όμοια" μεταξύ τους. Η συνένωση μπορεί να γίνει μεταξύ δύο cluster που έχουν μόνο ένα δεδομένο, μεταξύ cluster που έχουν περισσότερα δεδομένα είτε ανάμιξη αυτών των δύο δηλαδή ένα cluster που έχει ένα μόνο δεδομένο μπορεί να συνενωθεί με cluster που έχει περισσότερα του ενός. Το ζήτημα σε αυτή τη μέθοδο είναι να οριστούν σαφώς δύο σχέσεις. Η πρώτη είναι ένα μέτρο ομοιότητας (ή απόστασης) μεταξύ δύο μεμονωμένων σημείων και η δεύτερη είναι μέτρο ομοιότητας (ή απόστασης) μεταξύ δύο διαφορετικών clusters. Το πιο συνηθισμένο κριτήριο για την ομοιότητα μεταξύ δύο σημείων είναι η ευκλείδεια απόσταση αλλά φυσικά έχουν προταθεί περισσότερα. Όσον αφορά τις μετρικές ομοιότητας μεταξύ διαφορετικών clusters τα πιο συνήθη είναι τα εξής:

- **Μονή Σύνδεση (single linkage)**:

$$d(C_1, C_2) = \min_{i \in C_1, i' \in C_2} d(x_i, x_{i'}) \qquad \text{Σχέση 2.δ}$$

- **Πλήρης Σύνδεση (complete linkage)**:

$$d(C_1, C_2) = \max_{i \in C_1, i' \in C_2} d(x_i, x_{i'}) \qquad \text{Σχέση 2.ε}$$

- **Σύνδεση του μέσου όρου (average linkage)**:

$$d(C_1, C_2) = \frac{1}{n_1 n_2} \sum_{i \in C_1} \sum_{i' \in C_2} d(x_i, x_{i'}) \qquad \text{Σχέση 2.στ}$$

Όπου τα i, i' είναι αντίστοιχα οι i-οστές παρατηρήσεις στα cluster C_1,C_2 και συμβολίζουν τις

παρατηρήσεις x_i,x_i'. Το d συμβολίζει το μέτρο ομοιότητας μεταξύ δύο μεμονωμένων δεδομένων. Τα n_1,n_2 είναι το πλήθος των παρατηρήσεων στα cluster C_1, C_2 αντίστοιχα. Στην πρώτη μέθοδο ζητούμενο είναι η "μικρότερη" ομοιότητα μεταξύ δύο μεμονωμένων παρατηρήσεων ανάμεσα σε δυο clusters, στη δεύτερη η "μεγαλύτερη" ομοιότητα μεταξύ δύο μεμονωμένων παρατηρήσεων ενώ στην τρίτη ζητούμενο είναι ο μέσος όρος ομοιοτήτων μεταξύ όλων των παρατηρήσεων ανάμεσα στα δύο clusters.

2.3 Συσταδοποίηση υπό περιορισμούς (Constrained clustering)

Όπως ειπώθηκε και σε προηγούμενο κεφάλαιο η συσταδοποίηση με περιορισμούς (constrained clustering) αποτελεί έναν νέο τομέα έρευνας στο χώρο της ήμι-επιβλεπομένης μάθησης (semi-supervised clustering) [NaCi18]. Δίνει τη δυνατότητα στο χρήστη να ενσωματώσει μια ήδη προϋπάρχουσα γνώση που υπάρχει από τα δεδομένα και να τα χειριστεί κατάλληλα, ομαδοποιώντας τα με πιο ακριβές τρόπο. Οι πληροφορίες δεν δίνονται πλέον ως ξεχωριστά δεδομένα αλλά ως ένας πίνακας που περιέχει τις σχέσεις μεταξύ κάθε διαφορετικού ζεύγους δεδομένων και ονομάζονται περιορισμοί. Οι περιορισμοί αυτοί εκφράζουν τη δυνατότητα των δεδομένων να ανήκουν η όχι στην ίδια συστάδα. Το constrained clustering [BaDaWa08] δίνει την ικανότητα στο χρήστη να λάβει υπόψιν του σχέσεις και πληροφορίες των δεδομένων που αγνοούνταν στη συμβατική συσταδοποίηση ή την επιβλεπόμενη ομαδοποίηση. Εντούτοις δεν είναι ο χρήστης ή ο προγραμματιστής αυτός που θα αποφασίσει ποια θα είναι τα clusters αλλά ο αλγόριθμος. Ο προγραμματιστής απλά θα δώσει κάποιες επιπλέον πληροφορίες που θα βοηθήσουν τον αλγόριθμο να κάνει πιο ακριβή εκπαίδευση στα δεδομένα.

2.3.1 Περιορισμοί (Constraints)

Υπάρχουν δύο ειδών περιορισμοί που μπορούν να εξαχθούν από τα δεδομένα:

- **Θετικός περιορισμός (Must-Link)**, ML (a,b), που υποδεικνύει ότι δύο δεδομένα πρέπει να ομαδοποιηθούν στο ίδιο cluster.

- **Αρνητικός περιορισμός (Cannot-Link)**, CL (a,b), που υποδεικνύει ότι δύο δεδομένα πρέπει να ομαδοποιηθούν σε διαφορετικά clusters.

Υπάρχουν δύο κατηγορίες αλγορίθμων στην οικογένεια του constrained clustering. Στην πρώτη κατηγορία ανήκουν οι αλγόριθμοι εκείνοι που λαμβάνουν με μεγάλη αυστηρότητα το σετ των περιορισμών και δεν επιδέχονται να παραβιάζεται ούτε ένας όπως πχ ο Cop k-means. Αν δηλαδή την ώρα που εξελίσσεται η διαδικασία, δύο δεδομένα που συνδέονται με έναν σύνδεσμο ML κατηγοριοποιηθούν σε διαφορετικές συστάδες, τότε ο αλγόριθμος δεν θα έχει καν έξοδο αλλά απλά ένα μήνυμα ότι δεν μπόρεσε να πραγματοποιηθεί η συσταδοποίηση. Στην δεύτερη κατηγορία ανήκουν αλγόριθμοι που λαμβάνουν υπόψιν τους το σετ των περιορισμών και το χρησιμοποιούν ως οδηγό για να πραγματοποιήσουν τη συσταδοποίηση όσο γίνεται με μεγαλύτερη ακρίβεια. Επιδέχονται όσο το δυνατόν λιγότερους περιορισμούς στα clusters. Τέτοιοι αλγόριθμοι είναι οι pck-means , constrained agglomerative clustering, constrained spectral clustering. Ο στόχος τους είναι να έχουν πάντα έξοδο η οποία θα βγάζει ακριβείς συστάδες σεβόμενες τους περιορισμούς χωρίς

όμως να απαγορεύει την ενσωμάτωση ορισμένων περιορισμών για την επίτευξη αυτού.

Από μαθηματικής απόψεως η σχέση ML (a,b) είναι μεταβατική, ML (a,b) ^ ML (b,c) → ML (a,c), καθώς επίσης και συμμετρική και τετριμμένα ανακλαστική. Έτσι αποτελεί σχέση ισότητας. Αντιθέτως η σχέση CL (a,b) δεν είναι μεταβατική, είναι συμμετρική και τετριμμένα είναι μη ανακλαστική [BaDaWa08]. Παρόλα αυτά είναι δυνατόν να εξάγουμε γνώση και από το σετ των cannot-link αν το συνδυάσουμε με το σετ των must-link. Οι λίστες των ML δημιουργούν κλάσεις ισότητας (equivalence classes) και η ύπαρξη ML υπονοεί ότι όλα τα στοιχεία εντός των κλάσεων πρέπει να βρίσκεται στην ίδια συστάδα. Η ύπαρξη ενός CL μεταξύ ενός δεδομένου της κλάσης και ενός άλλου δεδομένου δημιουργεί αυτόματα και cannot links μεταξύ του δεύτερου δεδομένου και όλων των δεδομένων που βρίσκονται στην ίδια κλάση με το πρώτο. Μαθηματικά αυτό εκφράζεται ως εξής: CL (a,b) ^ ML (b,c) → CL (a,c).

2.4 Αλγόριθμοι που υλοποιούν συσταδοποίηση υπό περιορισμούς (Constrained Clustering Algorithms)

Υπάρχουν δύο είδη Αλγορίθμων που εμπίπτουν στην κατηγορία του constrained clustering και ο διαχωρισμός αυτός εξαρτάται από τον τρόπο που οι αλγόριθμοι διαχειρίζονται τις πληροφορίες που περιέχονται στους περιορισμούς. Οι δύο κατηγορίες αλγορίθμων είναι οι εξής:

- **Αλγόριθμοι βασισμένοι στους περιορισμούς (constraint-based algorithms)**: Σε αυτήν την κατηγορία τα βήματα που ακολουθούνται στον αλγόριθμο, όπως η αρχικοποίηση των clusters και η ανάθεση των δεδομένων στα διάφορα clusters, μεταβάλλονται με τέτοιο τρόπο ώστε να ενσωματώσουν τους διάφορους περιορισμούς και να δημιουργούν συσταδοποίηση που είναι σύμφωνη με αυτούς. Οι περιορισμοί έχουν άμεσο αντίκτυπο στον τρόπο που διεξάγονται τα βήματα του αλγορίθμου.

- **Αλγόριθμοι βασισμένοι στην απόσταση (distance-based algorithms)**: Σε αυτήν την κατηγορία οι πληροφορίες που παρέχονται από τους περιορισμούς χρησιμοποιούνται προκειμένου να διαμορφωθεί μια μετρική απόστασης μεταξύ των δεδομένων. Στόχος δηλαδή των αλγορίθμων αυτών είναι η μετρική απόστασης. Δεδομένα που ενώνονται μεταξύ τους με θετικούς συνδέσμους (ML link) έρχονται πιο κοντά μεταξύ τους, ενώ δεδομένα που συνδέονται με αρνητικούς συνδέσμους (CL link) διαχωρίζονται και απομακρύνονται. Έτσι διαμορφώνονται με ανάλογο τρόπο και τα clusters.

2.4.1 Constrained Complete Link

Το constrained complete link (CCL) [KlKaMa02] αποτελεί έναν distance-based αλγόριθμο που βασίζεται πάνω στους CL και ML συνδέσμους που υπάρχουν μεταξύ των δεδομένων. Αποτελεί έναν ιεραρχικό συναθροιστικό αλγόριθμο.

Σε αυτόν τον αλγόριθμο τα δεδομένα ξεκινούν ως ξεχωριστά clusters και ενοποιούνται μεταξύ τους ανάλογα με το ποια είναι "πιο όμοια" κάθε φορά. Η απόσταση μεταξύ των clusters ορίζεται ως η μεγαλύτερη απόσταση μεταξύ των σημείων τους. Σε μαθηματικούς όρους

dist(ω,ω')=max{dist (x,x')| $x_{\epsilon}\omega$, $x'_{\epsilon}\omega'$ } Δημιουργείται έτσι μια ιεραρχία από clusters η οποία μετά το πέρας του αλγορίθμου μπορεί να χρησιμοποιηθεί αναλόγως. Χάριν αποδοτικότητας του αλγορίθμου οι αποστάσεις μεταξύ των clusters αποθηκεύονται σε μία δομή προκειμένου να μην υπολογίζονται συνεχώς και να πραγματοποιούνται μόνο οι απαραίτητες αλλαγές. Όταν δύο clusters i,j συγχωνευθούν τότε στην δομή που είναι αποθηκευμένες οι αποστάσεις των clusters αλλάζουν μόνο οι αποστάσεις μεταξύ του νέου cluster και οποιουδήποτε άλλου. Πιο συγκεκριμένα για κάποιο cluster k η νέα τιμή που θα αποθηκευθεί θα είναι η μεγαλύτερη απόσταση μεταξύ του cluster k και οποιουδήποτε από τα cluster που μόλις συγχωνεύθηκαν δηλαδή max{dist (i,k), dist (j,k)}. Ο ψευδοκώδικας παρατίθεται παρακάτω:

```
COMPLETE LINK (CL)
Είσοδος: X, δεδομένα προς συσταδοποίηση
Έξοδος: Βήματα, Τα βήματα που ακολουθήθηκαν για την
διεκπεραίωση του  αλγορίθμου;

1 για i←1 μέχρι n κάνε: ωi←{xi}
2 D←Υπολόγισε τις αποστάσεις μεταξύ των clusters(Ω)
3 Όσο (Ω > 1) κάνε:
    4 (ωi,ωj)←Πάρε τις κοντινότερες αποστάσεις(D)
    5 Πρόσθεσε ((i,j), Βήματα)
    6 ΣυγχώνευσεΣυστάδες (i,j,Ω,D)
7 τέλος
```

Ο αλγόριθμος αυτός σε αυτήν τη μορφή δεν λαμβάνει υπ' όψιν του τους περιορισμούς που υπάρχουν μεταξύ των διάφορων δεδομένων. Μια μικρή παραλλαγή μετά την αρχικοποίηση του πίνακα D, δηλαδή μετά το πέρας του βήματος 2, προκειμένου να ενσωματώσουμε τέτοιου είδους περιορισμούς θα ήταν η εξής:

- Αν δύο δεδομένα συνδέονται με έναν θετικό σύνδεσμο (ML link) τότε η απόσταση μεταξύ των cluster τους (σε αυτή τη φάση το κάθε δεδομένο αποτελεί διαφορετικό cluster) τίθεται στο 0. Διαισθητικά τα φέρνουμε πιο κοντά αφού μηδενίζουμε τη μεταξύ τους απόσταση.

- Αν δύο δεδομένα συνδέονται με έναν αρνητικό σύνδεσμο (CL link) τότε η απόσταση μεταξύ των cluster τους τίθεται στη μεγαλύτερη δυνατή απόσταση (∞). Διαισθητικά τα απομακρύνουμε μεγαλώνοντας τη μεταξύ τους απόσταση.

Οι αλλαγές που γίνονται στον πίνακα D έχουν ως στόχο τη μεταβολή στη συμπεριφορά του αλγορίθμου προκειμένου να είναι σύμφωνη με τους διάφορους περιορισμούς. Προκαλείται έτσι μια αλλαγή σε επίπεδο χώρου καθώς τα σημεία των οποίων η απόσταση τίθεται στο 0 έχουν μεγαλύτερη πιθανότητα να ομαδοποιηθούν στην ίδια συστάδα ενώ εκείνα των οποίων η απόσταση τίθεται στο άπειρο έχουν μικρότερη πιθανότητα να κατηγοριοποιηθούν στην ίδια συστάδα. Όσον αφορά τους θετικούς συνδέσμους, είναι πιθανό να δημιουργηθούν παραβιάσεις στην τριγωνική ανισότητα. Το πρόβλημα αυτό μπορεί να ξεπεραστεί υπολογίζοντας το μικρότερο μονοπάτι μεταξύ των σημείων x,x' θέτοντας αυτήν την απόσταση στον πίνακα D. Το μονοπάτι αυτό θα πρέπει να αποτελείται είτε από τα σημεία x,x' είτε από οποιοδήποτε άλλο σημείο x'' που συνεπάγεται έναν ML σύνδεσμο. Ο ψευδοκώδικας που δείχνει την παραλλαγή του πιο πάνω αλγορίθμου είναι ο εξής:

```
CONSTRAINED COMPLETE LINK (CCL)
Είσοδος:  X, δεδομένα προς συσταδοποίηση;
          k, αριθμός των clusters;
          ML and CL, θετικοί και αρνητικοί περιορισμοί
          που πρέπει να ληφθούν υπ' όψιν;
Έξοδος:   Βήματα, Τα βήματα που ακολουθούνται από τον
          αλγόριθμο

1 για i←1 μέχρι n κάνε: ωi←{xi}
2 D← Υπολόγισε τις αποστάσεις μεταξύ των clusters(Ω)
3 ΕπίβαλεΠεριορισμούς (D,ML,CL)
4 Όσο (Ω > 1) κάνε:
     5 (ωi,ωj)← Πάρε τις κοντινότερες αποστάσεις(D)
     6 Πρόσθεσε ((i,j), Βήματα)
     7 ΣυγχώνευσεΣυστάδες (i,j,Ω,D)
8 τέλος

Συνάρτηση ΕπίβαλεΠεριορισμούς (D,ML,CL)
     Είσοδος: D, Δομή που αποθηκεύονται οι αποστάσεις
              μεταξύ των clusters;
              ML and CL, θετικοί και αρνητικοί
              περιορισμοί που πρέπει να ληφθούν υπ' όψιν

     9 Για κάθε (x,x')ϵML κάνε: D(x,x')←0
     10 ΔιάδωσεMustLinks (D, ML)
     11 Για κάθε (x,x')ϵCL κάνε: D(x,x')←∞
τέλος
```

2.4.2 Φασματική συσταδοποίηση υπό περιορισμούς (Spectral Clustering with Imposed constraints)

Στην ίδια λογική με τον CCL προτάθηκε ένας νέος αλγόριθμος ο οποίος μπορούσε να εισάγει αρνητικούς και θετικούς συνδέσμους σε έναν αλγόριθμο φασματικής συσταδοποίησης (spectral clustering algorithm) [KlKaMa03].

Αρχικά από τα δεδομένα που έχουμε ως είσοδο δημιουργούμε έναν πίνακα γειτνίασης A ο οποίος στη θέση i,j έχει την τιμή a_{ij} που δηλώνει την ομοιότητα του στοιχείου i με το στοιχείο j. Τα ιδιοδιανύσματα του πίνακα N, ο οποίος προέρχεται από τον πίνακα A ύστερα από κατάλληλη επεξεργασία, χρησιμοποιούνται για να απεικονίσουν τα δεδομένα σε διανύσματα στο χώρο R^k. Η σχέση σύμφωνα με την οποία γίνεται η επεξεργασία του πίνακα A είναι η εξής:

$$N = \frac{1}{d_{max}}(A + d_{max}I - D) \qquad \text{Σχέση 2.ζ}$$

Όπου D είναι διαγώνιος πίνακας όπου στη διαγώνιο έχει τα εξής στοιχεία

- $d_{ii} = \sum_{j=1}^{n} a_{ij}$

I είναι ο μοναδιαίος πίνακας , d_{max} είναι το μέγιστο στοιχείο του πίνακα D.

Ο αλγόριθμος SCIC (spectral clustering with imposed constraints) επίσης αποτελεί έναν distance-based αλγόριθμο ο οποίος έχει το πλεονέκτημα ότι δεν χρειάζεται διάδοση των περιορισμών του σε αντίθεση με τον CCL. Οι περιορισμοί που χρειάζεται να επιβάλλουμε στον πίνακα A, ύστερα από την αρχικοποίηση του, είναι οι εξής:

- Αν μεταξύ δύο δεδομένων x_i,x_j υπάρχει ένας θετικός σύνδεσμος (ML link) τότε θα πρέπει στη θέση i,j αλλά και στη θέση j,i του πίνακα A να θέσουμε την τιμή ίση με 1.

- Αν μεταξύ δύο δεδομένων x_i,x_j υπάρχει ένας αρνητικός σύνδεσμος (ML link) τότε θα πρέπει στη θέση i,j αλλά και στη θέση j,i του πίνακα A να θέσουμε την τιμή ίση με 0.

Ο ψευδοκώδικας είναι ο εξής:

```
SPECTRAL CLUSTERING WITH IMPOSED CONSTRAINTS (SCIC)
Είσοδος:  X, δεδομένα προς συσταδοποίηση;
          k, αριθμός των clusters;
          ML and CL, θετικοί και αρνητικοί περιορισμοί
          που πρέπει να ληφθούν υπ' όψιν
Έξοδος: = Ω={ω1,ω2,...,ωk}, διαμέριση των δεδομένων

1 Α←ΥπολόγισεΠίνακαΓειτνίασης(X)
2 ΕπίβαλεΠεριορισμούς (A,ML,CL)
3 N←ΥπολόγισεN (A)
4 Y←ΑπόκτησεΑπεικονίσεις (N,d)
5 Για κάθε i←1 έως n κάνε: pi←i^th γραμμή του Y
6 Ω'←Cluster ({p1,p2...pn},k)
7 Για κάθε ωi'εΩ' κάνε:
     Για κάθε pjεωi' κάνε: Επισύναψε (xj,ωi)

function ΕπίβαλεΠεριορισμούς (A,ML,CL)
     Είσοδος : A, Πίνακας Γειτνίασης; ML and CL, θετικοί
     και αρνητικοί περιορισμοί  που πρέπει να ληφθούν
     υπ' όψιν
     8 Για κάθε (xi,xj)εML κάνε: a(i,j)←a (j,i)←1
     9 Για κάθε (xi,xj)εCL κάνε: a(i,j)←a (j,i)←0
τέλος
```

2.4.3 COP k-means

Ο ψευδοκώδικας πάνω στον οποίο στηρίζεται ο cop k-means [Ba13] είναι ο εξής:

1. Τυχαία επισύναψε κάθε στοιχείο σε μια αρχική συστάδα.
2. Για κάθε χαρακτηριστικό j και για κάθε συστάδα k υπολόγισε το $x_{kj,}$ το μέσο του χαρακτηριστικού j στην συστάδα k.
3. Επισύναψε κάθε δεδομένο i σε μια νέα συστάδα C_i ως ακολούθως:

$$C_i = \arg\min_{k \in D_{ik}} \sum_{j=1}^{p} (x_{ij} - \bar{x}_{kj})^2 \qquad \text{Σχέση 2.η}$$

   D_{ik}={κανένας περιορισμός δεν παραβιάζεται όταν το δεδομένο i επισυνάπτεται στη συστάδα k}

4. Επανάλαβε τα βήματα 2 και 3 μέχρι ο αλγόριθμος να συγκλίνει. Ο αλγόριθμος αποτυγχάνει αν $D_{ik} = \emptyset$ για οποιοδήποτε i σε οποιοδήποτε βήμα της διαδικασίας.

Ο cop k-means αποτελεί έναν άπληστο αλγόριθμο που είναι ταυτόσημος με τον συμβατικό k-means. Η μόνη τους διαφορά είναι ότι ο πρώτος επισυνάπτει τα δεδομένα στο κοντινότερο cluster λαμβάνοντας υπόψιν όλους τους περιορισμούς και μη επιτρέποντας τον παραβιασμό έστω και ενός, ενώ ο δεύτερος δεν λαμβάνει υπόψιν τους περιορισμούς. Εξαιτίας του άπληστου αυτού κριτηρίου που χρησιμοποιεί ο cop k-means σε πολλές περιπτώσεις δεν μπορεί να βρει καν λύση στο πρόβλημα.

2.4.4 PCk-means

Ο PCk-means [EdBr13] αποτελεί επίσης μια παραλλαγή του συμβατικού k-means και σκοπός του είναι να εισάγει τους διάφορους περιορισμούς που δεν λαμβάνει υπόψιν του ο συμβατικός αλγόριθμος. Χρησιμοποιεί έτσι τους θετικούς και αρνητικούς περιορισμούς για να επηρεάσει τόσο την αρχικοποίηση των συστάδων όσο και και την ανάθεση των δεδομένων σε αυτές. Αποτελεί λοιπόν έναν αλγόριθμο βασισμένο στους περιορισμούς (constraint-based algorithm).

Έτσι οι Basu, Davidson, Wagstaff που εισήγαγαν αυτόν τον αλγόριθμο το 2004 πρότεινε μια νέα τεχνική για την αρχικοποίηση των συστάδων. Παίρνουμε το μεταβατικό κλείσιμο των θετικών (must-link) και αρνητικών συνδέσμων (cannot-link). Το κέντρο της κάθε μίας από τις k μεγαλύτερες γειτονιές που δημιουργούνται χρησιμοποιείται στην αρχικοποίηση των k συστάδων. Αν οι γειτονιές αυτές είναι λιγότερες από k τότε το επόμενο cluster αρχικοποιείται με ένα δεδομένο που συνδέεται με αρνητικούς συνδέσμους (cannot-link) με όλες τις υπόλοιπες γειτονιές. Τα

υπόλοιπα clusters αρχικοποιούνται με τυχαίες διαταραχές του κέντρου όλων των δεδομένων.

Όσον αφορά την ανάθεση των δεδομένων στις διάφορες συστάδες ο Basu et al πρότεινε μια νέα αντικειμενική συνάρτηση, η οποία θα έπρεπε να ελαχιστοποιηθεί κατά την διαδικασία της συσταδοποίησης, την J_{pckm}:

$$\mathcal{J}_{\text{pckm}}(\Omega) = \frac{1}{2}\sum_{i=1}^{k}\sum_{x\in\omega_i}\|x-\overline{\omega_i}\|^2 + \sum_{(x_i,x_j)\in ML} w_{ij}\mathbb{1}[l_i \neq l_j] + \sum_{(x_i,x_j)\in CL} \bar{w}_{ij}\mathbb{1}[l_i = l_j]$$

Σχέση 2.θ

Ο πρώτος όρος της αντικειμενικής συνάρτησης είναι ίδιος με την αντικειμενική συνάρτηση του συμβατικού αλγορίθμου και αποτελεί ουσιαστικά το κριτήριο για το πόσο συμπαγείς είναι οι συστάδες. Ο δεύτερος όρος μετρά τον βαθμό στον οποίο τηρούνται οι θετικοί περιορισμοί (must link) προσθέτοντας ένα πέναλτι w_{ij} σε περίπτωση που ένας τέτοιος σύνδεσμος μεταξύ δύο δεδομένων x_i και x_j παραβιάζεται δηλαδή οι ετικέτες τους είναι διαφορετικές. Ο τρίτος όρος μετρά τον βαθμό στον οποίο τηρούνται οι αρνητικοί περιορισμοί (cannot link) προσθέτοντας ένα πέναλτι w_{ij}' σε περίπτωση που ένας τέτοιος σύνδεσμος παραβιάζεται δηλαδή οι ετικέτες τους είναι ίδιες.

Παρακάτω παρατίθεται ο ψευδοκώδικας του αλγορίθμου στον οποίο γίνεται φανερό ειδικά στη γραμμή 6 ότι κατά την ανάθεση των δεδομένων λαμβάνονται υπ' όψιν τα πιθανά πέναλτι όταν παραβιάζονται ορισμένοι περιορισμοί. Παρόλο που τα πέναλτι μπορεί να είναι διαφορετικά για κάθε περιορισμό σε αυτόν τον ψευδοκώδικα. Ανάλογα με τον αριθμό των γειτονιών ο αλγόριθμος αυτός μπορεί να είναι εξαρτώμενος από διάφορες αρχικές συνθήκες (πχ τυχαιότητα και διαταραχές στο κέντρο όλων των δεδομένων).

```
PAIRWISE CONSTRAINED K-MEANS (PCKM)
Είσοδος : X, δεδομένα προς συσταδοποίηση;
          k, αριθμός των clusters;
          ML and CL, θετικοί και αρνητικοί
          περιορισμοί που πρέπει να ληφθούν υπ' όψιν
          w, το πέναλτι στους περιορισμούς
Έξοδος : Ω={ω1,ω2...ωk}, διαμέριση των δεδομένων

1 Μεταβατικό κλείσιμο (M,C)
2 ΑρχικοποίησεPCKM (k,X,ML,CL)
3 Όσο η σύγκλιση δεν έχει επιτευχθεί κάνε:
    4 Για κάθε ωϵΩ κάνε: ΕπαναυπολογισμόΚέντρων (ω)
    5 Για κάθε xϵX κάνε:
        6 i←argmin_{jϵ1...k} ( 1/2|| x-ωj ||^2+
          Πέναλτι (x,ω,ML,CL))
        7 Επισύναψε (x,ωi)
    8 τέλος
9 τέλος

Συνάρτηση ΑρχικοποίησεPCKM (k,X,ML,CL)
Είσοδος : k, ο αριθμός των clusters;
          X, δεδομένα προς συσταδοποίηση;
          ML and CL, θετικοί και αρνητικοί περιορισμοί
```

```
     10 {N1,N2,...Nν}←ΓειτονιεςΤαξινομημένεςκαταΜέγεθος
        (ML,CL)
     11 Αν k≤ν τότε:
          12 Για i←1 έως k κάνε:
               Αρχικοποίησε (ωi,Κέντρο(Ni))
     13 Αλλιώς
          14 Για i←1 έως ν κάνε:
               Αρχικοποίησε (ωi,Κέντρο (Ni))
          15 Αν υπάρχει x συνδεδεμένο με CL με όλα τα Ni
             τότε: Αρχικοποίησε (ων+1,x)
          16 Για κάθε ω το οποίο ακόμα δεν έχει
             αρχικοποιηθεί κάνε:
               17 Αρχικοποίησε (ω, ΤυχαίαΔιαταραχή
               (Κέντρο(X)))
          18 τέλος
     19 τέλος
τέλος

Συνάρτηση Πέναλτι (x,ω,ML,CL,w)
Είσοδος : x, ένα δεδομένο;
          ω, ένα cluster;
          ML and CL, θετικοί και αρνητικοί περιορισμοί;
          w, πέναλτι στους περιορισμούς
Έξοδος: p, το πέναλτι που δημιουργείται αν το x
        επισυναφθεί στο ω
          20 p←0
          21 Για κάθε (x,x')εML κάνε: Αν x δεν ανήκει ω
          τότε: p←p+w
          22 Για κάθε (x,x')εCL κάνε: Αν xεω τότε:
          p←p+w
          23 Επέστρεψε p
τέλος
```

ΚΕΦΑΛΑΙΟ 3 -Συναφής Βιβλιογραφία

` Στο κεφάλαιο αυτό γίνεται παράθεση των σχετικών άρθρων και δημοσιεύσεων που ασχολούνται με επίλυση προβλημάτων σε ασύρματα δίκτυα ανίχνευσης και επεξηγούνται ορισμένες έννοιες και ορισμοί πολύ σημαντικοί για την συγκεκριμένη διπλωματική. Αρχικά γίνεται ανάλυση του δικτύου πάνω στο οποίο στηριχτήκαμε, της μορφής των δεδομένων καθώς και του τρόπου επεξεργασίας τους. Επίσης επεξηγούνται οι διάφορες μετρικές που χρησιμοποιούνται για την αξιολόγηση των αλγορίθμων τόσο όσον αφορά το πείραμα αλλά και για την μεταξύ τους σύγκριση. Τέλος, αναφέρεται από τη βιβλιογραφία ένας κατανεμημένος αλγόριθμος που χρησιμοποιείται σε τέτοιου είδους πειράματα και σε επόμενα κεφάλαια γίνεται σύγκριση με δικές μας προτεινόμενες μεθόδους.

3.1 Θέματα ιδιωτικότητας σε προβλήματα εξόρυξης δεδομένων και βάσεων

Η εξόρυξη δεδομένων αποτελεί ένα πεδίο μελέτης που μπορεί να εφαρμοστεί σε διάφορα προβλήματα που εγείρουν ζητήματα ασφάλειας. Μια καινοτόμα τεχνική αποτελεί η κατασκευή ταξινομητών (decision-tree classifiers) σε δεδομένα στα οποία οι αρχικές τιμές του δείγματος έχουν αλλοιωθεί από μια συνάρτηση τυχαίας λειτουργίας (randomizing function) [AgSr00]. Ενώ δεν είναι δυνατό να εκτιμηθούν οι αρχικές τιμές, εντούτοις με μία διαδικασία ανακατασκευής, που προτείνεται στο άρθρο αυτό, είναι δυνατό να εκτιμηθεί με μεγάλη ακρίβεια η κατανομή των αρχικών δεδομένων. Συνεπώς, δοθέντος n το πλήθος τελικών τιμών $x_1+y_1,x_2+y_2,\ldots,x_n+y_n$ όπου $x_1, \ldots,x_n$ οι αρχικές τιμές που ακολουθούν ίδια κατανομή, $y_1,\ldots,y_n$, οι τιμές αλλοίωσης που ακολουθούν επίσης ίδια κατανομή, συνήθως την κατανομή Gauss η την ομοιόμορφη, και γνωρίζοντας την συσσωρευτική συνάρτηση κατανομής F_Y της ανεξάρτητης μεταβλητής Y, στόχος είναι να εκτιμηθεί η συνάρτηση κατανομής F_x της ανεξάρτητης μεταβλητής X. Γι αυτό το σκοπό και με τη βοήθεια αλγορίθμων, όπως Local και Byclass, δημιουργούνται ακριβή μοντέλα πρόβλεψης.

Η προστασία ευαίσθητων δεδομένων αποτελεί ένα υψίστης σημασίας θέμα ειδικά τη σημερινή εποχή με τη δημιουργία εφαρμογών που απαιτούν όλο και περισσότερα προσωπικά δεδομένα αλλά και την ραγδαία αύξηση των κακόβουλων επιθέσεων που γίνεται εναντίον τους. Γι αυτό το λόγο χρησιμοποιούνται διάφοροι μέθοδοι αλλοίωσης των δεδομένων αυτών ούτως ώστε ακόμα κι αν υποκλαπούν να μην είναι σε μορφή αξιοποιήσιμη από τους εισβολείς. Ωστόσο είναι αναγκαίο πολλές φορές τα δεδομένα αυτά να είναι διαθέσιμα για στατιστική ανάλυση. Γι αυτόν ακριβώς το λόγο μελετήθηκε η αλλοίωση δεδομένων με πιθανοτική κατανομή [LiChLi85]. Η τεχνική αυτή προσδίδει ένα πολύ σημαντικό πλεονέκτημα καθώς τα αλλοιωμένα δεδομένα εμφανίζουν ασυμπτωματικά τα ίδια χαρακτηριστικά καθώς έχουν την ίδια κατανομή με τα αρχικά. Η μέθοδος αυτή αποτελείται από 3 κύρια βήματα. Το πρώτο βήμα είναι η αναγνώριση της συνάρτησης πυκνότητας των αρχικών δεδομένων και η εκτίμηση των παραμέτρων που σχετίζονται με αυτή. Αυτό το βήμα επιτυγχάνεται είτε επιλέγοντας μία συνάρτηση πυκνότητας από ένα προκαθορισμένο σύνολο όπως Poisson, εκθετική, κανονική, γάμμα και άλλες σύμφωνα με κάποια τεστ μέγιστης απόκλισης Kolmogorov- Simirnov είτε ακολουθείται μια τεχνική σύμφωνα με την οποία τα αρχικά δεδομένα χωρίζονται σε κάποια διαστήματα μετρώντας τη συχνότητα που εμφανίζονται τα δεδομένα μέσα σε αυτά. Το δεύτερο βήμα είναι η δημιουργία αλλοιωμένων δεδομένων από την εκτιμώμενη συνάρτηση πυκνότητας. Οι παράμετροι που εκτιμήθηκαν στο πρώτο βήμα χρησιμοποιούνται για να παραχθούν τα καινούρια δεδομένα. Το τρίτο βήμα είναι η αντιστοίχηση και αντικατάσταση των αλλοιωμένων δεδομένων στη θέση των αρχικών. Στο βήμα

αυτό τα τελικά και αρχικά δεδομένα μοιράζονται ασυμπτωματικά τα ίδια χαρακτηριστικά και την ίδια συνάρτηση πυκνότητας. Παίρνοντας ένα δείγμα X_n με n αρκετά μεγάλο από τα τελικά δεδομένα βλέπουμε ότι ο μέσος του X_n συγκλίνει με το μέσο του πληθυσμού X και το δείγμα αυτό γίνεται ασυμπτωματικά ίδιο με ένα δείγμα n στοιχείων από τα αρχικά.

Ένα πολύ σημαντικό ζήτημα στις βάσεις δεδομένων αποτελεί η ασφάλεια των δεδομένων και ο τρόπος επεξεργασίας τους. Στις βάσεις είναι σημαντικό να υπάρχει έλεγχος όχι μόνο σε τι είδους πληροφορίες έχει πρόσβαση ο χρήστης αλλά και στον τρόπο με τον οποίο μπορεί να επεξεργαστεί αυτό το κομμάτι των δεδομένων [Mi76]. Είναι αναγκαίο να υπάρχει ένας έλεγχος στην συμπεριφορά των προγραμμάτων που χρησιμοποιούν οι χρήστες για να αλληλεπιδράσουν με τη βάση. Έτσι, εισήχθηκε η έννοια του subscheme όπου ο χρήστης δεν έχει το δικαίωμα να δει ολόκληρη τη βάση δεδομένων αλλά ένα υποσύνολο αυτής. Ο χρήστης επιλέγει, λοιπόν, ένα subscheme της βάσης συνδέοντας το με μία γλώσσα L, η οποία έχει συγκεκριμένες δυνατότητες επεξεργασίας της βάσης, και δημιουργείται μια νέα γλώσσα L_Σ . Ο χρήστης είναι πλέον εξουσιοδοτημένος να αλληλεπιδράσει με τη βάση σύμφωνα με τους κανόνες που ορίζονται για το συγκεκριμένο subscheme και τη γλώσσα L_Σ.

3.2 Πρωτόκολλα μετάδοσης δεδομένων με γνώμονα την ενέργεια και την βελτίωση της ιδιωτικότητας

Πολύ σημαντική είναι επίσης και η κατανάλωση ενέργειας στα ασύρματα δίκτυα. Γι αυτό το λόγο έχει μελετηθεί και κατασκευασθεί ένα μοντέλο επικοινωνίας βασισμένο σε γεγονότα για ασύρματα multi-hop δίκτυα [CeFlSu03]. Το δίκτυο διαρθρώνεται σε ένα δέντρο μετάδοσης γεγονότων στο οποίο οι κόμβοι εγγράφονται μόνο σε γεγονότα που τους "ενδιαφέρουν" και η μετάδοση τους γίνεται είτε από τα ανώτερα ιεραρχικά επίπεδα προς τα κατώτερα είτε αντίστροφα. Η παρουσία ενός δρομολογητή που διανέμει δυναμικά τα διαστήματα χρόνου, τα οποία υποδεικνύουν τον τύπο των δεδομένων και την καθοδική ή ανοδική πορεία τους στο δέντρο,και σύμφωνα με δύο αλγορίθμους, τον deterministic και τον speculative, διασφαλίζει την μείωση της συνολικής ενέργειας. Οι κόμβοι καταναλώνουν ενέργεια μόνο όταν χρειάζεται να στείλουν ή να λάβουν ένα γεγονός που τους "ενδιαφέρει" αλλιώς εισέρχονται σε κατάσταση εξοικονόμησης ενέργειας. Το τίμημα μιας τέτοιας τεχνικής είναι φυσικά η αύξηση της καθυστέρησης στη μετάδοση των συμβάντων.

Η ανάγκη για αποδοτική μετάδοση των δεδομένων και ασφαλή πρόσβαση στα δεδομένα του δικτύου έχουν οδηγήσει στη δημιουργία των data-centric sensor networks (DCSN). Ωστόσο διάφορα προβλήματα ασφάλειας που δημιουργούνται από επίδοξους εισβολείς οδήγησαν στη δημιουργία μιας νέας γενιάς δικτύων των pDCS, τα οποία προσφέρουν διαφορετικά επίπεδα ασφάλειας, τα οποία βασίζονται σε διαφορετικά κρυπτογραφικά κλειδιά [ShZhZh09]. Επιπλέον στη συγκεκριμένη μελέτη προτείνονται και δύο τεχνικές για query optimization, η Euclidean Steiner Tree (EST) και η Keyed Bloom Filter (KBF).Το δίκτυο είναι χωρισμένο σε κελιά και η γενικότερη λειτουργία του pDCS έχει ως εξής, όταν ένα κελί u ανιχνεύσει ένα γεγονός E, 1) Το κελί u αποφασίζει την τοποθεσία του κελιού αποθήκευσης v μέσω μιας συνάρτησης κατακερματισμού, 2) Κρυπτογραφεί το μήνυμα που περιέχει το γεγονός με το κλειδί του κελιού του, 3) Προωθεί το πακέτο προς τον προορισμό σύμφωνα με τεχνικές και πρωτόκολλα ασφάλειας, 4) Το κελί αποθήκευσης v το αποθηκεύει τοπικά και 5) αν κάποιο εξουσιοδοτημένο sink ενδιαφέρεται για το γεγονός E που μετέδωσε το u, τότε εκείνο εντοπίζει το v και στέλνει ένα query. Όσον αφορά το query optimization στην πρώτη τεχνική, το EST, υπάρχουν κάποια κελιά που ονομάζονται Steiner cells διαφορετικά από τα κελιά αποθήκευσης και είναι οργανωμένα σε μορφή EST δέντρου με κορυφή το mobile sink που δημιουργεί το query. Ο κόμβος αυτός προωθεί το μήνυμα στα παιδιά

του. τα οποία κατασκευάζουν EST υποδέντρα και επαναπροωθούν το μήνυμα, ώσπου να φτάσει στους κόμβους αποθήκευσης. Στη δεύτερη τεχνική, το KBF, έχουμε ένα σύνολο $S=s_1,s_2,\ldots,s_n$, k συναρτήσεις κατακερματισμού και ένα string από m bits που αρχικοποιούνται στο 0. Για κάθε $s \in S$, το βάζουμε ως είσοδο σε κάθε μία από τις k συναρτήσεις και αποκτούμε τις τιμές hi (s) ($1 \leq i \leq k$). Τα bits που αντιστοιχούν σε αυτές τις τιμές θέτονται στην τιμή 1. Επιπλέον χρησιμοποιούνται τα κλειδιά των κελιών για να κρυπτογραφηθούν τα id των κόμβων αποθήκευσης πριν αυτά εισέλθουν στον αλγόριθμο.

3.3 Αντιμετώπιση προβλημάτων ιδιωτικότητας και διαθεσιμότητας στα συστήματα αποθήκευσης δεδομένων

Ένα πολύ σημαντικό μέρος ενός ασύρματου δικτύου ανίχνευσης αποτελούν οι κόμβοι αποθήκευσης και γενικότερα το σύστημα σύμφωνα με το οποίο γίνεται η αποθήκευση των δεδομένων. Μια πολύ σημαντική μελέτη που έγινε σχετικά με το ζήτημα αυτό, είναι το Safestore,το οποίο αποτελεί ένα κατανεμημένο σύστημα αποθήκευσης σχεδιασμένο να διατηρεί μακροπρόθεσμα τα δεδομένα ανεξάρτητα από οποιαδήποτε αστοχία δικτύου, ανθρωπίνου λάθους η κακόβουλης επίθεσης [KoAlDa07]. Προτείνεται έτσι μια καινοτόμα αρχιτεκτονική 2 επιπέδων που αποτελείται από έναν τοπικό server ο οποίος λειτουργεί ως μνήμη ενδιάμεσης αποθήκευσης και ένα σύστημα απομακρυσμένο με πολλαπλούς παρόχους αποθήκευσης. Πλεονεκτήματα της αρχιτεκτονικής αυτής αποτελούν η αποδοτική διάδοση των δεδομένων μεταξύ των διάφορων παρόχων όπως επίσης και ο αποτελεσματικός από άκρη σε άκρη έλεγχος αυτών για πιθανή απώλεια δεδομένων. Τέλος προσφέρει αποθήκευση με κόστος, απόδοση και διαθεσιμότητα συγκρίσιμη με συμβατικά συστήματα αποθήκευσης.

Οι κόμβοι και τα συστήματα αποθήκευσης στα δίκτυα ανίχνευσης θα πρέπει να διατηρούν τα δεδομένα και να δίνουν πρόσβαση σε αυτά ανεξάρτητα από τα προβλήματα που μπορεί να προκύψουν είτε στους εξυπηρετητές (server-side) είτε προς την πλευρά του χρήστη (client-side). Γι αυτό το λόγο έγινε μελέτη και δημιουργήθηκε η PASIS αρχιτεκτονική [GaKhBi01] η οποία επιλύει προβλήματα των εξυπηρετητών, όπως οι συνδέσεις στους servers, κωδικοποιώντας τα δεδομένα με σχήματα κατωφλίου και κάνοντας τους servers πιο ασφαλείς. Τα self-securing συστήματα που προτείνονται σε αυτή τη μελέτη επιλύουν client-side προβλήματα με την παρακολούθηση των προσβάσεων σε αυτά και την επεξεργασία των δεδομένων. Χρησιμοποιούνται p-m-n σχήματα κατωφλίου όπου τα δεδομένα σπάνε σε n μέρη, οποιαδήποτε m από αυτά μπορούν να δημιουργήσουν τα αρχικά δεδομένα και τα p δεν δίνουν κάποια πληροφορία. Η αρχιτεκτονική του PASIS αποτελείται από ένα κατανεμημένο σύστημα αποθήκευσης και κάποιους client-side agents προς την πλευρά του χρήστη που συλλέγουν τα μέρη των δεδομένων που είναι μοιρασμένα στο σύστημα αποθήκευσης και τα συνδυάζει χρησιμοποιώντας τα σχήματα κατωφλίου.

3.4 Αντιμετώπιση προβλημάτων εξαγωγής τοποθεσίας της πηγής των δεδομένων και του mobile sink

Η παρακολούθηση και ανάλυση μηνυμάτων ενός ασύρματου δικτύου από επίδοξους εισβολείς αποτελεί πολύ σημαντικό πρόβλημα καθώς μπορεί να οδηγήσει σε πληθώρα επιθέσεων. Το πρόβλημα αυτό γίνεται ακόμα πιο έντονο ιδιαίτερα όταν ο επίδοξος εισβολέας έχει τη δυνατότητα να παρακολουθήσει όλο το δίκτυο. Μελέτη που έγινε πάνω σε αυτό το θέμα εισήγαγε

την έννοια της ψεύτικης κίνησης μέσα στο δίκτυο με δύο διαφορετικές μεθόδους την proxy-based filtering scheme (PFS) και την tree-based filtering scheme (TFS) [YaShZh08] που όμως είναι δύσκολες και ακριβές ως προς την υλοποίηση. Στην πρώτη μέθοδο επιλέγονται κάποιοι από τους κόμβους ανίχνευσης ως proxy κόμβοι οι οποίοι έχουν την ικανότητα να φιλτράρουν τα πακέτα. Τα μηνύματα στο δίκτυο στέλνονται σε διαστήματα που ακολουθούν την εκθετική κατανομή για να μην ξεχωρίζει η ψεύτικη κίνηση από τα πραγματικά γεγονότα. Όταν ένα ψεύτικο πακέτο φτάσει σε έναν proxy κόμβο τότε απορρίπτεται. Αντιθέτως όταν φτάσει ένα πραγματικό συμβάν τότε επανακρυπτογραφείται και στέλνεται στο σταθμό βάσης για να αποθηκευτεί. Στην περίπτωση που δεν έρθει κάποιο πραγματικό συμβάν τότε στέλνεται ένα κρυπτογραφημένο ψεύτικο μήνυμα. Η δεύτερη μέθοδος έχει ως στόχο να επιλύσει το πρόβλημα του μεγάλου αριθμού proxy κόμβων οργανώνοντας τους σε μία δενδρική μορφή. Έτσι το μήνυμα φιλτράρεται σε πολλαπλούς proxy κόμβους και μειώνεται η κίνηση στο δίκτυο. Φυσικά η μέθοδος αυτή αυξάνει το χρόνο που κάνει ένα πραγματικό γεγονός να φτάσει στο σταθμό βάσης.

Εξίσου σημαντικό πρόβλημα που μπορεί να δημιουργηθεί από την παρακολούθηση της κίνησης του δικτύου από κόμβο σε κόμβο αποτελεί η εξαγωγή της τοποθεσίας της πηγής. Μεγάλο πεδίο έρευνας αποτελούν τα πρωτόκολλα επικοινωνίας των κόμβων μέσα στο δίκτυο και πολλά έχουν δημιουργηθεί για σκοπούς ασφάλειας όπως το phantom routing [KaZhTr05]. Στην μελέτη αυτή γίνεται ανάλυση ήδη υπαρχόντων πρωτοκόλλων όπως το flooding routing, στο οποίο ένας κόμβος που επιθυμεί να μεταδώσει ένα μήνυμα το προωθεί σε όλους τους γείτονες του και αυτοί με τη σειρά τους στους δικούς τους δημιουργώντας μια πλημμύρα μηνυμάτων στο δίκτυο. Εξετάζεται επίσης το single path routing στο οποίο οι κόμβοι προωθούν τα μηνύματα μόνο σε ένα γείτονα, ο οποίος αποφασίζεται μέσω διαφόρων τεχνικών, και το routing with fake sources στο οποίο εισάγονται διάφορες ψεύτικες πηγές που δημιουργούν πλαστή κίνηση στο δίκτυο προκειμένου να μπερδέψουν τον επιτιθέμενο. Οι τεχνικές αυτές είτε είναι εύκολο να γίνουν αντιληπτές από τον εισβολέα είτε καταναλώνουν περίσσια ενέργεια. Έτσι, προτείνεται μία νέα τεχνική το phantom routing, στο οποίο εισάγεται μια πηγή φάντασμα. Υπάρχουν δύο βήματα, 1) η φάση του random walk προς την πηγή-φάντασμα και 2) εφαρμογή του flooding/single-path routing για την μετάδοση του μηνύματος προς τον προορισμό.

Όλες οι παραπάνω τεχνικές ενδέχεται να είναι αρκετά ευάλωτες ειδικά όταν ο εισβολέας μπορεί να παρακολουθήσει όλο το δίκτυο και έχει γνώση αυτού. Για να αποφευχθούν τυχών περιπτώσεις διαρροής πληροφοριών και εξαγωγές τοποθεσίας, υπό αυτές τις συνθήκες, έχουν προταθεί δύο τεχνικές, το periodic collection και το source simulation [MeLiWr07]. Στην πρώτη τεχνική κάθε κόμβος ανεξάρτητα, ανά τακτά χρονικά διαστήματα, στέλνει δεδομένα σε μία λογική συχνότητα ανεξάρτητα αν πρόκειται για πραγματικά γεγονότα η ψεύτικα μηνύματα. Όμως παρόλο την ασφάλεια που προσδίδει στο δίκτυο, καταναλώνει μεγάλα ποσά ενέργειας σε real-time εφαρμογές. Στη δεύτερη τεχνική ένα σύνολο από ψεύτικα κινητά αντικείμενα εισέρχονται και εξομοιώνονται στο δίκτυο σαν να ήταν αληθινά. Με αυτό τον τρόπο ο επιτιθέμενος δεν έχει τη δυνατότητα να ξεχωρίσει τους αληθινούς στόχους από τους ψεύτικους.

Ένα άλλο μεγάλο ζήτημα που αντιμετωπίζουν τα ασύρματα δίκτυα ανίχνευσης αποτελούν οι επιθέσεις από επίδοξους εισβολείς. Τα δίκτυα τέτοιου τύπου αποτελούνται από αισθητήρες ανίχνευσης και σταθμούς βάσης που συλλέγουν τα δεδομένα. Έχουν μελετηθεί, λοιπόν, δύο τύποι επιθέσεων που έχουν ως στόχο τους σταθμούς αυτούς, καθώς επίσης και διάφορες μέθοδοι για την αντιμετώπισή τους [DeHaMi,04]. Η πρώτη οικογένεια επιθέσεων αφορά την απομόνωση και την παρεμπόδιση της επικοινωνίας μεταξύ του σταθμού βάσης και των υπόλοιπων κόμβων η οποία μπορεί να αντιμετωπισθεί με την εγκαθίδρυση πολλαπλών μονοπατιών σε πολλαπλούς σταθμούς βάσης και με τεχνικές όπως one-way hash chains και echo-back algorithm. Οι σταθμοί βάσης παράγουν ένα μήνυμα REQ το οποίο πλημμυρίζει το δίκτυο και λαμβάνεται από τους κοντινότερους κόμβους, οι οποίοι με τη σειρά τους το προωθούν στους κόμβους-παιδιά τους, δημιουργώντας έτσι πολλαπλά μονοπάτια προς τους σταθμούς βάσης. Για να αποφευχθεί η περίπτωση ένας επιτιθέμενος να δημιουργήσει ένα κακόβουλο REQ μήνυμα ακολουθείται η

τεχνική του one way hash chains. Έχοντας μια μη αντιστρέψιμη συνάρτηση F δημιουργούμε μια one way hash chain $K_n,K_{n-1},\ldots,K_0$ όπου $K_{i-1}=F(K_i)$. Κάθε σταθμός βάσης παράγει έναν τυχαίο αριθμό α δημιουργώντας την αλληλουχία $H^{\alpha}=<K_n^{\alpha},K_{n-1}^{\alpha},...K_0^{\alpha}>$ και κάθε κόμβος ανίχνευσης έχει εξαρχής το K_0^{α}. Έτσι σε κάθε REQ που έρχεται μπορεί να επαληθεύει την αλληλουχία των αριθμών. Όσον αφορά την τεχνική του echo-back algorithm, κάθε φορά που ένας κόμβος ανίχνευσης λαμβάνει ένα REQ από έναν άλλον, στέλνει ταυτόχρονα και ένα echo message στον δεύτερο περιμένοντας την απάντησή του. Με αυτόν τον τρόπο μπορεί με ασφάλεια να αναγνωρίσει τους γείτονές του. Η δεύτερη οικογένεια επιθέσεων αφορά την ανάλυση της κίνησης προς έναν κόμβο βάσης ώστε να εξαχθεί η τοποθεσία του. Τεχνικές προκειμένου να αντιμετωπισθούν τέτοιου είδους επιθέσεις αποτελεί το hop by hop encryption/decryption και sending rate control. Στην πρώτη μέθοδο κάθε κόμβος έχει και από ένα cluster key για να μπορεί να κρυπτογραφεί το μήνυμα που στέλνει. Έτσι κάθε φορά που ένα μήνυμα στέλνεται από έναν κόμβο σε έναν άλλο αποκρυπτογραφείται και κρυπτογραφείται ξανά. Στη δεύτερη τεχνική ελέγχεται ο ρυθμός αποστολής κάθε κόμβου καθώς αν είχαμε έναν ομοιόμορφο ρυθμό αποστολής τότε οι κόμβοι ανίχνευσης που είναι πιο κοντά στο σταθμό βάσης θα είχαν μεγαλύτερο ρυθμό κάτι που θα βοηθούσε τον επιτιθέμενο να εξάγει την τοποθεσία του.

ΚΕΦΑΛΑΙΟ 4 -Περιγραφή του προβλήματος

4.1 Ορισμοί και πλαίσιο του προβλήματος

Τα τελευταία χρόνια τα ασύρματα δίκτυα ανίχνευσης (WSN) [InKy15] έχουν χρησιμοποιηθεί σε πληθώρα εφαρμογών και κυρίως στην παρακολούθηση του περιβάλλοντα χώρου. Η ευρεία χρησιμοποίηση τους οδήγησε σε μεγάλη ποσότητα δεδομένων ανίχνευσης καθώς και σε ανάγκη για τεράστιες ποσότητες ενέργειας. Έτσι, αναπτύχθηκε μια νέα γενιά δικτύων τα λεγόμενα data-centric sensor networks (DCSN) [XuXu16]. Στα δίκτυα αυτά, τα δεδομένα που προέρχονται από τους κόμβους ανίχνευσης αποθηκεύονται σε μερικούς κόμβους που είναι προορισμένοι να αποθηκεύουν τα εκάστοτε δεδομένα. Ένας κινητός κόμβος (mobile sink) περνάει περιστασιακά και συλλέγει τα αποθηκευμένα δεδομένα.

Τα δίκτυα αυτά, παρόλο που είναι πιο αποδοτικά, είναι αρκετά ευάλωτα σε διάφορες επιθέσεις εισβολής ή παρεμβολής από διάφορους επίδοξους εισβολείς. Στόχος τους είναι είτε να παραβιάσουν τα δεδομένα κάποιου κόμβου αποθήκευσης είτε να παύσουν τη λειτουργία του προκειμένου τα δεδομένα του να μην είναι διαθέσιμα στο υπόλοιπο δίκτυο. Προκειμένου να μετριαστούν τέτοιου είδους επιθέσεις έχουν προταθεί διάφορες κρυπτογραφικές μέθοδοι ώστε να διασφαλίσουν την ακεραιότητα και την εμπιστευτικότητα των δεδομένων του δικτύου. Ωστόσο, οι τεχνικές αυτές δεν είναι αρκετές για να διασφαλίσουν πλήρως την ασφάλεια τέτοιου είδους δικτύων καθώς επίσης εισάγουν έξτρα πολυπλοκότητα και υπολογιστικό κόστος. Επομένως, έχουν μελετηθεί και έχουν αναπτυχθεί διάφορες άλλες τεχνικές που δεν στηρίζονται σε κρυπτογραφικές μεθόδους και εκμεταλλεύονται αποκλειστικά τα δεδομένα από την ανίχνευση της τοποθεσίας του περιβάλλοντα χώρου.

4.2 Παράταξη Δικτύου

Το δίκτυο που παρατάσσεται στην συγκεκριμένη έρευνα για την ανίχνευση κινητού στόχου στον περιβάλλοντα χώρο αποτελείται από 3 ειδών κόμβους:

1. **Κόμβοι ανίχνευσης (sensor node)**: Το δίκτυο αποτελείται από n_n κόμβους ανίχνευσης και είναι διατεταγμένοι σε ένα επίπεδο χώρο (πρόβλημα 2 διαστάσεων) στις θέσεις $x_1,x_2,\ldots,xn_n$. Ορίζεται λοιπόν το σύνολο $S_n=\{x_i\}_{i\epsilon[1,\ldots,n_n]}$ που αποτελείται από τους κόμβους ανίχνευσης. Όλοι οι κόμβοι είναι πανομοιότυποι καθώς έχουν την ίδια ακτίνα ανίχνευσης r_s καθώς και την ίδια ακτίνα επικοινωνίας r_c. Το μόνο που κάνουν είναι να παρακολουθούν τον περιβάλλοντα χώρο και μόλις ανιχνεύσουν τον κινητό στόχο στέλνουν ένα μήνυμα σε έναν κόμβο αποθήκευσης (ή σε περισσότερους αν αποφασίσουν να το αναπαράγουν για λόγους διαθεσιμότητας). Η διαδικασία της ανίχνευσης έχει ως εξής. Οι κόμβοι μπορούν να ανιχνεύσουν τον κινητό στόχο σε περίπτωση που εκείνος περάσει σε απόσταση $||q(t)-xn_i||\leq r_s$ όπου $q(t)$ είναι η θέση του κινητού την χρονική στιγμή t. Η πληροφορία του μηνύματος που στέλνεται είναι μόνο ότι πέρασε εντός της ακτίνας ανίχνευσης χωρίς να υπάρχει καμία άλλη πληροφόρηση. Η μέτρηση λοιπόν είναι ένα κύκλος ακτίνας r_s και το μόνο που μπορεί

να γνωρίζει κανείς βλέποντας αυτό το μήνυμα είναι ότι το κινητό βρισκόταν κάπου μέσα σε αυτόν τον κύκλο την χρονική στιγμή t. Οι κόμβοι δεν δύνανται να αποθηκεύσουν πληροφορίες λόγω έλλειψης μνήμης και το μόνο που γνωρίζουν επιπλέον είναι οι σχετικές θέσεις των γειτόνων τους.

2. **Κόμβοι αποθήκευσης (storage node)**: Το δίκτυο αποτελείται από n_s κόμβους αποθήκευσης που είναι διατεταγμένοι σε επίπεδο χώρο στις θέσεις $y_1,y_2,...y_{n_s}$ και κάνουμε την υπόθεση ότι το πλήθος των κόμβων είναι κατά πολύ μικρότερο από το πλήθος των κόμβων ανίχνευσης $n_s << n_n$. Ορίζεται έτσι το σύνολο $S_s=\{y_i\}_{i\in[1...ns]}$. Διαθέτουν επαρκή μνήμη,μεγάλα αποθέματα μπαταρίας και είναι υπεύθυνοι να υποδέχονται τα μηνύματα από τους κόμβους ανίχνευσης και να επεξεργάζονται καταλλήλως τις πληροφορίες. Γι αυτό το λόγο παρά τις κρυπτογραφικές μεθόδους που χρησιμοποιούνται στο δίκτυο, εκείνοι θα πρέπει να έχουν πρόσβαση στο μη κρυπτογραφημένο κείμενο. Τέλος, έχουν τη δυνατότητα να κάνουν κάποιου είδους φιλτράρισμα στα πακέτα ώστε να προστατεύονται από κακόβουλες επιθέσεις.

3. **Κινητό αντικείμενο (mobile sink)**: Ανά τακτά χρονικά διαστήματα το κινητό αντικείμενο (μπορεί να είναι και περισσότερα) περνά δίπλα από τους κόμβους αποθήκευσης και μαζεύει τα όποια δεδομένα έχουν. Είναι εξοπλισμένα με ειδικό hardware για να έχουν επαυξημένη προστασία. Κάνουμε την υπόθεση ότι είναι αξιόπιστα και δεν πρόκειται να παραβιαστούν από επίδοξο εισβολέα.

Όλοι οι κόμβοι του δικτύου πέραν του κινητού αντικειμένου (mobile sink) μπορούν να παραβιαστούν και θεωρούνται αναξιόπιστοι. Ο εισβολέας μπορεί να παραβιάσει μέχρι ένα συγκεκριμένο αριθμό κόμβων g το πλήθος και ως εναρκτήριο σημείο σε αυτήν την έρευνα έχουμε το g=1. Οι εισβολείς ενδιαφέρονται αποκλειστικά για την παραβίαση κόμβων αποθήκευσης διότι μέσω αυτών μπορούν να υποκλέψουν πολύ ευαίσθητες πληροφορίες και μυστικά κλειδιά κρυπτογράφησης. Οι εισβολείς δεν έχουν καθολική εικόνα του δικτύου παρά μόνο γνωρίζουν ένα κομμάτι του μέσω των πληροφοριών που έχουν υποκλέψει. Επιπλέον, οι κόμβοι μπορούν εξαιτίας διάφορων παρεμβολών ή παραβιάσεων να μην είναι διαθέσιμοι και να υπάρχει αστοχία στο δίκτυο. Έτσι τα δεδομένα που προορίζονται γι αυτόν είτε απλά χάνονται είτε υποκλέπτονται από εισβολείς. Υπάρχει λοιπόν διαρροή πληροφοριών και έλλειψη δεδομένων που είναι χρήσιμα για την παρακολούθηση του περιβάλλοντα χώρου W.

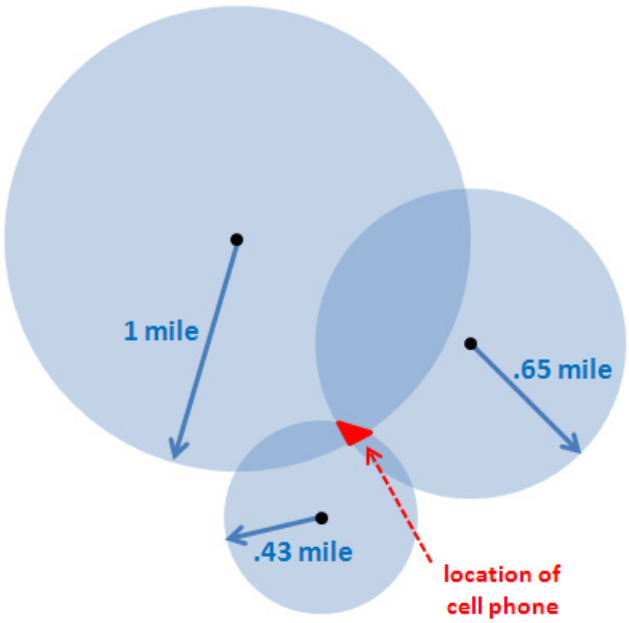
4.3 Καταστάσεις Πληροφορίας (Information states)

4.3.1 Επίπεδο αβεβαιότητας και I-states

Υπάρχουν πολλοί τρόποι για να οριστεί η έννοια της ασφάλειας σε υπολογιστικά συστήματα και δίκτυα. Το πιο σημαντικό που καθορίζει την ιδιωτικότητα σε ένα σύστημα πληροφοριακών συστημάτων είναι τα δεδομένα να είναι προσβάσιμα μόνο σε εκείνους που είναι εξουσιοδοτημένοι να τα διαχειρίζονται και να τα επεξεργάζονται. Όταν όμως αναφερόμαστε σε ασφάλεια που αφορά δεδομένα τοποθεσίας στόχων τα πράγματα διαφέρουν αρκετά. Η έννοια της ασφάλειας σε τέτοιου είδους δεδομένα δεν έχουν να κάνουν μόνο με το ποιος έχει πρόσβαση σε αυτά αλλά και με τι ακρίβεια εντοπίζεται η τοποθεσία του στόχου πχ σε επίπεδο χιλιομέτρων,

μέτρων, εκατοστών και ούτω καθεξής. Στην περίπτωση αυτή το περιεχόμενο των μηνυμάτων είναι πιο σημαντικό από την ποσότητα των μηνυμάτων.

Για να οριστεί και κωδικοποιηθεί η έννοια της ακρίβειας σύμφωνα με την οποία εντοπίζεται η θέση του κινητού στόχου εισάγεται η έννοια των information states (I-states). Η μέθοδος αυτή χρησιμοποιείται τόσο στην ρομποτική όσο και στην κινητή τηλεφωνία. Ο τρόπος με τον οποίο γίνεται ο εντοπισμός των κινητών τηλεφώνων από τις κεραίες, σύμφωνα με την μέθοδο του triangulation [Bo17, Da17], είναι πανομοιότυπος με τον τρόπο που τα I-states εντοπίζουν τον κινητό στόχο για διαφορετικές χρονικές στιγμές. Όπως φαίνεται και στο σχήμα 1, το κινητό τηλέφωνο εντοπίζεται εντός της ακτίνας της κάθε κεραίας. Με αυτόν τον τρόπο μπορεί να εξαχθεί η τοποθεσία του που βρίσκεται στην περιοχή που επικαλύπτονται οι 3 κύκλοι.



Σχήμα 1: I-states [Search Engine Land]

Τα I-states είναι το σύνολο των πιθανών θέσεων στις οποίες μπορεί να βρίσκεται ο κινητός στόχος σε συγκεκριμένες χρονικές στιγμές και οι οποίες είναι σύμφωνες με τις μετρήσεις που έρχονται από τους κόμβους ανίχνευσης. Ένα μεγάλο πλεονέκτημα της μεθόδου αυτής σε σχέση με άλλες, είναι ότι δεν απαιτεί προηγούμενη γνώση της τοποθεσίας του κινητού αντικειμένου και μπορεί να υπολογιστεί από κατάλληλη επεξεργασία των μηνυμάτων που στέλνονται την συγκεκριμένη χρονική στιγμή.

Για να επεξηγηθεί καλύτερα ο τρόπος με τον οποίο γίνεται η ανίχνευση και παρακολούθηση του κινητού στόχου, υποθέτουμε ότι για ένα προηγούμενο χρονικό διάστημα Δt_i οι κόμβοι ανίχνευσης προώθησαν συνολικά m μηνύματα που αφορούσαν το στόχο,

$$\{ (O_1,t_1), (O_2,t_2),\ldots, (O_m,t_m)\}$$

Όπου O_i αντιστοιχεί σε ένα κύκλο που περιέχει την πραγματική κατάσταση του στόχου την χρονική στιγμή t_i. Η θέση του κινητού στόχου τότε q' είναι έγκυρη και συνεπής αν υπάρχει μία συνεχής τροχιά $q: [0,\Delta t_i] \rightarrow W$ τέτοια ώστε:

- $dq/dt \leq v_{max}$ *για κάθε* $t \in [0, \Delta ti]$, όπου v_{max} είναι η μέγιστη ταχύτητα του στόχου.
- $q\ (t_i) \in O_i$ *για κάθε* $i \in [1,m]$.
- $q\ (\Delta ti) = q'$.

Ορίζονται έτσι οι έννοιες I-state (t) και V (t):

- **I-state (t)**: Αποτελεί το σύνολο των θέσεων στόχων που είναι συνεπείς με τα μηνύματα που αναφέρονται σε χρονικά στιγμιότυπα πριν την χρονική στιγμή t
- **V(t)**: Δηλώνει το χώρο που περικλείεται από το I-state τη χρονική στιγμή t.

Ως χώρος ορίζεται το κομμάτι εκείνο που βρίσκεται στην τομή των μηνυμάτων O_i που προέρχονται από τους κόμβους ανίχνευσης (κόκκινη περιοχή στο σχήμα 1). Όσο μεγαλύτερη είναι η περιοχή τότε ο κινητός στόχος μπορεί να βρίσκεται οπουδήποτε σε αυτήν τη μεγαλύτερη περιοχή. Με αυτόν τον τρόπο κωδικοποιείται η ακρίβεια της τοποθεσίας.

4.3.2 Υπολογισμός των I-states

Υποθέτουμε ότι έως τη χρονική στιγμή t το I-state έχει υπολογιστεί και έτσι έχουμε την αρχική του κατάσταση η(0)=W. Από εκεί και πέρα η ανανέωση του I-state εξαρτάται από από το εάν στάλθηκαν η όχι νέα μηνύματα:

- Αν από τη χρονική t_1 έως τη χρονική στιγμή t_2 δεν στάλθηκαν νέα μηνύματα, τότε το νέο I-state $\eta(t_2)$ υπολογίζεται από το προηγούμενο εφαρμόζοντας το άθροισμα minkowski [DeTe15, Kr14] μεταξύ του $\eta(t_1)$ και κύκλου ακτίνας $r=(t_2-t_1)v_{max}$. Για να επεξηγηθεί λίγο καλύτερα αυτή η μέθοδος αρκεί να αναφερθεί ότι το κινητό στο χρονικό διάστημα t_1 έως t_2 θα μπορούσε να είχε κινηθεί προς πάσα κατεύθυνση και να είχε διανύσει απόσταση μικρότερη η ίση του $(t_2-t_1)v_{max.}$ Έτσι με αυτόν τον υπολογισμό επεκτείνεται το προηγούμενο I-state προκειμένου να συμπεριληφθεί η πιθανότητα το κινητό να έχει μετακινηθεί. Αν δεν εφαρμοζόταν το άθροισμα minkowski θα ήταν δυνατό ένα νέο μήνυμα που θα έρθει μια μεταγενέστερη χρονική στιγμή να συμπεριλάμβανε μια τελείως διαφορετική περιοχή που δεν είχε σχέση με το προηγούμενο I-state και έτσι να χανόταν τελικά η τοποθεσία του κινητού στόχου.

- Σε περίπτωση που ένα νέο μήνυμα (O,t) ληφθεί τότε το καινούριο I-state η(t) θα είναι η τομή του O με το παλιό. Με αυτόν τον τρόπο το νέο μήνυμα δίνει την πληροφορία για την καινούρια τοποθεσία του στόχου και γίνεται η σύνδεση με την προγενέστερη κατάσταση.

Ο υπολογισμός αυτός γίνεται στους κόμβους αποθήκευσης οι οποίοι είναι υπεύθυνοι να δέχονται τις πληροφορίες από τους κόμβους ανίχνευσης, να τους αποθηκεύουν και να επεξεργάζονται κατάλληλα τα δεδομένα. Κάθε κόμβος αποθήκευσης δεν επικοινωνεί με τους υπόλοιπους ούτε έχει γνώση των πληροφοριών που διαχειρίζονται εκείνοι. Έτσι κάθε κόμβος αποθήκευσης υπολογίζει το δικό του I-state $\eta_j(t)$. Έτσι αν υπάρχουν n κόμβοι αποθήκευσης τότε θα υπάρχουν και n I-states. Όμως υπάρχει και το "master" I-state η*(t) το οποίο υπολογίζεται από όλα τα μηνύματα όλων των κόμβων αποθήκευσης και ουσιαστικά περιέχει τη γνώση και τις

πληροφορίες όλου του δικτύου. Έτσι στην πραγματικότητα υπάρχουν n+1 I-states στο δίκτυο. Σε κανονικές συνθήκες λειτουργίας του δικτύου το κινητό αντικείμενο (mobile sink) είναι το μόνο που μπορεί να συλλέξει όλες τις πληροφορίες από όλους τους κόμβους αποθήκευσης και να έχει πρόσβαση στο "master" I-state. Μόνο σε περίπτωση παραβίασης η μη διαθεσιμότητας του κόμβου στο δίκτυο δεν μπορεί το κινητό αντικείμενο (mobile sink) να το ανακτήσει. Αν κάποιος επιτιθέμενος καταφέρει και έχει πρόσβαση στο "master" I-state τότε μπορεί να γνωρίζει ανά πάσα στιγμή που βρίσκεται ο κινητός στόχος και αποτελεί την χειρότερη περίπτωση παραβίασης του δικτύου.

4.4 Μετρικές αξιολόγησης

Σε αυτήν την μελέτη ορίζονται 3 κριτήρια για την αξιολόγηση της αποδοτικότητας της προτεινόμενης μεθόδου για την ασφαλή μετάδοση των δεδομένων. Επιπλέον ορίζεται 1 κριτήριο για την σύγκριση και την αξιολόγηση μεταξύ των αλγορίθμων που χρησιμοποιούνται. Τα κριτήρια αυτά είναι τα εξής:

1. **Στιγμιαία Ασφάλεια (Privacy):** Υποθέτοντας ότι ένας εισβολέας έχει παραβιάσει ένα κόμβο αποθήκευσης i και έχει πρόσβαση σε ευαίσθητα δεδομένα, είναι λογικό και σαφές ότι ως μέτρο ασφάλειας θεωρούμε το ποσοστό των δεδομένων που είναι ακόμα ασφαλή. Ο εισβολέας έχει πρόσβαση σε ένα σεβαστό ποσοστό των δεδομένων που ισούται με το κλάσμα $\eta^*(t)/\eta_j(t)$. Το $\eta^*(t)$ που είναι η γνώση όλου του δικτύου είναι λογικό να είναι το πολύ ίση ή μικρότερη σαν μέγεθος από κάθε I-state των κόμβων αποθήκευσης αφού προέρχεται από την τομή τους. Έτσι λαμβάνουμε υπόψιν τη χειρότερη περίπτωση, δηλαδή το $\eta_j(t)$ να είναι το ελάχιστο δυνατό περιέχοντας όσο τον δυνατόν πιο ευαίσθητες πληροφορίες μεγαλώνοντας έτσι και την αναλογία στο κλάσμα. Το κριτήριο ασφάλειας για μια δεδομένη χρονική στιγμή t είναι:

$$P = 1 - \frac{V(\eta^*(t))}{\min_{i \in S_s} V(\eta_i(t))} \qquad \text{Σχέση 4.α}$$

Όσο το $V(\eta_i(t))$ μικραίνει και πλησιάζει το $V(\eta^*(t))$ το κλάσμα τείνει προς το 1 και το P προς το 0 που σημαίνει ότι κάποιος κόμβος αποθήκευσης έχει γνώση πολύ κοντινή στην ολική γνώση του δικτύου. Έτσι το επίπεδο της ασφάλειας χειροτερεύει. Όταν το $V(\eta_i(t))$ μεγαλώνει και απομακρύνεται από το $V(\eta^*(t))$ το κλάσμα γίνεται πολύ μικρό και το P πλησιάζει το 1. Έτσι το επίπεδο της ασφάλειας ανεβαίνει.

2. **Διαθεσιμότητα (Availability)**: Υποθέτοντας ότι ένας εισβολέας έχει παραβιάσει έναν κόμβο αποθήκευσης i σε ένα δίκτυο με n τέτοιους κόμβους, θεωρούμε ως μέτρο διαθεσιμότητας τα δεδομένα που είναι ακόμα διαθέσιμα στο δίκτυο. Όταν ένας κόμβος δεν είναι διαθέσιμος τότε το I-state του δεν είναι διαθέσιμο στο δίκτυο. Ωστόσο, η γνώση που έχει μείνει ακόμα διαθέσιμη είναι η τομή των υπόλοιπων n-1 I-states. Έτσι λαμβάνουμε και πάλι την χειρότερη δυνατή περίπτωση και ορίζουμε ως μέτρο διαθεσιμότητας για δεδομένη χρονική στιγμή t:

$$A = \frac{V(\eta^*(t))}{\max_{i \in S_s} V(\bigcap_{j \in S_s - \{i\}} \eta_j(t))} \qquad \text{Σχέση 4.β}$$

Όσο ο παρανομαστής μεγαλώνει και απομακρύνεται από το V(η*(t)) σημαίνει ότι η τομή των υπόλοιπων κόμβων αποθήκευσης είναι πολύ μεγάλη και στην οριακή περίπτωση που που γίνεται άπειρα μεγάλη και το κλάσμα τείνει προς το 0 τότε έχουμε την χειρότερη περίπτωση της διαθεσιμότητας καθώς όλα τα μηνύματα στέλνονται προς έναν συγκεκριμένο κόμβο που αν παραβιαστεί, ο εισβολέας έχει την ολική γνώση του δικτύου. Στην περίπτωση που ο παρανομαστής μικραίνει και πλησιάζει προς το V(η*(t)) σημαίνει ότι οποιοσδήποτε κόμβος αποθήκευσης και να παραβιαστεί τότε όλοι οι υπόλοιποι έχουν γνώση κοντά στην ολική γνώση του δικτύου και το κλάσμα τείνει προς το 1.

3. **Ενέργεια (Energy)**: Ορίζουμε ως E(i) τον αριθμό των μηνυμάτων που προωθούνται η δημιουργούνται από τον κόμβο ανίχνευσης i στο χρονικό διάστημα μεταξύ t=0 και t=T.

$$E = \frac{1}{T} \sum_{i=1}^{n_n} E(i) \qquad \text{Σχέση 4.γ}$$

Το μοντέλο αυτό είναι επαρκές να αναπαραστήσει και τα μηνύματα που στέλνονται αλλά και τα μηνύματα που λαμβάνονται πολλαπλασιάζοντας το E με έναν συντελεστή α που περικλείει την ενέργεια που καταναλώνεται και από τον κόμβο που στέλνει το μήνυμα αλλά και από τους γείτονες που το λαμβάνουν.

Όσον αφορά το κριτήριο σύμφωνα με το οποίο γίνεται η σύγκριση μεταξύ των αλγορίθμων αποτελεί ο αριθμός των conflicts μετά την εφαρμογή του αλγορίθμου. Ως conflict στην μελέτη αυτή ορίζουμε το γεγονός ότι δύο κόμβοι ανίχνευσης που είναι γειτονικοί στον SPG έχουν το ίδιο label δηλαδή ανήκουν στην ίδια συστάδα και στέλνουν τα δεδομένα τους στην ίδιο κόμβο αποθήκευσης. Όσο μικρότερος είναι ο αριθμός αυτός τόσο καλύτερος είναι ο αλγόριθμος που εφαρμόσαμε.

4.5 Κατανεμημένος αλγόριθμος διάδοσης πληροφοριών

Το βασικό πρόβλημα που έπρεπε αρχικά να επιλυθεί ήταν σε ποιο κόμβο αποθήκευσης θα έστελνε ο κάθε κόμβος ανίχνευσης προκειμένου να μεγιστοποιηθεί η ασφάλεια και η διαθεσιμότητα του δικτύου διατηρώντας την ενέργεια σε χαμηλά επίπεδα. Το πρόβλημα αυτό μπορεί να μοντελοποιηθεί και να αναχθεί σε ένα πρόβλημα χρωματισμού γράφου. Για αυτό το σκοπό ορίζεται μια συνάρτηση ανάθεσης χρωμάτων. Κάθε κόμβος αποθήκευσης αντιστοιχεί σε ένα χρώμα το οποίο είναι μοναδικό και κανένας άλλος κόμβος αποθήκευσης του δικτύου δεν μπορεί να έχει το ίδιο. Το πρόβλημα λοιπόν μετασχηματίζεται στην ανάθεση χρωμάτων στους κόμβους

ανίχνευσης προκειμένου εκείνοι να στέλνουν τα δεδομένα τους με τρόπο που να εξασφαλίζει την ασφάλεια το δίκτυο. Ορίζεται λοιπόν η συνάρτηση ανάθεσης χρώματος C η οποία αντιστοιχεί τον κάθε κόμβο ανίχνευσης σε έναν η περισσότερος κόμβους αποθήκευσης:

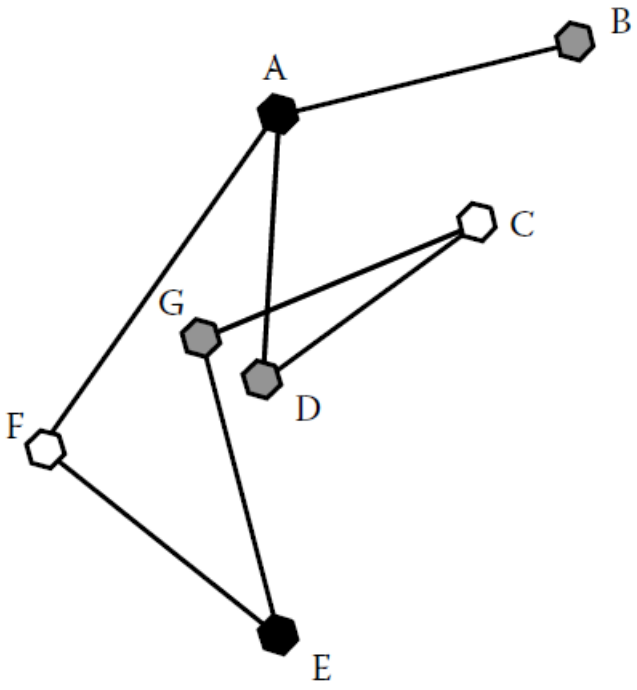
$$C: S_n \rightarrow 2^{Ss} \qquad \text{Σχέση 4.δ}$$

Όπου το 2^{Ss} είναι το δυναμοσύνολο του S_s. Το πρόβλημα λοιπόν είναι να βρεθεί η συνάρτηση αυτή που μεγιστοποιεί την μετρική της ασφάλειας και διαθεσιμότητα του δικτύου αλλά ταυτόχρονα ελαχιστοποιεί την μετρική της ενέργειας. Τα 3 κριτήρια αξιολόγησης που προαναφέρθηκαν βρίσκονται σε σύγκρουση μεταξύ τους γι αυτό και θα πρέπει να βρεθεί ένα σημείο ισορροπίας μεταξύ τους.

4.5.1 Spatial Privacy Graph (Γράφος χωρικής ιδιωτικότητας)

Όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο το περιεχόμενο το μηνυμάτων είναι πολύ πιο σημαντικό από το πλήθος τους και το πλήθος των κόμβων που τα παράγουν, όσον αφορά τη συγκεκριμένη μελέτη. Για παράδειγμα ο συνδυασμός (τομή) των μηνυμάτων που προέρχονται από 2 μόλις κόμβους ανίχνευσης μπορεί να είναι πολύ πιο σημαντικός από τον συνδυασμό μηνυμάτων από 4 κόμβους. Έτσι, είναι πολύ πιο σημαντικό οι 2 κόμβοι να στείλουν τα δεδομένα τους σε ξεχωριστούς κόμβους αποθήκευσης ενώ είναι σχετικά ανώδυνο για τους 4 κόμβους να στείλουν τα δικά τους στον ίδιο κόμβο αποθήκευσης. Η τοπολογική τους θέση στο δίκτυο και η σχετική απόσταση μεταξύ τους είναι αυτή που κρίνει το πόσο ευαίσθητα δεδομένα μπορούν να παράγουν σε σχέση με τους γειτονικούς τους κόμβους. Γι αυτόν ακριβώς τον σκοπό κατασκευάστηκε ο SPG (γράφος χωρικής ιδιωτικότητας) προκειμένου να αναγνωρίζει αυτά τα ζευγάρια κόμβων ανίχνευσης τα οποία μπορούν να εντοπίσουν τον κινητό στόχο με αρκετά μεγάλη ακρίβεια.

Ένα σύνολο από κόμβους ανίχνευσης S σχηματίζει SPG $G_p= (S,E_p)$ όπου ένα ζευγάρι (x_i,x_j) συνδέεται στον γράφο με μια ακμή e_{ij}, αν και μόνο αν σχηματίζουν ένα ζευγάρι ασφάλειας. Δοθέντος ενός βαθμωτού μεγέθους α που λογίζεται ως παράγοντας ασφάλειας, ένα ζευγάρι ασφάλειας αποτελείται από ένα ζευγάρι κόμβων ανίχνευσης που απέχουν μεταξύ τους απόσταση $d \in [2r_s-\alpha,2r_s]$. Διαισθητικά αυτά είναι τα ζεύγη που δημιουργούν ευαίσθητα δεδομένα.



Σχήμα 2: Spatial Privacy Graph[XuXu16]

Όπως φαίνεται και στο σχήμα 2 κάποιοι κόμβοι που είναι είτε πολύ κοντινοί μεταξύ τους είτε μακρινοί δεν συνδέονται με ακμή στον SPG παρά μόνο εκείνοι που απέχουν μεταξύ τους απόσταση μέσα στο εύρος που ορίστηκε προηγουμένως.

4.5.2 Αλγόριθμος walkthrough

Ο SPG το μόνο που κάνει στο δίκτυο είναι να αναγνωρίζει εκείνα τα ζευγάρια κόμβων που αποτελούν ζευγάρι ασφάλειας και πρέπει να στείλουν τα δεδομένα τους σε ξεχωριστούς κόμβους αποθήκευσης. Για να αποφασίσει όμως ο κάθε ένας σε ποιον κόμβο τελικά θα στείλει, έχει προταθεί ένας κατανεμημένος αλγόριθμος ο οποίος εκτελείται από κάθε κόμβο ανίχνευσης ξεχωριστά. Δοθέντος ενός γράφου SPG $G_p= (S,E_p)$, η έξοδος του κατανεμημένου αλγορίθμου είναι ένας χρωματιστός γράφος $G_c= (S,E_p,C)$. Ουσιαστικά αυτό που κάνει ο αλγόριθμος είναι να αναθέτει ένα χρώμα σε κάθε έναν κόμβο ανίχνευσης σύμφωνα με την συνάρτηση ανάθεσης χρωμάτων C που έχει οριστεί όπου $C=\{c_{x_i}|c_{x_i}=C(x_i)\}_{\forall x_i \in S}$. Ό νέος αυτός χρωματιστός γράφος πρέπει να ικανοποιεί τις εξής δύο απαιτήσεις:

- **Έγκυρος**: Για κάθε ακμή του γράφου $e_{ij} \in E_P$, οι κόμβοι της x_i και x_j θα πρέπει να έχουν

διαφορετικά χρώματα $\mathbf{c}_{x_i} <> \mathbf{c}_{x_j}$.

- **Επιτρεπτός**: Το χρώμα κάθε κόμβου ανίχνευσης θα πρέπει να είναι ένα από τα χρώματα των κόμβων αποθήκευσης.

Ο επιτρεπτός και έγκυρος γράφος αποτελεί τον βασικό στόχο του κατανεμημένου αυτού αλγορίθμου. Ωστόσο, για οποιοδήποτε δοθέν πλήθος κόμβων ανίχνευσης και δοθέν πλήθος κόμβων αποθήκευσης δεν είναι πάντα δυνατό να έχουμε ως έξοδο έναν τέτοιο γράφο. Συνεπώς αυτό το πρόβλημα πολλές φορές δεν έχει λύση. Γι αυτό το λόγο ο αλγόριθμος θα κοιτάει πρώτα να είναι έγκυρος ο γράφος και μετά επιτρεπτός.

Ο αλγόριθμος έχει ως εξής. Αρχικά κάθε κόμβος αποθήκευσης αντιστοιχίζεται με ένα μοναδικό χρώμα από το 1 έως το n_s. Έπειτα, κάθε κόμβος ανίχνευσης αναθέτει το χρώμα μόνο του, εκτελώντας τον κατανεμημένο αλγόριθμο, στηριγμένο στα χρώματα των γειτόνων του. Ως γείτονες θεωρούνται οι κόμβοι εκείνοι που συνδέονται άμεσα μαζί του στον SPG. Ο κάθε κόμβος ξεκινά με ένα ανέφικτο χρώμα για παράδειγμα προσθέτοντας το ID του στο n_s, που είναι το πλήθος των κόμβων αποθήκευσης. Με αυτόν τον τρόπο όλοι οι κόμβοι, πριν ξεκινήσει ο αλγόριθμος, έχουν ανατεθεί σε ανύπαρκτα χρώματα δηλαδή χρώματα που δεν αντιστοιχούν σε κόμβους αποθήκευσης. Ο αλγόριθμος σταματά όταν κανένα χρώμα δεν ανανεώνεται μεταξύ δύο διαδοχικών επαναλήψεων.

Αρχικά σε κάθε επανάληψη, ο κόμβος x_i στέλνει ένα μήνυμα (I_{x_j}, c_{x_j}) σε κάθε γείτονα στον SPG ανακοινώνοντας το τωρινό του χρώμα c_{x_j} και το ID του, που συμβολίζεται με I_{x_j}. Ενώ την ίδια στιγμή δέχεται μηνύματα από τους γείτονες του μαθαίνοντας τα δικά τους χρώματα $\{c_{xi}\}_{x_i \in Nbr}$. Σε κάθε επανάληψη μόνο ένας κόμβος ανίχνευσης που ικανοποιεί τις ακόλουθες συνθήκες επιτρέπεται να ανανεώσει το χρώμα του:

1. Δεν του έχει ανατεθεί ένα επιτρεπτό χρώμα
2. Το ID του χρώματός του είναι μεγαλύτερο από το ID των χρωμάτων όλων των γειτόνων του

Είναι πολύ σημαντικό να επεξηγηθεί ο τρόπος με τον οποίο γίνεται η ανανέωση των χρωμάτων στους κόμβους. Ορίζεται η συνάρτηση ΑνανέωσηΧρωμάτων () η οποία ψάχνει να βρει ένα χρώμα το οποίο ικανοποιεί τις παρακάτω συνθήκες:

1. **Επιτρεπτό**: Το νέο χρώμα θα πρέπει να είναι χρώμα κάποιου κόμβου αποθήκευσης, $c'_{x_j} \in \{1,\ldots..,n_s\}$.
2. **Έγκυρο**: Κανένας από τους γειτονικούς κόμβους στον SPG δεν έχει επιλέξει αυτό το χρώμα $c'_{x_j} <> \{c_{x_j}\}_{x_j \in Nbr}$.
3. Κοντινότερος: Μεταξύ όλων των έγκυρων και επιτρεπτών χρωμάτων, επιλέγεται ο κόμβος αποθήκευσης που απέχει τα λιγότερα hop counts από τον συγκεκριμένο κόμβο ανίχνευσης.

Όπως προ ειπώθηκε δεν υπάρχει πάντα λύση στο πρόβλημα όταν ο αριθμός των κόμβων αποθήκευσης και ο αριθμός των κόμβων ανίχνευσης είναι δοσμένος. Έτσι και ο αλγόριθμος είναι σαφές ότι δεν μπορεί πάντα να έχει ως έξοδο έναν γράφο που είναι και έγκυρος και επιτρεπτός. Στις

περιπτώσεις αυτές, η συνάρτηση Ανανέωση Χρωμάτων (), σε οποιονδήποτε κόμβο ανίχνευσης δεν μπορεί να χρωματίσει σύμφωνα με τους όρους που έχουμε θέσει, επιστρέφει ως έξοδο το $-|c_{x_i}|$, δηλαδή τον αντίθετο του αρχικού του χρώματος. Στο τέλος της διαδικασίας οι κόμβοι ανίχνευσης που δεν έχουν ένα έγκυρο και επιτρεπτό χρώμα, επιλέγουν τυχαία ένα από τα επιτρεπτά χρώματα των κόμβων αποθήκευσης ανεξάρτητα από το χρώμα των γειτόνων τους. Ο αλγόριθμος σταματά όταν κανένας κόμβος ανίχνευσης δεν μπορεί να ανανεώσει περαιτέρω το χρώμα του.

Επιπλέον ισχύει το ακόλουθο λήμμα: Ο αλγόριθμος τερματίζει μετά από |S| επαναλήψεις με έναν έγκυρο (αλλά όχι απαραίτητα επιτρεπτό) χρωματισμένο γράφο G_c (S,E_p,C), όπου |S| είναι το πλήθος των κόμβων ανίχνευσης.

Παρακάτω ακολουθεί ο ψευδοκώδικας του αλγορίθμου walkthrough:

```
Κατανεμημένος αλγόριθμος (walkthrough)
Απαίτηση: Είσοδος:
          Nbr: Σύνολο γειτόνων
       I_o: ID του κόμβου ανίχνευσης
       Διαδικασίες:
      1: c_o = I_o+n_s;
      2: επανάλαβε:
      3:    Ανακοίνωσε (I_o, c_o);
      4:    {c_xi}_xi∈Nbr = ΛήψηΑνακοίνωσης ();
      5:    Αν c_o > n_s και c_o > max{c_xi}_xi∈Nbr τότε:
      6:        c_o = ΑνανέωσηΧρωμάτων ({c_xi}_xi∈Nbr);
      7:    end if
      8: until ΔενυπάρχειΑλλαγή (co) and
         ΔενυπάρχειΑλλαγή ({c_xi}_xi∈Nbr)
```

Τέλος για λόγους συγχρονισμού και προς επίλυση ζητημάτων που αναδύονται από κατανεμημένους αλγορίθμους είναι σημαντικό να εξασφαλιστεί ότι κάθε κόμβος ανίχνευσης αποφασίζει αν θα αλλάξει το χρώμα του αφού ολοκληρωθεί η διαδικασία των ανακοινώσεων.

4.6 Αναπαραγωγή μηνυμάτων

Στον αλγόριθμο που περιγράφηκε στην προηγούμενη ενότητα, κυρίαρχα θέματα αποτελούσαν η ασφάλεια και η κατανάλωση ενέργειας του δικτύου. Ωστόσο, η διαθεσιμότητα των κόμβων αποθήκευσης ανά πάσα στιγμή είναι εξίσου πολύ σημαντική. Σε κανονικές συνθήκες λειτουργίας του δικτύου, το κινητό αντικείμενο (mobile sink) είναι σε θέση να συλλέξει τα δεδομένα όλων των κόμβων και να εξάγει το η*(t), το οποίο είναι η ολική γνώση του δικτύου. Παρόλα αυτά σε πολλές περιπτώσεις, είτε λόγω αστοχίας υλικού είτε λόγω εισβολής, ορισμένοι κόμβοι δεν είναι σε θέση να στείλουν τα δεδομένα τους έχοντας έτσι διαρροή και έλλειψη πληροφοριών. Γι αυτόν ακριβώς τον λόγο είναι σημαντικό ο συνδυασμός (τομή) των δεδομένων των εναπομεινάντων κόμβων αποθήκευσης να μας δίνει γνώση που να είναι αρκετά κοντά στην ολική γνώση η*(t). Είναι αναγκαίο να εφαρμοστεί μια τεχνική που θα μεγιστοποιούσε την διαθεσιμότητα των δεδομένων χωρίς όμως να οδηγούσε σε μεγάλη αύξηση κατανάλωσης ενέργειας. Ένας λογικός τρόπος θα ήταν να έχουμε αναπαραγωγή (replication) των μηνυμάτων, που

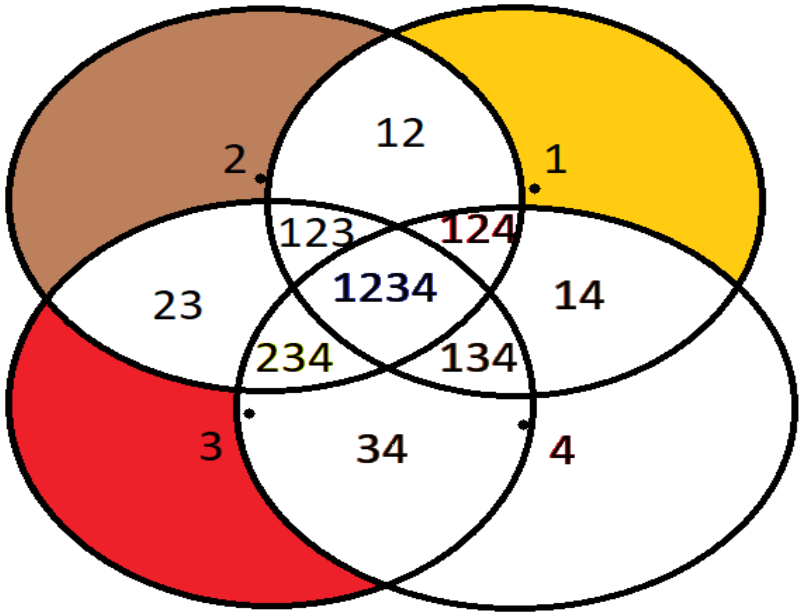
στέλνουν οι κόμβοι ανίχνευσης, δηλαδή να στέλνουν ένα αντίγραφο σε κάποιον άλλον κόμβο αποθήκευσης από αυτόν που είναι προορισμένοι να στείλουν.

Η διαδικασία λοιπόν προς επίτευξη αυτού του στόχου και υλοποίησης της συγκεκριμένη τεχνικής έχει ως εξής. Μόνο οι κόμβοι που αποτελούν ζευγάρι ασφάλειας (privacy pair) έχουν τη δυνατότητα να αναπαράγουν τα μηνύματα τους. Ο λόγος είναι ότι οι κόμβοι που δεν αποτελούν ζευγάρι ασφάλειας είναι τοποθετημένοι ανάμεσα από τους κόμβους των ζευγαριών ασφάλειας και επομένως η τομή των μηνυμάτων τους είναι μεγαλύτερη και χωρίς ευαίσθητη πληροφορία. Επιπλέον, η ενέργεια δαπανάται στα πραγματικά σημαντικά μηνύματα. Επιπροσθέτως, προκειμένου να διατηρηθεί μια ισορροπία ανάμεσα στην διαθεσιμότητα και την ασφάλεια εισάγουμε τον παράγοντα του πιθανοτικού διπλασιασμού p. Κάθε κόμβος που είναι μέρος ενός ζευγαριού ασφάλειας αναπαράγει το μήνυμα του με μία πιθανότητα p. Όταν λοιπόν ο κόμβος ανιχνεύσει τον κινητό στόχο, παράγει έναν τυχαίο αριθμό στο διάστημα [0,1]. Μόνο αν ο αριθμός αυτός είναι μικρότερος από το p θα αναπαραχθεί το μήνυμα και θα σταλθεί σε έναν δεύτερο κόμβο αποθήκευσης. Αν το p=0 τότε έχουμε τέλεια ασφάλεια και μηδενική διαθεσιμότητα καθώς κανένα μήνυμα δεν θα αναπαραχθεί. Αν p=1 τότε έχουμε την τέλεια διαθεσιμότητα καθώς όλα τα μηνύματα θα αναπαραχθούν αλλά θα έχουμε χαμηλότερα επίπεδα ασφάλειας.

Για να αποφύγουμε το γεγονός ότι συγκεκριμένοι κόμβοι ανίχνευσης θα στέλνουν σε συγκεκριμένους κόμβους αποθήκευσης, τα ζευγάρια ασφάλειας κάθε φορά που ανιχνεύουν τον κινητό στόχο διαλέγουν τυχαία τους δεύτερους κόμβους ανίχνευσης στους οποίους θα στείλουν το αντίγραφο των δεδομένων τους.

4.7 Γενική περιγραφή του προβλήματος

Το βασικό πρόβλημα που πραγματεύεται σε αυτή εδώ η μελέτη είναι ο ορισμός μιας πολιτικής μετάδοσης δεδομένων ανάμεσα στους κόμβους ανίχνευσης και κόμβους αποθήκευσης. Πρέπει να οριστεί με σαφήνεια η διαδικασία με την οποία οι διάφοροι κόμβοι, που ανιχνεύουν τον κινητό στόχο, θα αποφασίζουν σε ποιον κόμβο αποθήκευσης θα στείλουν τα δεδομένα τους. Ο στόχος της παρούσας διπλωματικής είναι διπλός. Ο πρώτος αποτελεί την επιλογή του ελάχιστου αριθμού των κόμβων αποθήκευσης ώστε να έχουμε μηδενικό αριθμό conflicts στο δίκτυο, δηλαδή να προσεγγιστεί ο χρωματικός αριθμός [WolframMathworld]. Ο δεύτερος αποτελεί την επίλυση του εξής προβλήματος. Ο ελάχιστος αριθμός conflicts ανεξάρτητα από τον αριθμό των κόμβων αποθήκευσης ούτως ώστε να μεγιστοποιηθεί η μετρική της ασφάλειας (Privacy) διατηρώντας τις άλλες δύο μετρικές, διαθεσιμότητα (availability) και ενέργεια (energy) σε ικανοποιητικά επίπεδα. Η στόχευση της διπλωματικής αποτελεί το κριτήριο της ιδιωτικότητας που είναι το πλέον σημαντικό για την προστασία των δεδομένων έναντι των άλλων δύο. Παρακάτω στο σχήμα 3 δίνεται ένα παράδειγμα των προβληματικών περιοχών που δημιουργούνται από γειτονικούς κόμβους που στέλνουν τα δεδομένα τους σε ίδιους κόμβους:



Σχήμα 3: Υλοποίηση κλασσικής θεωρίας χρωματισμού γράφου

Υποθέτοντας ότι στο δίκτυο υπάρχουν 3 κόμβοι αποθήκευσης και άρα 3 διαθέσιμα χρώματα, τα οποία επισυνάπτονται στους κόμβους 1,2,3 αντίστοιχα, δεν υπάρχει ελεύθερο χρώμα για τον κόμβο 4. Έτσι, ο κόμβος αυτός είναι αναγκασμένος να επιλέξει ένα από τα χρώματα που ήδη χρησιμοποιούνται. Είτε επιλέξει το κόκκινο, κίτρινο η καφέ χρώμα σχηματίζονται οι εξής επικίνδυνες περιοχές α (34,234,134,1234), β (14,134,124,1234),γ (124,234,1234) που δημιουργούν πρόβλημα ιδιωτικότητας στο δίκτυο

4.8 Μαθηματική περιγραφή του προβλήματος

Έστω ότι έχουμε m το πλήθος κόμβους ανίχνευσης, n το πλήθος κόμβους αποθήκευσης και ένα κινητό στόχο l ο οποίος μετακινείται εντός δικτύου. Επιπλέον, κακόβουλος εισβολέας έχει τη δυνατότητα να υποκλέψει έως και g=1 κόμβους αποθήκευσης αποκτώντας την πρόσβαση σε κλειδιά κρυπτογράφησης και ευαίσθητα δεδομένα. Στόχος είναι ο ορισμός μιας συνάρτησης Χρωματισμού C: $S_m \rightarrow 2^{Sn}$ τέτοια ώστε να ελαχιστοποιεί τα conflicts που δημιουργούνται ανάμεσα στους κόμβους ανίχνευσης του δικτύου μεγιστοποιώντας ταυτόχρονα την μετρική privacy του δικτύου.

Στον πρώτο στόχο της διπλωματικής ζητούμενο είναι να βρεθεί το $\min_k$, όπου k ο αριθμός των clusters στο δίκτυο τέτοιο ώστε:

$$\sum_{i,j} conflicts_{ij} = 0$$ Σχήμα 4.ε

όπου τα conflicts ορίζονται ως:

$$conflicts_{ij} = \begin{cases} 1, C(x_i) = C(x_j) \\ 0, C(x_i) \neq C(x_j) \end{cases}$$ Σχήμα 4.στ

Όσον αφορά το δεύτερο στόχο ζητούμενο είναι να βρούμε για κάθε k, ακόμα και για εκείνα που είναι μικρότερα από το min_k, τον ελάχιστο αριθμό των conflicts που δημιουργούνται στο δίκτυο:

$$\min_k \left\{ \sum_{i,j} conflicts_{ij} \right\}, \forall x_{i,j} \in Cluster_k$$ Σχήμα 4.ζ

4.9 Όρια Βέλτιστων Λύσεων

Τέτοιου είδους προβλήματα χρωματισμού γράφου δεν είναι πάντοτε επιλύσιμα και υπόκεινται σε ορισμένα όρια. Αν έχουμε ένα δοσμένο αριθμό κόμβων ενός γράφου n αλλά και δοσμένο αριθμό χρωμάτων k τότε το πρόβλημα της ανάθεσης χρωμάτων στους διάφορους κόμβους του γράφου δεν έχει πάντοτε λύση. Γι αυτό λοιπόν και στη συγκεκριμένη μελέτη δεν είναι πάντα επιτρεπτό να διαχωρίζουμε τους κόμβους ανίχνευσης σε ένα επιθυμητό αριθμό ομάδων χωρίς να υπάρχουν ούτε δύο γειτονικοί κόμβοι που να ανήκουν στην ίδια ομάδα. Γι αυτό σκοπός είναι να προσεγγίσουμε τον χρωματικό αριθμό και να φτάσουμε με τους υπάρχοντες αλγορίθμους συσταδοποίησης στο μικρότερο δυνατό αριθμό ομάδων χωρίς να υπάρχει κανένα conflict.

4.9.1 Χρωματικός αριθμός

Ο χρωματικός αριθμός ενός γράφου G είναι ο μικρότερος αριθμός χρωμάτων που απαιτούνται για να χρωματίσουμε τον γράφο αυτό, με τέτοιο τρόπο ώστε να μην υπάρχει ούτε ένα ζεύγος γειτονικών κόμβων που να έχουν το ίδιο χρώμα μεταξύ τους [WolframMathworld]. Αυτό σημαίνει ότι είναι ο ελάχιστος δυνατός αριθμός k για να επιλυθεί το πρόβλημα χρωματισμού με k χρώματα ώστε κανένας κόμβος να μην έχει το ίδιο χρώμα με κάποιο γειτονικό του. Τις περισσότερες φορές στη βιβλιογραφία συμβολίζεται είτε ως χ(G) είτε ως γ(G). Οι γράφοι που δεν έχουν καμία ακμή (empty graphs) έχουν χρωματικό αριθμό 1 ενώ εκείνοι που μπορούν να χωριστούν σε δύο ομάδες τέτοιες ώστε κάθε κόμβος να μην είναι γειτονικός με κάποιον κόμβο της ίδιας ομάδας (bipartite graphs) έχουν χρωματικό αριθμό 2.

Το χρωματικό πολυώνυμο $\pi_G(z)$ αποτελεί ένα πολυώνυμο που κωδικοποιεί το πλήθος των διαφορετικών τρόπων με τους οποίους μπορούμε να χρωματίσουμε τους κόμβους ενός γράφου G. Ο χρωματικός αριθμός είναι ο μικρότερος θετικός αριθμός z τέτοιος ώστε για το χρωματικό πολυώνυμο να ισχύει $\pi_G(z)>0$. Ο υπολογισμός του πολυωνύμου αυτού αποτελεί NP-complete πρόβλημα καθώς δεν υπάρχει κάποιος εύκολος και φορμαλιστικός τρόπος να εξαχθεί. Έχουν διεξαχθεί αρκετές μελέτες στο ερευνητικό αυτό πεδίο και έχουν προταθεί αρκετοί αλγόριθμοι όπως

ο column generation algorithm [MeTr95] που μπορεί να επιλύσει τους περισσότερους μικρούς και μεσαίου μεγέθους γράφους. Ο χρωματικός αριθμός πρέπει να είναι μεγαλύτερος η ίσος του αριθμού clique [BrJa13, Sz15]. Όσον αφορά του τέλειους γράφους (perfect graphs), για κάθε υπογράφο του ο χρωματικός αριθμός ισούται με το μεγαλύτερο αριθμό των κατά ζεύγη γειτονικών κόμβων του υπογράφου αυτού. Ένας γράφος, που ο αριθμός clique είναι ίσος με τον χρωματικό, ονομάζεται weakly perfect. Ένας γράφος με χρωματικό αριθμό ≤ 2 ονομάζεται bicolorable (δίχρωμο) ενώ εκείνος με χρωματικό αριθμό ≤ 3 ονομάζεται threecolorable (τριών χρωμάτων). Γενικά ένας γράφος με χρωματικό αριθμό ίσον με k ονομάζεται k-chromatic ενώ ένας με χρωματικό αριθμό ≤ k ονομάζεται k-colorable.

4.9.2 Θεώρημα Brooks

Το θεώρημα του Brooks υποδηλώνει μία σχέση μεταξύ του μέγιστου βαθμού ενός γράφου και του χρωματικού του αριθμού [Sc16,Si15]. Ένας μη κατευθυνόμενος γράφος λέγεται συνεκτικός (connected) αν για κάθε ζευγάρι κορυφών υπάρχει διαδρομή που τις συνδέει. Σύμφωνα με αυτό το θεώρημα, σε έναν συνεκτικό γράφο [Wi12] στον οποίο κάθε κόμβος έχει το πολύ Δ γείτονες, οι κόμβοι μπορούν να χρωματιστούν με μόνο Δ χρώματα, εκτός από 2 περιπτώσεις, οι πλήρεις γράφοι και οι εκείνοι με κύκλους περιττού μήκους, απαιτούν Δ+1 χρώματα. Το θεώρημα ονομάστηκε από τον R.Leonard Brooks ο οποίος δημοσίευσε την απόδειξη. Χρωματισμός με εφαρμογή του συγκεκριμένου θεωρήματος ονομάζεται Δ-χρωματισμός.

Απόδειξη:

Σε έναν biconnected graph [Wi12, St10], οι συνεκτικές συνιστώσες του μπορούν να χρωματιστούν ξεχωριστά και αργότερα να συνδυαστούν οι χρωματισμοί. Εάν ο γράφος έχει έναν κόμβο v με βαθμό μικρότερο του Δ, τότε ένας άπληστος αλγόριθμος που χρωματίζει τους πιο απομακρυσμένους κόμβους από τον v πριν από τους πιο κοντινούς, χρησιμοποιεί το πολύ Δ χρώματα. Το πιο δύσκολο κομμάτι της απόδειξης αποτελεί οι συνεκτικοί Δ-regular γράφοι [Wi12, St10], δηλαδή εκείνοι που οι κόμβοι τους έχουν ίδιο αριθμό γειτόνων ίσο με Δ, και ειδικότερα εκείνοι που έχουν Δ ≥ 3. Και σε αυτήν την περίπτωση μπορεί να αποδειχθεί το θεώρημα Brooks δημιουργώντας ένα spanning tree τέτοιο ώστε δύο μη κοντινοί γείτονες u και w της ρίζας v να είναι φύλλα του δέντρου. Με έναν άπληστο αλγόριθμο ο οποίος ξεκινάει από τους u και w και επεξεργάζεται τους εναπομείναντες κόμβους του spanning tree με bottom-up (από πάνω προς τα κάτω) λογική τελειώνοντας στον v, χρησιμοποιεί το πολύ Δ χρώματα. Για κάθε κόμβο που είναι χρωματισμένος, έχει έναν αχρωμάτιστο πατέρα, έτσι οι ήδη χρωματισμένοι του γείτονες δεν μπορούν να χρησιμοποιήσουν όλα τα εναπομείναντα χρώματα, ενώ στον κόμβο v οι δύο γείτονες u και w έχουν ίδια χρώματα άρα και πάλι μόνο ένα ελεύθερο χρώμα παραμένει για τον v.

4.9.3 Όριο του Hoffman

Το θεώρημα του Hoffman [SFU] για τον χρωματικό αριθμό έχει ως εξής: Υποθέτουμε ότι έχουμε γράφο G= (V,E) όπου V={1,2,3...n} και E το σύνολο των ακμών του. Έχουμε έναν μη αρνητικό, συμμετρικό nxn πίνακα W ο οποίος είναι διάφορος του μηδενός και έχει μηδενικά μόνο στις θέσεις στις οποίες δεν υπάρχει ακμή. Επιπλέον $\lambda_1,\lambda_2,...,\lambda_n$ αποτελούν οι ιδιοτιμές του πίνακα W. Αποδεικνύεται για τον χρωματικό αριθμό ότι ισχύει το εξής:

$$\chi_v(G) \geq 1 - \frac{\lambda_n}{\lambda_1}.$$ Σχήμα 4.η

4.9.4 Lovasz theta συνάρτηση

Μια ορθοκανονική αναπαράσταση ενός γράφου G= (V,E) είναι ένα σύνολο από μοναδιαία διανύσματα κόμβων(v_1, ..., v_n), τέτοια ώστε κάθε ζεύγος κόμβων να είναι γειτονικά ή κάθετα μεταξύ τους. Καλούμε δύο κόμβους i και j "παρόμοιους" αν i = j ή αν i, j ∈ E. Τότε (v_1, ..., v_n) είναι μια ορθοκανονική αναπαράσταση του G αν $<v_i , v_j >=0$ για κάθε ζεύγος i,j που δεν είναι "παρόμοια" μεταξύ τους. Όλοι οι γράφοι έχουν ορθοκανονική αναπαράσταση, από τη στιγμή που μπορούμε να πάρουμε ένα σύνολο από μοναδιαία διανύσματα του G σε αυτή τη μορφή.

Ορισμός: Ας υποθέσουμε ότι (u_1, … , u_n) αποτελεί μια ορθοκανονική αναπαράσταση και C είναι σύνολο όλων των μοναδιαίων διανυσμάτων στο R_n. Η τιμή αυτής της αναπαράστασης ορίζεται ως:

$$\min_{c \in C} \max_{1 \le i \le n} \frac{1}{\langle c, u_i \rangle^2}$$ Σχέση 4.θ

Η Lovasz theta συνάρτηση ϑ(G) [Ri13] ορίζεται ως η μικρότερη τιμή όλων των πιθανών τιμών των ορθοκανονικών αναπαραστάσεων του γράφου G.

Σύμφωνα με το "θεώρημα sandwich" του Lovasz, ο αριθμός Lovasz κυμαίνεται μεταξύ δύο άλλων των οποίων ο υπολογισμός αποτελεί NP-complete πρόβλημα. Πιο συγκεκριμένα

$$\omega (G) \le \vartheta (G) \le \chi (G)$$ Σχέση 4.ι

όπου ω(G) αποτελεί τον αριθμό clique [BrJa13, Sz15] ενός γράφου G και χ(G) αποτελεί ο χρωματικός του αριθμός.

4.9.5 Κλασματικός χρωματισμός

Ο κλασματικός χρωματισμός [Wolfram] αποτελεί ένα πρωτοποριακό θέμα στην θεωρία γράφων γνωστός ως κλασματική θεωρία γράφων. Αποτελεί μια γενίκευση του κλασικού χρωματισμού γράφων. Στην κλασσική θεωρία χρωματισμού κάθε κόμβος αντιστοιχίζεται με ένα και μόνο χρώμα και οι γείτονες του επίσης θα πρέπει να επισυναφθούν με ένα μοναδικό χρώμα, διαφορετικό από εκείνον. Στον κλασματικό χρωματισμό όμως, ένα σύνολο χρωμάτων επισυνάπτεται σε κάθε κόμβο του γράφου. Επίσης ο περιορισμός ανάμεσα στα διαφορετικά χρώματα των γειτόνων ισχύει και σε αυτήν την περίπτωση. Ο κλασματικός χρωματικός αριθμός ενός γράφου G συμβολίζεται ως χ_f (G) και έχει τις εξής ιδιότητες:

$$\chi_f(G) \ge n(G)/\alpha(G)$$ Σχέση 4.κ

$$\omega(G) \le \chi_f(G) \le \chi(G).$$ Σχέση 4.λ

και τελικά

$$\frac{\chi(G)}{1+\ln\alpha(G)} \leq \chi_f(G) \leq \chi(G) \qquad \text{Σχέση 4.μ}$$

όπου n(G) το πλήθος των κόμβων , α(G) είναι ο αριθμός ανεξαρτησίας του γράφου δηλαδή το μέγιστο σύνολο ανεξαρτησίας ενός γράφου, που αποτελείται από τους κόμβους που δημιουργούν υπογράφους δίχως ακμή, ω(G) ο αριθμός clique και χ(G) ο κλασσικός χρωματικός αριθμός.

ΚΕΦΑΛΑΙΟ 5 -Προτεινόμενη μέθοδος επίλυσης

Στο κεφάλαιο αυτό γίνεται παρουσίαση της προτεινόμενης μεθόδου καθώς και ανάλυση των διάφορων βημάτων που ακολουθούνται. Επιπλέον, επεξηγείται το πως η μέθοδος αυτή ανάγεται σε μια μέθοδο κλασματικού χρωματισμού που με συνέπεια την βελτίωση του επιπέδου ασφάλειας του δικτύου σε σχέση με τους κλασσικούς χρωματισμούς γράφων που έχουν προταθεί σε αντίστοιχες μελέτες.

5.1 Αλγόριθμος Δυναμικού Χρωματισμού (ΑΔΧ)

Το πρώτο στάδιο της μεθόδου αποτελεί η στατική ανάλυση του δικτύου και η προσέγγιση του χρωματικού αριθμού. Μέσω της εφαρμογής διαφόρων αλγορίθμων constrained clustering εντοπίζεται ο αριθμός των χρωμάτων k τέτοιος ώστε να μην υπάρχει κανένα conflict ανάμεσα στους κόμβους του δικτύου. Έτσι διασφαλίζουμε ότι όλοι οι κόμβοι x_i θα έχουν διαφορετικό χρώμα από τους γείτονες τους x_j, οι οποίοι ορίζονται ως οι κόμβοι που απέχουν από τον x_i απόσταση που ανήκει μέσα σε ένα ορισμένο διάστημα $2r_s-\alpha\leq|x_i-x_j|\leq2r_s$, όπου α μία παράμετρος ασφάλειας και r_s η ακτίνα ανίχνευσης. Οι αλγόριθμοι που υλοποιούνται και εμπειρικά ταιριάζουν καλύτερα σε αυτού του είδους τα δίκτυα αποτελούν οι walkthrough, constrained agglomerative algorithm, pck-means και η επιλογή του αριθμού των χρωμάτων γίνεται ανάλογα με τα χαρακτηριστικά του δικτύου, αριθμός κόμβων αποθήκευσης, μέσο πλήθος γειτόνων κόμβου κτλ.

Τις περισσότερες φορές όμως σε πραγματικά δίκτυα το πλήθος n των κόμβων αποθήκευσης και γενικότερα των εξυπηρετητών (servers) είναι μικρότερο από τον αριθμό των χρωμάτων k και έτσι υπάρχουν ομάδες που δεν αντιστοιχούν σε κάποιο εξυπηρετητή υπαρκτού χρώματος. Στην περίπτωση λοιπόν που n<k τότε εφαρμόζεται το δεύτερο στάδιο της μεθόδου που είναι ο δυναμικός χρωματισμός του δικτύου. Στους κόμβους αποθήκευσης ανατίθεται από ένα χρώμα που είναι και μοναδικό. Οι κόμβοι ανίχνευσης μπορούν και γνωρίζουν ποιοι είναι οι κόμβοι, που απέχουν απόσταση έως και 2rs από αυτούς, καθώς και το χρώμα που τους έχει ανατεθεί. Υπάρχουν 3 τρόποι με τους οποίους μπορεί να γίνει η επιλογή των κόμβων που θα αλλάζουν το χρώμα τους δυναμικά:

1. **Επιλεκτικός (selective)**: Οι κόμβοι ανίχνευσης του δικτύου ομαδοποιούνται σε n το πλήθος clusters όσοι δηλαδή και οι κόμβοι αποθήκευσης. Εξετάζοντας μία μία τις ακμές του γράφου και εντοπίζοντας εκείνες στους οποίους οι κόμβοι ανήκουν στην ίδια ομάδα, το χρώμα n+1 επισυνάπτεται σε έναν από αυτούς τους δύο κόμβους.

2. **Άμεσος (direct)**: Γίνεται μία ταξινόμηση των k ομάδων ως προς το πλήθος και επιλέγονται οι k-n μικρότερες σε πλήθος ομάδες στους κόμβους των οποίων και ανατίθεται το μη υπαρκτό χρώμα n+1

3. **Υβριδικός (hybrid)**: Ο τρόπος αυτός επιλογής των κόμβων αποτελεί έναν συνδυασμό των προηγούμενων δύο. Γίνεται ταξινόμηση των k ομάδων ως προς το πλήθος. Υπάρχουν i-n+1, με $n\leq i\leq k$, τρόποι να χρωματιστούν οι κόμβοι. Στην περίπτωση που επιλεγεί i=n η μέθοδος συμπίπτει με την selective ενώ αν i=k η μέθοδος συμπίπτει με την direct. Στις ενδιάμεσες περιπτώσεις, χρωματίζονται οι κόμβοι εκείνοι που δημιουργούν conflicts και τα clusters εκείνα που έχουν τους λιγότερους κόμβους ώσπου να φτάσουμε σε n το πλήθος ομάδες.

Έτσι καταλήγουμε σε n+1 ομάδες στις οποίες οι n έχουν το μεγαλύτερο πλήθος κόμβων ανίχνευσης του δικτύου, δεν έχουν γειτονικούς κόμβους που να έχουν το ίδιο χρώμα και στέλνουν τα δεδομένα τους σε κάποιο υπαρκτό κόμβο αποθήκευσης. Έτσι παραμένει μόνο η ομάδα n+1 που δεν έχει τα χαρακτηριστικά των προηγούμενων ομάδων. Οι κόμβοι λοιπόν της ομάδας αυτής έχουν μείνει ουσιαστικά χωρίς χρώμα και γι αυτό το λόγο θα αλλάζουν το χρώμα τους δυναμικά ανάλογα με το ποιοι γείτονες τους την κάθε χρονική στιγμή συνανιχνεύουν τον κινητό στόχο. Ακολουθεί ένα ενδιάμεσο στάδιο προεπεξεργασίας (Preprocessing) στο οποίο εξετάζεται αν οι κόμβοι της ομάδας n+1, μετά τον μετασχηματισμό των ομάδων, μπορούν να αποκτήσουν ένα υπαρκτό χρώμα ανάλογα με αυτό των γειτόνων τους.

Το τελικό στάδιο του αλγορίθμου αποτελεί και την ουσία του που είναι ο δυναμικός χρωματισμός. Αν λοιπόν την χρονική στιγμή ti ανιχνεύουν τον κινητό στόχο d το πλήθος γείτονες, με d<n τότε ο κόμβος x_i, που του έχει ανατεθεί το χρώμα n+1, έχει επιλογή ανάμεσα από τα n-d εναπομείναντα χρώματα. Αν ισχύει η περίπτωση που d>n τότε ο κόμβος x_i επιλέγει το χρώμα του λιγότερο χρησιμοποιούμενου κόμβου αποθήκευσης, υπολογίζοντας το πλήθος των κόμβων ανίχνευσης που στέλνουν στον κάθε κόμβο αποθήκευσης. Παρακάτω φαίνονται τα βήματα του αλγορίθμου συνοδευόμενα και με ψευδοκώδικα:

1. **Είσοδος**:k, αριθμός ομάδων δίχως conflicts
   Amount:[k_1,k_2,…,k_k], πλήθος κόμβων κάθε μιας από τις k ομάδες
   n, αριθμός επιθυμητών ομάδων (αριθμός εξυπηρετητών)

   Ταξινόμηση k ομάδων σε φθίνουσα σειρά ως προς το πλήθος:

```
1)sorting (Amount)
```

   Έξοδος: η ταξινομημένη δομή Amount με φθίνουσα σειρά πλήθους των ομάδων

2. **Είσοδος**: Amount:[k_1,k_2,…,k_k], ταξινομημένο πλήθος κόμβων
   Colors:[C_1,C_1,…,C_{ns}], χρώματα όλων των κόμβων ανίχνευσης πλήθος n_s
   n, αριθμός επιθυμητών ομάδων (αριθμός εξυπηρετητών)
   strategy:[direct,selective,mixed], στρατηγική επιλογής κόμβων
   Edges:[(a,b), (a,c)…, (d,w)], σύνολο ακμών στο γράφο

   Μετασχηματισμός χρωμάτων:

```
Αν strategy="direct":
  1)Αν Colors[i]<=n τότε Συνέχισε
  2)Αλλιώς Colors[i]=n+1
Αλλιώς αν strategy="selective":
  1)Για κάθε link∈Edges κάνε:
  2)   t1=link[i][0]
  3)   t2=link[i][1]
  4)   Αν Colors[t1]=Colors[t1] κάνε:
  5)           Colors[t1]=n+1
```

```
Αλλιώς:
  1) Βήματα 1-5 selective          /*hybrid*/
       2) Βήματα 1-2 direct
```

Έξοδος: η δομή colors με τα χρώματα μετασχηματισμένα

3. **Είσοδος**: Nbr, σύνολο κόμβων ανίχνευσης που γειτνιάζουν
   Colors:[C_1,C_1,…,C_{ns}], χρώματα κόμβων μετά τον μετασχηματισμό

Προεπεξεργασία κόμβων που έχουν το n+1 χρώμα:

```
1)Αν colors[i]=n+1 τότε:
2)d=0

                                  /*πλήθος διαφορετικών
                                   χρωμάτων γειτόνων*/
3)Για κάθε x_i∈Nbr κάνε:
4)   Αν χρώμα διαφορετικό από των γειτόνων:
     d=d+1

5)Αν d<n τότε:
6)   επιλογή εναπομεινάντων χρωμάτων
```

Έξοδος: η δομή colors με τα χρώματα μετά την προεπεξεργασία

4. **Είσοδος**: Nd, σύνολο κόμβων που απέχουν έως και 2rs
   Colors:[C_1,C_2,…,C_{ns}], χρώματα κόμβων
   target, θέση κινητού στόχου
   n, αριθμός επιθυμητών ομάδων (αριθμός εξυπηρετητών)
   counters:[l_1,l_2,…,l_{dn}], μετρητής χρησιμοποιούμενων χρωμάτων

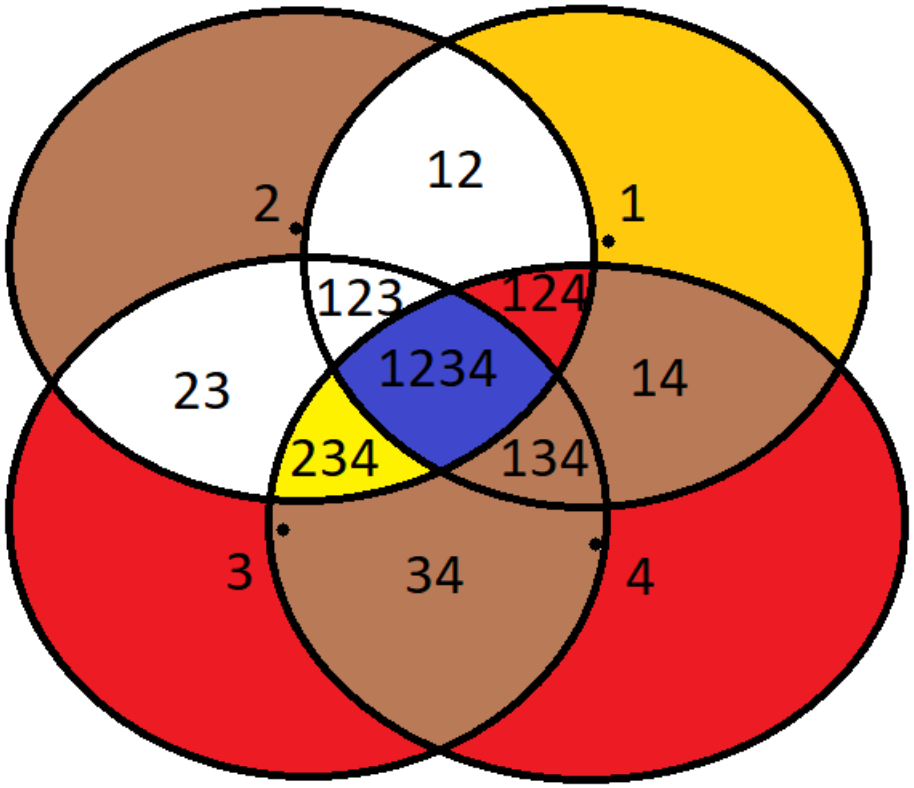
Δυναμικός Χρωματισμός:

```
1)Αν |target-x_i|<=rs τότε:
2)   Αν colors[i]<n+1 τότε:
3)        Στείλε μήνυμα (server με χρώμα i)
4)        για κάθε x_i∈Nd:{C(x_i)=n+1} κάνε:
5)             Στείλε μήνυμα (x_i)
6)   Αλλιώς:
7)        Για κάθε x_i∈Nd κάνε:
8)             d=0
9)             Αν χρώμα διαφορετικό από των
```

```
                υπολοίπων:
10)                  d=d+1
11)        Αν d<n τότε:
12)             επιλογή εναπομεινάντων χρωμάτων
13)         Αλλιώς:
14)             min (counters)
15)        Στείλε μήνυμα (server με χρώμα που
           επιλέχθηκε)
```

Έξοδος: παρακολούθηση κινητού στόχου και ενημερωμένοι servers

5.2 Συνδυασμός graph coloring και time-based coloring



Σχήμα 4: Συνδυασμός graph coloring και time-based coloring

Ο αλγόριθμος που παρατέθηκε στην προηγούμενη ενότητα, ουσιαστικά ανάγει το πρόβλημα σε μια συνδυαστική εφαρμογή κλασσικής θεωρίας χρωματισμού γράφων αλλά και μιας θεωρίας χρωματισμού βασισμένη στο χρόνο (time-based coloring). Για να γίνει πιο κατανοητή αυτή η πρόταση, θα επεξηγηθεί μέσω παραδείγματος. Στο Σχήμα 4 φαίνεται μια τοπολογία αισθητήρων που είναι τοποθετημένοι συμμετρικά σε σχήμα τετραγώνου και είναι πανομοιότυποι με την έννοια

ότι έχουν την ίδια ακτίνα ανίχνευσης, είναι κατασκευασμένοι από το ίδιο hardware, έχουν την ίδια χωρητικότητα μνήμης και επίσης μπορούν να εκπέμψουν μήνυμα έως την ίδια ακριβώς απόσταση. Η τοπολογία αυτή αποτελεί μέρος μιας ευρύτερης τοπολογίας δικτύου η οποία περιέχει 3 κόμβους αποθήκευσης (n=3). Σε κάθε επιμέρους περιοχή του σχήματος αναγράφεται ακριβώς ποιοι κόμβοι μπορούν να ανιχνεύσουν τον κινητό στόχο. Η κάθε περιοχή θα ονομάζεται με αυτόν τον τρόπο πχ αναφερόμενοι στην περιοχή 134, θα εννοούμε την περιοχή στην οποία ανιχνεύουν οι κόμβοι 1,3,4.

Στο πρώτο στάδιο του αλγορίθμου, που εξηγήσαμε αναλυτικά στην ενότητα 5.2, εφαρμόζονται οι αλγόριθμοι constrained clustering για να βρεθεί ο αριθμός χρωμάτων k τέτοιος ώστε να μην υπάρχει κανένα conflict στο δίκτυο. Στην περίπτωση αυτή ας υποθέσουμε ότι μετά την υλοποίηση αυτών των αλγορίθμων, έχουμε k=5 καθώς και ότι στους κόμβους 1,2,3 επισυνάφθηκαν τα αντίστοιχα υπαρκτά χρώματα κίτρινο, καφέ, κόκκινο (n=3) αντίστοιχα ενώ στον κόμβο 4 επισυνάφθηκε το μη υπαρκτό χρώμα 4 (n+1). Έτσι έχουμε εφαρμόσει ορθώς την κλασσική θεωρία χρωματισμού γράφου αφού κανένας κόμβος ανίχνευσης δεν έχει το ίδιο χρώμα με κάποιον γειτονικό του. Ωστόσο ενώ οι 3 πρώτοι κόμβοι έχουν αντιστοιχηθεί σε κάποιον υπαρκτό κόμβο αποθήκευσης, ο 4ος κόμβος δεν μπορεί να στείλει κάπου τα μηνύματα του. Γι αυτόν ακριβώς το λόγο έρχεται η θεωρία χρωματισμού, που είναι βασισμένη στο χρόνο (time-based coloring), να συμπληρώσει αυτό το κενό.

Ο κόμβος 4 θα αλλάζει δυναμικά το χρώμα του ανάλογα με τη θέση του κινητού στόχου και το ποιοι γειτονικοί κόμβοι τον ανιχνεύουν κάθε χρονική στιγμή. Συνεπώς κάνει επιλογή μεταξύ των διαθέσιμων χρωμάτων, και κατά συνέπεια των κόμβων αποθήκευσης, που δεν χρησιμοποιούνται εκείνη τη χρονική στιγμή. Για παράδειγμα στην περιοχή 4 λόγω του ότι κανένας άλλος κόμβος δεν ανιχνεύει, θα μπορούσαμε να οδηγούμε τα μηνύματα προς οποιονδήποτε κόμβο αποθήκευσης. Στις περιοχές 134,14,34 ο κόμβος που δεν ανιχνεύει είναι ο 2 στην 134 και οι 2,4 στις 14,34 οπότε επισυνάπτεται στον κόμβο 4 το καφέ χρώμα, τις χρονικές στιγμές που ο κινητός στόχος βρίσκεται σε αυτές τις περιοχές, καθώς είναι είναι ένα από τα διαθέσιμα χρώματα. Αντίστοιχα στην περιοχή 124 επισυνάπτεται το κόκκινο χρώμα, αφού ο 3 δεν ανιχνεύει, ενώ στην περιοχή 234 για τον ίδιο λόγο επισυνάπτεται το κίτρινο χρώμα. Ιδιαίτερο ενδιαφέρον παρουσιάζει η μπλε περιοχή του σχήματος όπου όλοι οι κόμβοι ανιχνεύουν ταυτόχρονα το κινητό στόχο και όλα τα χρώματα χρησιμοποιούνται. Εκεί δεν μπορεί να γίνει επισύναψη κάποιου χρώματος που να μην συμπίπτει με κάποιο χρώμα γειτονικού του κόμβου. Έτσι ο κόμβος 4 επιλέγει αυθαίρετα κάποιο χρώμα και στέλνει τα δεδομένα του. Μια μικρή βελτίωση που θα μπορούσε να γίνει, είναι ο κόμβος 4, εφόσον γνωρίζει τους γείτονες του και την σχετική τους θέση ως προς αυτόν, να επιλέγει το χρώμα του κοντινότερου του γείτονα διότι η τομή μεταξύ τους είναι μεγαλύτερη από ότι με έναν πιο απομακρυσμένο ως προς αυτόν κόμβο. Έτσι ο κινητός στόχος μπορεί να εντοπιστεί σε μία μεγαλύτερη περιοχή. Για παράδειγμα η περιοχή που περικλείει τις 34,134,234,1234 είναι μεγαλύτερη από την περιοχή που περικλείει τις 124,234,1234 γι αυτό και θα μπορούσε να επιλέξει το κόκκινο χρώμα.

Παρόλο που βλέπουμε ότι υπάρχει μια περιοχή στο δίκτυο (περιοχή 1234), που είναι επικίνδυνη εντούτοις έχουμε καταφέρει να την περιορίσουμε σε μεγάλο βαθμό σε σχέση με μια στατική ανάθεση χρωμάτων. Στη δεύτερη περίπτωση η προβληματική περιοχή θα ήταν κάποια εκ των περιοχών α(34,234,134,1234), β(14,134,124,1234), γ(124,234,1234) οι οποίες είναι πολύ μεγαλύτερες και μάλιστα μπορούμε να παρατηρήσουμε ότι η 1234 αποτελεί υποσύνολο όλων των περιοχών α,β,γ. Έτσι με αυτόν τον τρόπο μπορούμε να αυξήσουμε το επίπεδο της ασφάλειας του δικτύου.

Η τοπολογία αυτή αποτελεί μέρος μιας ευρύτερης τοπολογίας που έχει χρησιμοποιηθεί για πειραματική μελέτη και παρατίθεται στο κεφάλαιο 6 (ενότητα 6.4). Η μέθοδος αυτή ενώ αναλύθηκε μέσω πολύ συγκεκριμένου παραδείγματος μπορεί να γενικευθεί και αυτό αποδεικνύεται μέσω πειραμάτων που διεξάγονται στο κεφάλαιο 6 (ενότητα 6.5).

ΚΕΦΑΛΑΙΟ 6 -Πειραματική αξιολόγηση και ανάλυση

Στο κεφάλαιο αυτό θα γίνει επεξήγηση και ανάλυση του τεχνικού μέρους της διπλωματικής. Αρχικά θα παρουσιαστεί η πλατφόρμα FIT IoT LAB που χρησιμοποιήθηκε για το πειραματικό μέρος της μελέτης αυτής. Παρείχε τις διαφορετικές τοπολογίες πάνω στις οποίες εφαρμόστηκαν οι σχετικοί αλγόριθμοι, που παρουσιάστηκαν στα προηγούμενα κεφάλαια, καθώς επίσης και τις διάφορες τροχιές του κινητού στόχου. Στη συνέχεια γίνεται παρουσίαση των r και python libraries που χρησιμοποιήθηκαν για την ανάλυση των δεδομένων του δικτύου καθώς επίσης για την επικοινωνία μεταξύ των κόμβων. Τέλος, παρουσιάζονται και επεξηγούνται τα πειραματικά αποτελέσματα τόσο της στατικής ανάλυσης όσο και της δυναμικής ανάλυσης του δικτύου.

6.1 FIT IoT LAB

Το FIT IoT LAB αποτελεί μια υποδομή η οποία χρησιμοποιείται για πειράματα με ασύρματους κόμβους ανίχνευσης και διάφορες ετερογενείς συσκευές [FIT-IoT]. Το FIT IoT LAB περιλαμβάνει πάνω από 1500 συσκευές ανίχνευσης οι οποίες είναι παραταγμένες σε 6 διαφορετικές τοποθεσίες στη Γαλλία. Πιο συγκεκριμένα, αποτελείται από 1791 ασύρματες συσκευές ανίχνευσης οι οποίες βρίσκονται στις τοποθεσίες Grenoble (640), Lille (332), Sanclay (230), Strasbourg (400), Paris (160), Lyon (29) όπως φαίνεται και στο Σχήμα 5. Παρέχει μεγάλη ποικιλία ασύρματων συσκευών οι οποίες διαθέτουν διαφορετικές αρχιτεκτονικές (MSP430, STM32 and Cortex-A8) και ολοκληρωμένα κυκλώματα ασύρματης επικοινωνίας (802.15.4 PHY @ 800 MHz or 2.4 GHz).

Αποτελεί ουσιαστικά ένα επιστημονικό testbed που παρέχει πλήρη έλεγχο των ασύρματων συσκευών και απευθείας πρόσβαση στα gateways από τα οποία οι ενδιαφερόμενοι ερευνητές μπορούν να εξάγουν διάφορες πληροφορίες για τους κόμβους του δικτύου τους όπως η κατανάλωση της ενέργειας, η καθυστέρηση από άκρο σε άκρο και η συμφόρηση. Διευκολύνει την διαδικασία της παράταξης δικτύων για επιστημονικό σκοπό όπως επίσης και την διαδικασία συλλογής και ανάλυσης των δεδομένων.

Πέρα όμως από τους στατικούς κόμβους ανίχνευσης, το FIT IoT LAB παρέχει μια πληθώρα κινητών κόμβων. Μέχρι στιγμής διαθέτει 15 κινητούς κόμβους στις εξής τοποθεσίες Grenoble (2), Lille (3), Strasbourg (10). Οι κόμβοι αυτοί έχουν προκαθορισμένες τροχιές και κινούνται στο έδαφος ανάμεσα από τους στατικούς κόμβους ανίχνευσης.

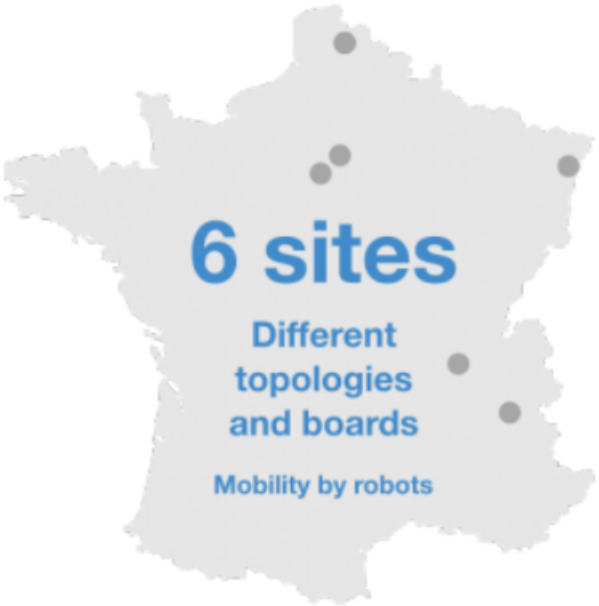
Σημαντικό επίσης αποτελεί το γεγονός ότι προσφέρει ευελιξία στο χρήστη όσον αφορά την υλοποίηση του πειράματός του. Αρχικά προσφέρονται 3 επίπεδα API με Drivers, Operating Systems, Communication Libraries, προκειμένου να μπορεί ο χρήστης να προγραμματίσει embedded εφαρμογές. Οι drivers προσφέρουν απομακρυσμένη πρόσβαση στους κόμβους ενώ τα Operating Systems and Communication Libraries υψηλότερου επιπέδου λειτουργίες. Τέλος, διατίθενται εργαλεία που διευκολύνουν την παρακολούθηση των κόμβων αλλά και την απομακρυσμένη πρόσβαση στην πλατφόρμα.

Η πρόσβαση μπορεί να επιτευχθεί με δύο τρόπους. Ο πρώτος είναι μέσω του Web portal όπου ο χρήστης μπορεί να έχει πρόσβαση μέσω ενός web browser ενώ ο δεύτερος μέσω command line εργαλείων όπως φαίνεται και στο Σχήμα 6. Επιπλέον, προσφέρονται SSH front end περιβάλλοντα ανά τοποθεσία τα οποία περιέχουν προεγκατεστημένα τα command line εργαλεία τα οποία παρέχουν μεγαλύτερη ευελιξία στο χρήστη.

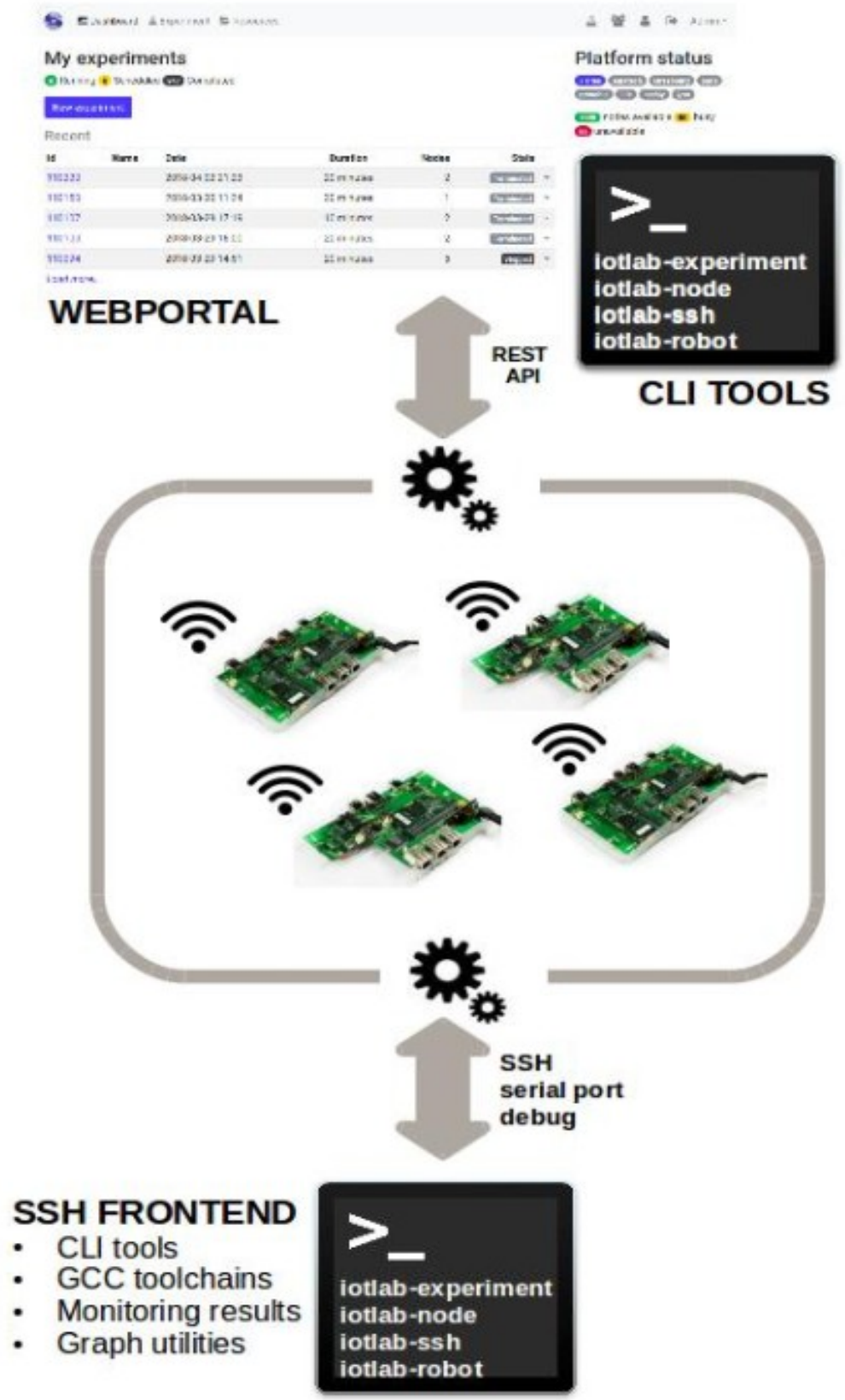
Η πλατφόρμα FIT IoT LAB παρέχει 3 ειδών στατικούς κόμβους, τους WSN430, τους M3

και τους A8. Οι WSN430 [GitHub] κόμβοι βασίζονται στην χαμηλής ενέργειας πλατφόρμα MSP430 με πλήρως λειτουργική ISM ασύρματη διασύνδεση και ένα σύνολο αισθητήρων ανίχνευσης. Όσον αφορά την ασύρματη διασύνδεση, έχουν αναπτυχθεί δύο διαφορετικές εκδόσεις, η version 1.3b στα 868MHz και η version 1.4 στα 2.4GHz. Διαθέτει 3 ειδών αισθητήρες , αισθητήρα θερμοκρασίας, ήχου και διάχυτου φωτός. Οι κόμβοι M3 [GitHub] είναι βασισμένοι στον μικροελεγκτή STM32 (ARM Cortex M3). Όπως και ο κόμβος WSN430, έτσι και ο M3, διαθέτει ένα σύνολο αισθητήρων όπως φωτός, πίεσης, γυρόμετρο, επιταχυνσιόμετρο/μαγνητόμετρο και ασύρματη διασύνδεση ATMEL στα 2.4 GHz. Ο κόμβος A8 είναι βασισμένος στο TI SITARA AM3505 που είναι ένας υψηλής επίδοσης ARM Cortex-A8 μικροεπεξεργαστής. Η ταχύτητα του ρολογιού του στα 600 MHz του δίνει τη δυνατότητα να τρέξει ενσωματωμένα Linux η Android. Επιτρέπει να τρέχει εφαρμογές σε συσκευές με σκοπό την συλλογή πληροφοριών. Και αυτός διαθέτει ένα σύνολο αισθητήρων όπως γυρόμετρο και μαγνητόμετρο/επιταχυνσιόμετρο.

Στις τοπολογίες που χρησιμοποιήθηκαν στο πειραματικό κομμάτι για την διεξαγωγή της μελέτης χρησιμοποιήθηκαν και τα 3 είδη των κόμβων ώστε να εξυπηρετήσουν διαφορετικούς σκοπούς. Οι κόμβοι A8 που επιτρέπουν απομακρυσμένη διασύνδεση μέσω ssh και είναι κατάλληλοι για την συλλογή των πληροφοριών χρησιμοποιήθηκαν κυρίως για να προσομοιώσουν τους κόμβους αποθήκευσης.



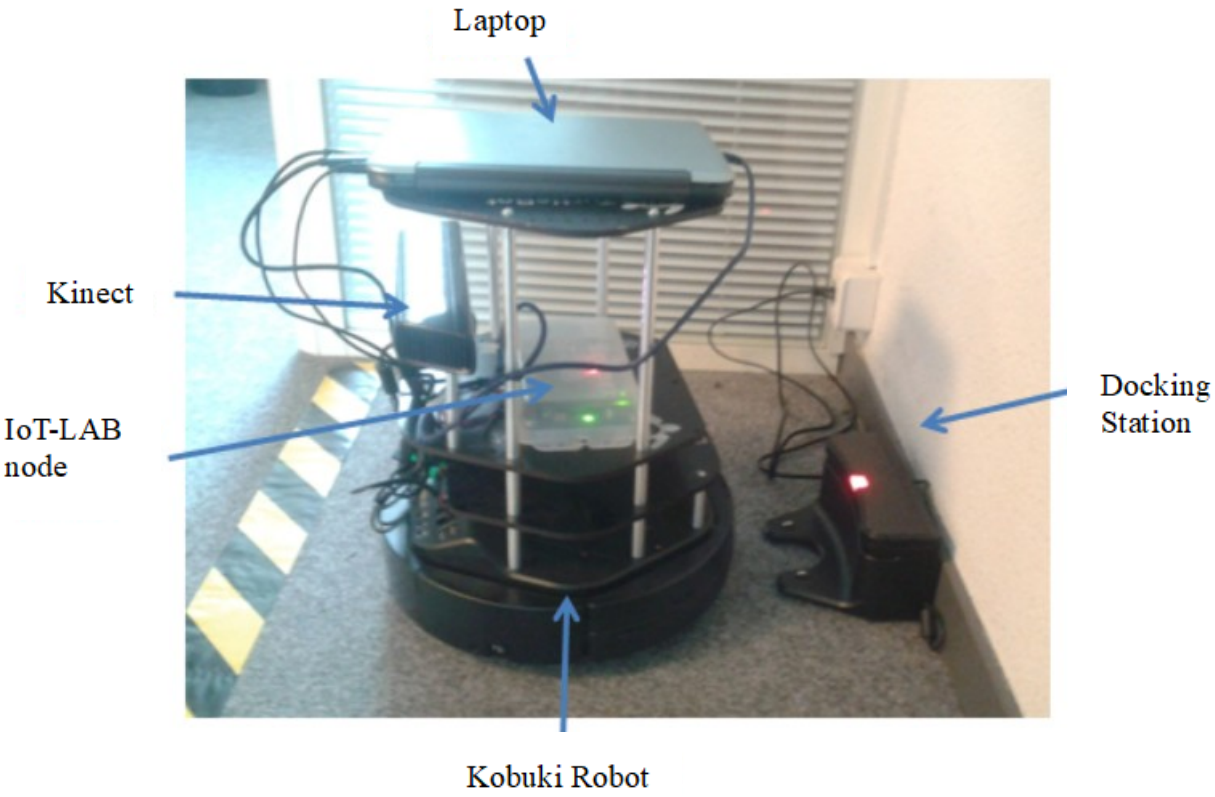
Σχήμα 5: Υποδομές FIT IoT LAB [FIT IoT LAB]



Σχήμα 6: Πρόσβαση στο FIT IoT LAB [FIT IoT LAB]

6.2 Turtlebot

Ως κινητός στόχος του πειράματος χρησιμοποιήθηκε ο κόμβος turtlebot που προσφέρει το FIT IoT LAB[GitHub]. Αποτελείται από ένα notebook, ένα kinect, έναν κόμβο FIT-IoT και έναν gateway όπως φαίνεται και στο σχήμα 7.

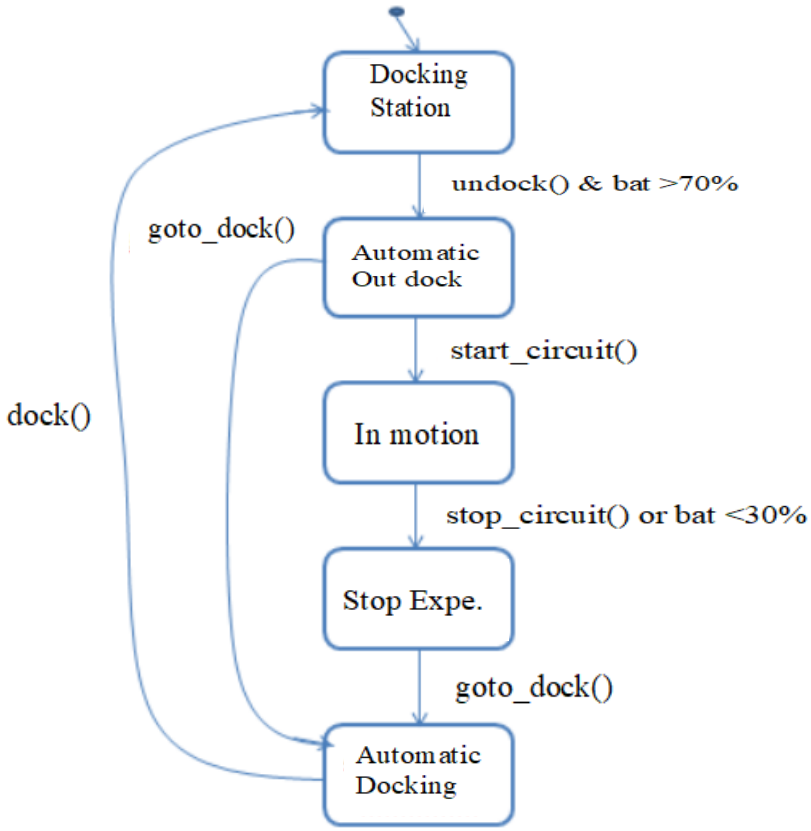


Σχήμα 7: Turtlebot [FIT IoT LAB]

Το notebook χρησιμοποιείται τόσο για την επικοινωνία του κόμβου με την υποδομή του FIT IoT LAB, μέσω wifi, όσο και για τον έλεγχο του ρομπότ μέσω του του λειτουργικού συστήματος ROS [ros]. Ο κινητός κόμβος είναι προσβάσιμος ακριβώς με τον ίδιο τρόπο όπως και οι στατικοί κόμβοι του δικτύου μέσω API.

Όσων αφορά την τροχιά του κινητού κόμβου τότε αυτή είναι προκαθορισμένη, διαφορετική για κάθε τοποθεσία (Strasbourg,Lille,Grenoble) και επιλέγεται από το προφίλ του χρήστη. Το ρομπότ μπορεί να χρησιμοποιηθεί μόνο όταν είναι αγκυροβολημένο στη βάση του. Η τροχιά του κινητού κόμβου, όταν ολοκληρωθεί, επαναλαμβάνεται αρκετές φορές ώσπου να καλυφθεί ο απαιτούμενος χρόνος του πειράματος. Όταν η διάρκεια του πειράματος ολοκληρωθεί, το ρομπότ επιστρέφει πίσω στη βάση του προκειμένου να φορτιστεί. Κατά τη διάρκεια της τροχιάς του έχει τη δυνατότητα να παρακάμπτει τα εμπόδια. Για να αποφευχθεί η πιθανότητα το κινητό αντικείμενο να σταματήσει την πορεία του κατά την διάρκεια του πειράματος λόγω έλλειψης μπαταρίας, ο χρήστης μπορεί να το χρησιμοποιήσει μόνο όταν είναι επαρκώς γεμισμένο (δηλαδή πάνω από 70%). Επιπλέον, εάν η διάρκεια του πειράματος είναι αρκετά μεγάλη σε σημείο να εξαντλήσει την μπαταρία (κάτω του 30%), τότε ο κόμβος διακόπτει την πορεία για να γυρίσει στη βάση όπου και

μπορεί να επαναφορτιστεί. Το σχήμα 8 δείχνει το διάγραμμα καταστάσεων του ρομπότ.



Σχήμα 8: Καταστάσεις turtlebot [FIT IoT LAB]

Όσον αφορά τα τεχνικά χαρακτηριστικά του turtlebot, έχει μέγιστη ταχύτητα 0.7 m/s οπότε η συγκεκριμένη ταχύτητα δηλώθηκε ως v_{max} για τις ανάγκες του πειράματος. Είναι εξοπλισμένο με δύο μπαταρίες, μία 3000mAh για το netbook και μία 4400mAh για τους τροχούς. Διαρκεί έως και 3 ώρες όταν είναι πλήρως γεμισμένη και μπορεί να επαναφορτιστεί σε 6 ώρες. Οι συντεταγμένες της τροχιάς του κινητού καταγράφονται σε ένα αρχείο .oml όπου και μπορούν να χρησιμοποιηθούν για πειραματικούς σκοπούς. Το αρχείο μεταφέρθηκε από το FIT IoT LAB με χρήση του προγράμματος WinSCP.

6.3 Python and r βιβλιοθήκες

Για τη διεξαγωγή του πειράματος και την συγγραφή του κώδικα χρησιμοποιήθηκαν ορισμένα βιβλιοθήκες των γλωσσών r και python προκειμένου να υλοποιηθεί το δίκτυο των κόμβων ανίχνευσης και κόμβων αποθήκευσης, η μεταξύ τους επικοινωνία, η εφαρμογή των αλγορίθμων συσταδοποίησης στις διάφορες τοπολογίες, η ανάλυση των δεδομένων και η

απεικόνιση των αποτελεσμάτων.

Αρχικά για το σωστό διάβασμα των διαφόρων αρχείων που προσφέρει το FIT IoT LAB, τα οποία ήταν σε μορφή json, χρησιμοποιήθηκε η βιβλιοθήκη json.

```
import json

with open ('data.json') as f:
    data=json.load (f)
```

Στη συνέχεια για την υλοποίηση του SPG χρησιμοποιήθηκε η βιβλιοθήκη της python networkx η οποία χρησιμοποιείται στη συγκεκριμένη γλώσσα για τη δημιουργία και την προσωμοίωση των γράφων.

```
import networkx as nx

G=nx.Graph()
G.add_node(1)
G.add_node(2)
G.add_edge(1, 2)
```

Ένα από τα σημαντικότερα ζητήματα της μελέτης υπήρξε η αξιόπιστη επικοινωνία μεταξύ των κόμβων αποθήκευσης και των κόμβων ανίχνευσης. Οι κόμβοι ανίχνευσης αποστέλλουν δεδομένα στους κόμβους αποθήκευσης, οι οποίοι δεν έχουν την δυνατότητα να επικοινωνήσουν μεταξύ τους, οπότε επιλέχθηκε το μοντέλο client-server. Οι κόμβοι αποθήκευσης λειτουργούν ως server, οι κόμβοι ανίχνευσης ως client και επικοινωνούν μεταξύ τους με το πρωτόκολλο ZMTP, το ZeroMQ Message Transfer Protocol [zeromq.org]. Χρησιμοποιήθηκε η βιβλιοθήκη zmq της python.

Client side

```
import zmq

context = zmq.Context()
socket = context.socket(zmq.REQ)
socket.connect("server ip")

socket.send(msg)
msg_in = socket.recv()
```

Server side

```
import zmq

context = zmq.Context()
socket = context.socket(zmq.REP)
socket.bind("server ip")
```

```
msg = socket.recv()
res=json.loads(msg)
socket.send(ack)
```

Εξίσου σημαντικό κομμάτι της διπλωματικής αποτελούν οι αλγόριθμοι constrained clustering που χρησιμοποιήθηκαν. Στα δεδομένα εφαρμόσαμε μία πληθώρα τέτοιων αλγορίθμων και καταλήξαμε σε 3 οι οποίοι εμπειρικά και μέσα από πολλές δοκιμές φάνηκε ότι ταιριάζουν πολύ περισσότερο στα συγκεκριμένα δεδομένα. Οι αλγόριθμοι που εφαρμόστηκαν ήταν οι ακόλουθοι:

1. Constrained hierarchical agglomerative :

```
from sklearn.cluster import AgglomerativeClustering

def hierarchical_agglomerative (num_clusters):
     model=AgglomerativeClustering.fit(data_matrix)
return model
```

Η ουσιαστική διαφορά του constrained agglomerative με τον απλό aggomerative clustering αλγόριθμο αποτελεί το γεγονός ότι η πρώτος παίρνει ως είσοδο έναν προκαθορισμένο πίνακα Α μεγέθους n_s x n_s ,όπου n_s το πλήθος των κόμβων ανίχνευσης. Ο πίνακας αυτός σε κάθε θέση έχει την απόσταση ανάμεσα στους αντίστοιχους κόμβους ανίχνευσης για παράδειγμα αν ο κόμβοι 2,3 απέχουν απόσταση 5 τότε θα έχουμε A[2][3]=5. Ωστόσο αν η απόσταση μεταξύ δύο κόμβων ανήκει στο εύρος [2rs-a,2rs] τότε η αντίστοιχη θέση του πίνακα τίθεται σε μία πολύ μεγάλη τιμή πχ max (A)+100. Έτσι κόμβοι που έχουν τέτοια απόσταση είναι λιγότερο πιθανό να μπουν στο ίδιο cluster.

Ως παράμετροι της συνάρτησης έχουμε τον προκαθορισμένο πίνακα που εξηγήσαμε παραπάνω, τον επιθυμητό αριθμό των clusters, και την παράμετρο average linkage η οποία ταίριαζε περισσότερο από τις single και complete καθώς λαμβάνει υπόψιν όλα τα σημεία ενός cluster.

2. Walkthrough:

Έγινε αναπαραγωγή του αλγορίθμου walkthrough που παρουσιάστηκε εκτενώς στο κεφάλαιο 3 (Ενότητα 3.6.2).Ο αλγόριθμος αυτός είναι κατανεμημένος και εκτελείται από τον κάθε κόμβο ανίχνευσης ξεχωριστά. Η αναπαραγωγή έγινε σε γλώσσα python και λόγω του ότι υπήρχε η ανάγκη οι κόμβοι να έχουν πρόσβαση σε κάποια κοινόχρηστα δεδομένα, εξομοιώθηκαν σαν threads χρησιμοποιώντας τις αντίστοιχες βιβλιοθήκες της python threading,multiprocessing.

3. PCk-means:

Για τον αλγόριθμο αυτό χρησιμοποιήθηκε η βιβλιοθήκη conclust της γλώσσας r. Η βιβλιοθήκη αυτή περιέχει 4 αλγορίθμους constrained clustering, τον pck-means, copk-means, lcvqe, ccls. Εξετάστηκαν και οι 4 αλγόριθμοι αλλά οι δύο τελευταίοι δεν εμφάνιζαν αποτελέσματα συγκρίσιμα με τους 3 αλγορίθμους που επιλέξαμε ενώ ο copk-means αποτελεί greedy algorithm [So18, Ka18] και δεν εμφανίζει κάποιο αποτέλεσμα αν βρει έστω και ένα conflict σε κάποιο cluster.

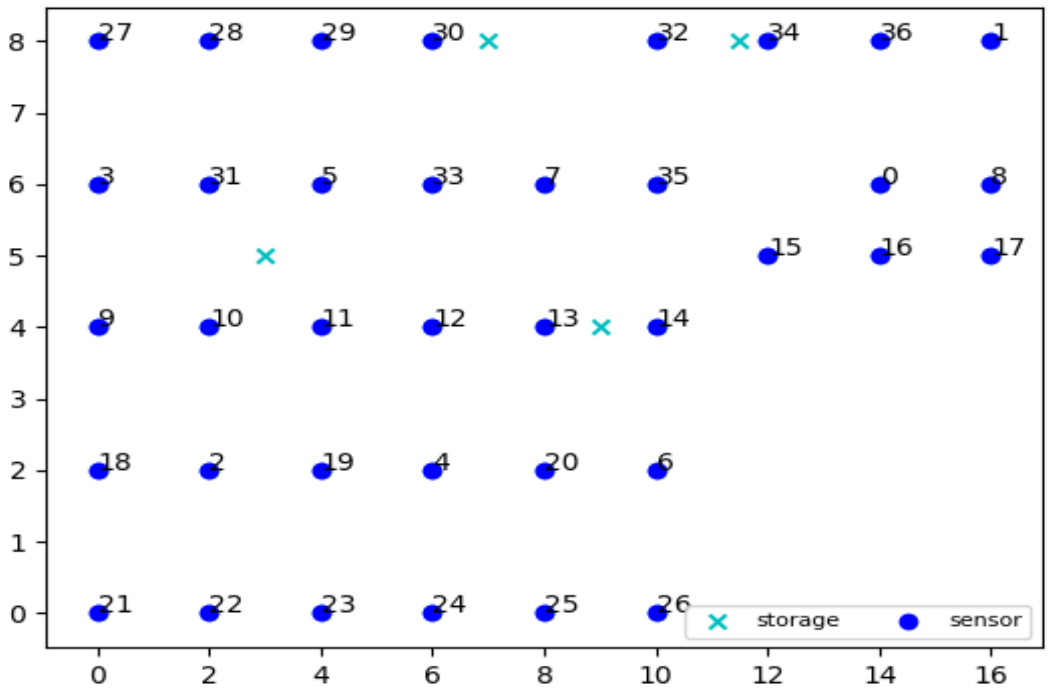
```
pred = mpckm(data,k,mustLink,cantLink)
pred
```

Ο αλγόριθμος παίρνει ως είσοδο τα δεδομένα data που είναι οι x,y συντεταγμένες των κόμβων ανίχνευσης του δικτύου, το k που είναι ο επιθυμητός αριθμός του πλήθους των cluster, το σύνολο mustLink που περιέχει ζεύγη κόμβων που πρέπει να μπουν στο ίδιο cluster (για τη δική μας μελέτη αυτό το σύνολο είναι το κενό []) και ένα σύνολο cantLink που περιέχει τα ζεύγη των κόμβων που δεν πρέπει να μπουν στο ίδιο cluster (εδώ μπαίνουν όλα τα ζεύγη των κόμβων που απέχουν απόσταση στο εύρος [2rs-a,2rs]).

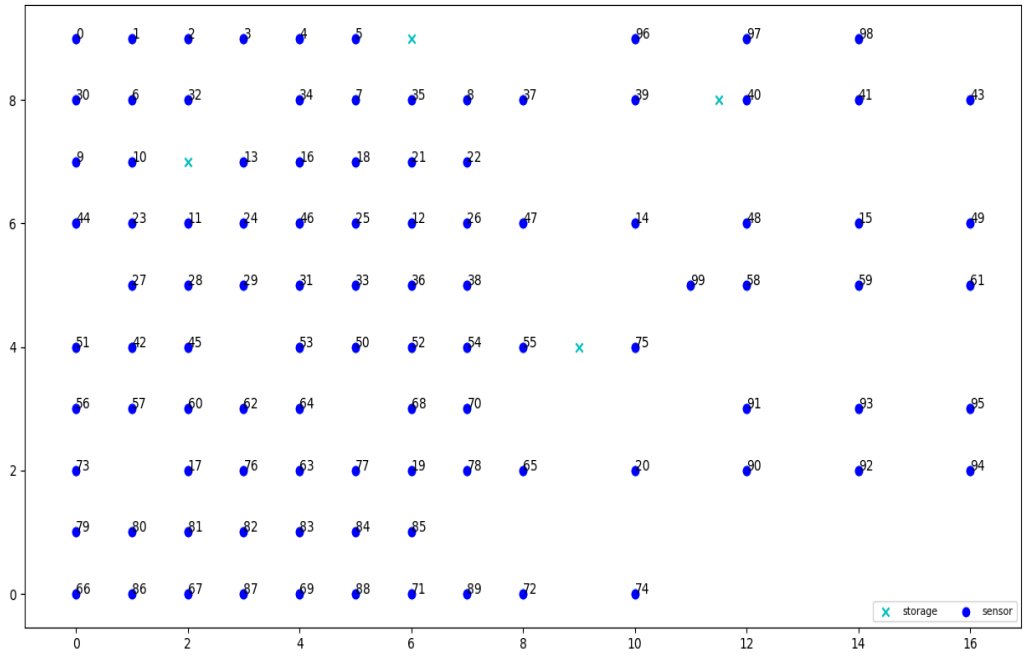
Οι αλγόριθμοι αυτοί έχουν ανοχή ως προς τα conflicts, με την έννοια ότι προσπαθούν να βρουν τον μικρότερο δυνατό αριθμό από conflicts ώστε να μεγιστοποιήσουν την αντικειμενική τους συνάρτηση. Αυτό σημαίνει ότι επιτρέπουν λύσεις που εμπεριέχουν conflicts άλλωστε αυτό είναι και το ζητούμενο της συγκεκριμένης μελέτης.

6.4 Πειραματική αξιολόγηση στατικής ανάλυσης

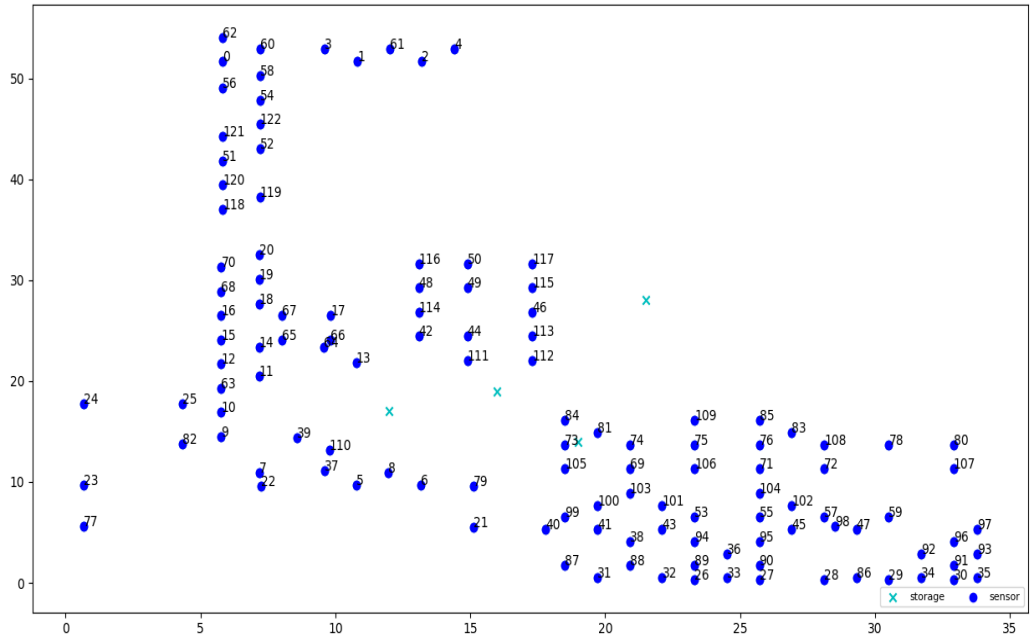
Για την διεξαγωγή των πειραμάτων χρησιμοποιήθηκαν 3 τοπολογίες από το FIT IoT LAB. Το πλήθος των κόμβων ήταν αντίστοιχα 37,100,123. Με σκούρο μπλε χρώμα απεικονίζονται οι κόμβοι ανίχνευσης ενώ με γαλάζιο οι κόμβοι αποθήκευσης (servers). Η μορφή των τοπολογιών φαίνεται παρακάτω:



Σχήμα 9: Τοπολογία με 37 κόμβους και 4 servers



Σχήμα 10: Τοπολογία με 100 κόμβους και 4 servers



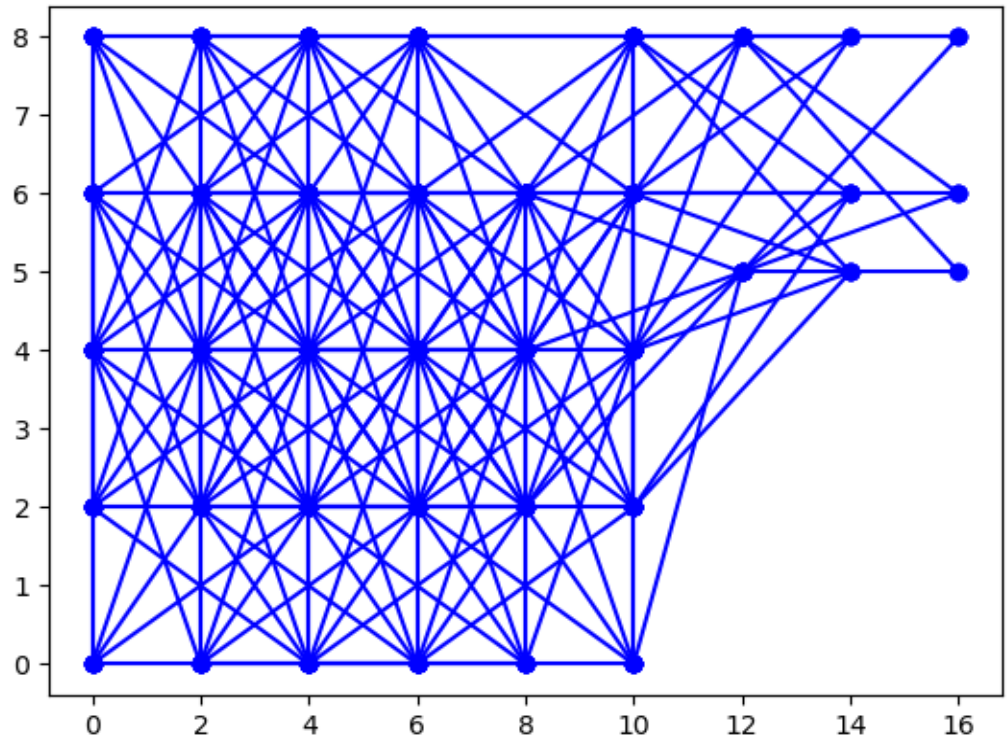
Σχήμα 11: Τοπολογία με 123 κόμβους 3 servers

Οι τρεις αυτές τοπολογίες χρησιμοποιήθηκαν προκειμένου να διεξαχθούν 8 πειράματα στατικής ανάλυσης για την καλύτερη αξιολόγηση των αλγορίθμων pck-means, walkthrough, constrained hierarchical agglomerative. Χρησιμοποιήθηκαν διαφορετικές τιμές για την ακτίνα ανίχνευσης των αισθητήρων r_s καθώς και τον παράγοντα ασφάλειας α προκειμένου να μεταβάλλουμε το πλήθος των κόμβων που συνανιχνεύουν κάθε χρονική στιγμή. Ως κριτήριο για την πυκνότητα ενός γράφου χρησιμοποιήθηκε ο μέσος όρος των γειτόνων των κόμβων ανίχνευσης του δικτύου, κατά τον SPG. Παρατίθενται τα αποτελέσματα ομαδοποιημένα ανά τοπολογία. Σε κάθε πείραμα, η πρώτη εικόνα δείχνει την τοπολογία και η δεύτερη γράφημα με το πλήθος των conflicts.

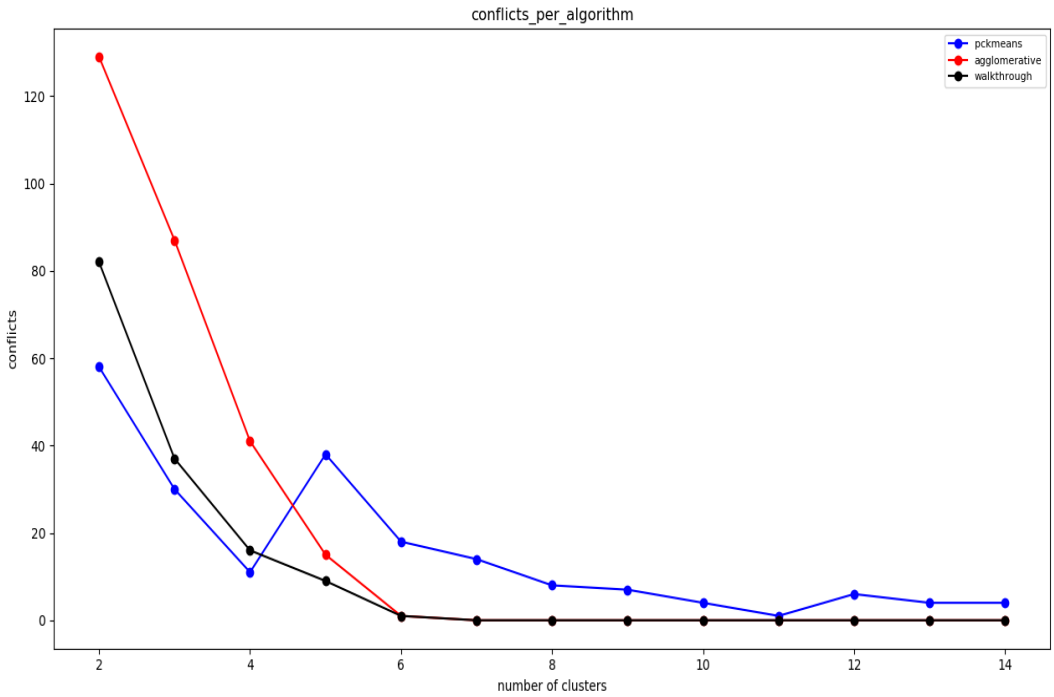
37 κόμβοι

μέσος όρος γειτόνων= 9.2,ακτίνα ανίχνευσης r_s=3,παράγοντας ασφάλειας α=2

dist < rs => συνανίχνευση κόμβων => ακμή στον SPG



Σχήμα 12: SPG της τοπολογίας των 37 κόμβων με r_s=3 και α=2

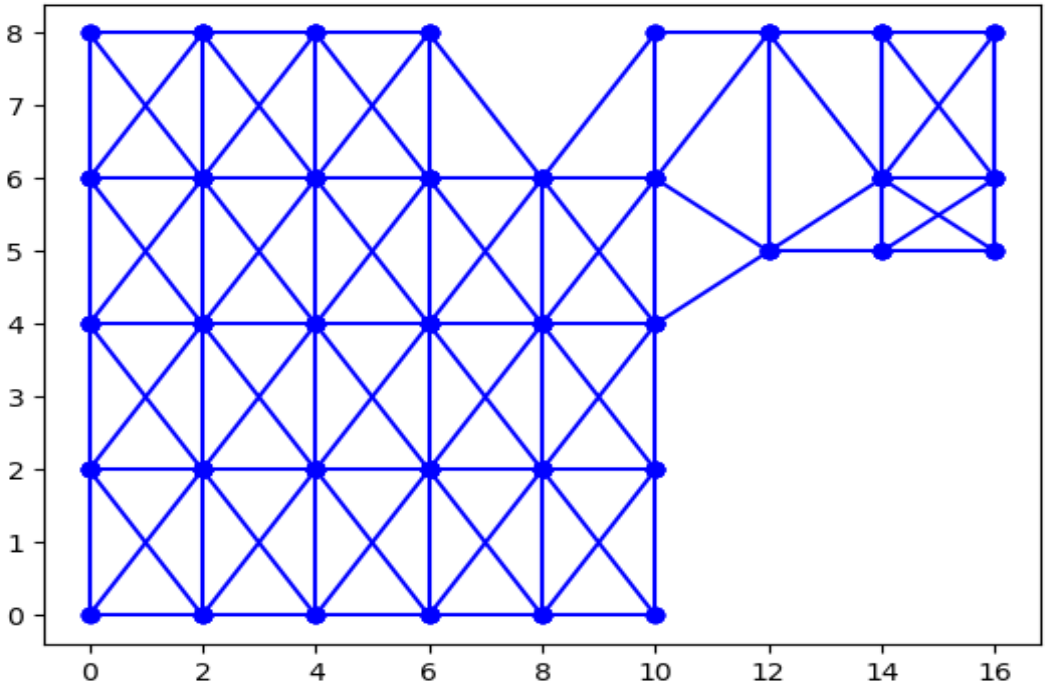


Σχήμα 13: Στατική ανάλυση τοπολογίας 37 κόμβων με r_s=3 και α=2

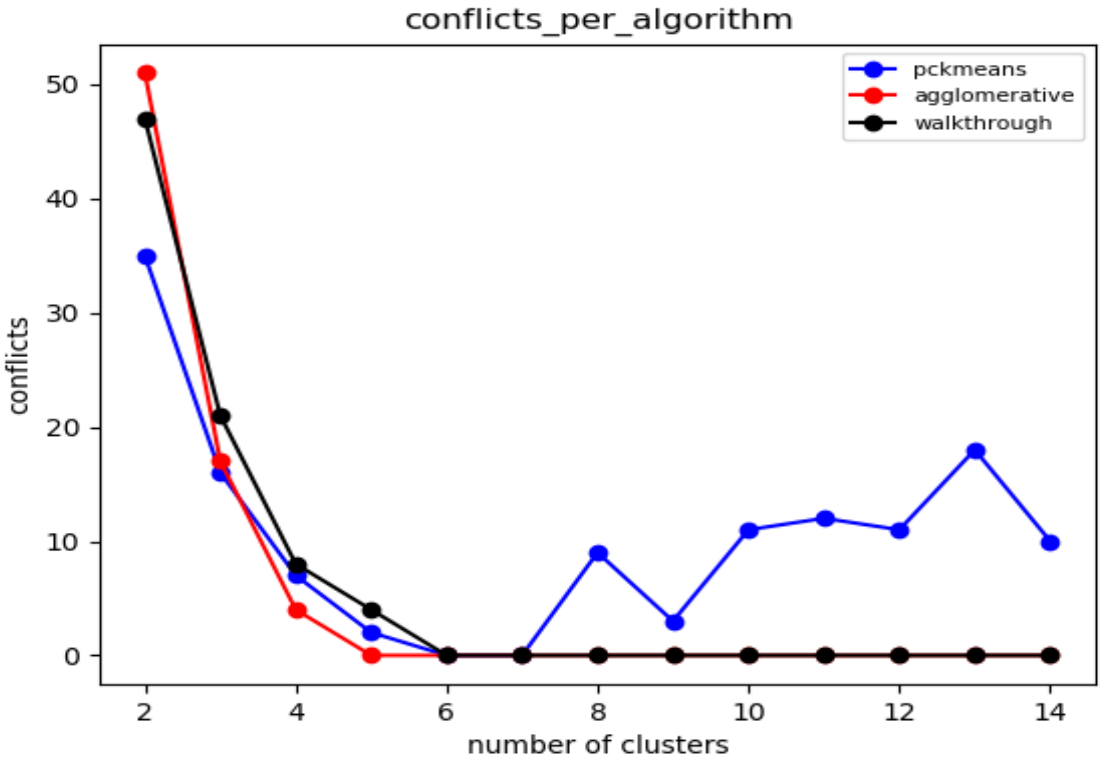
Παρατηρείται ότι ο pckmeans εμφανίζει πολύ λιγότερα conflicts έως και k=4 ενώ αργότερα

παρουσιάζει αστάθεια. Αυτό οφείλεται στην καλύτερη αρχικοποίηση που κάνει αφού σε αυτήν λαμβάνει υπ' όψιν του μόνο τους περιορισμούς του δικτύου. Ο walkthrough έχει καλύτερο ρυθμό σύγκλισης από τον agglomerative αλλά και οι δύο συγκλίνουν στο ίδιο σημείο k=6.

μέσος όρος γειτόνων= 5.7,ακτίνα ανίχνευσης r_s=1.5,παράγοντας ασφάλειας α=2



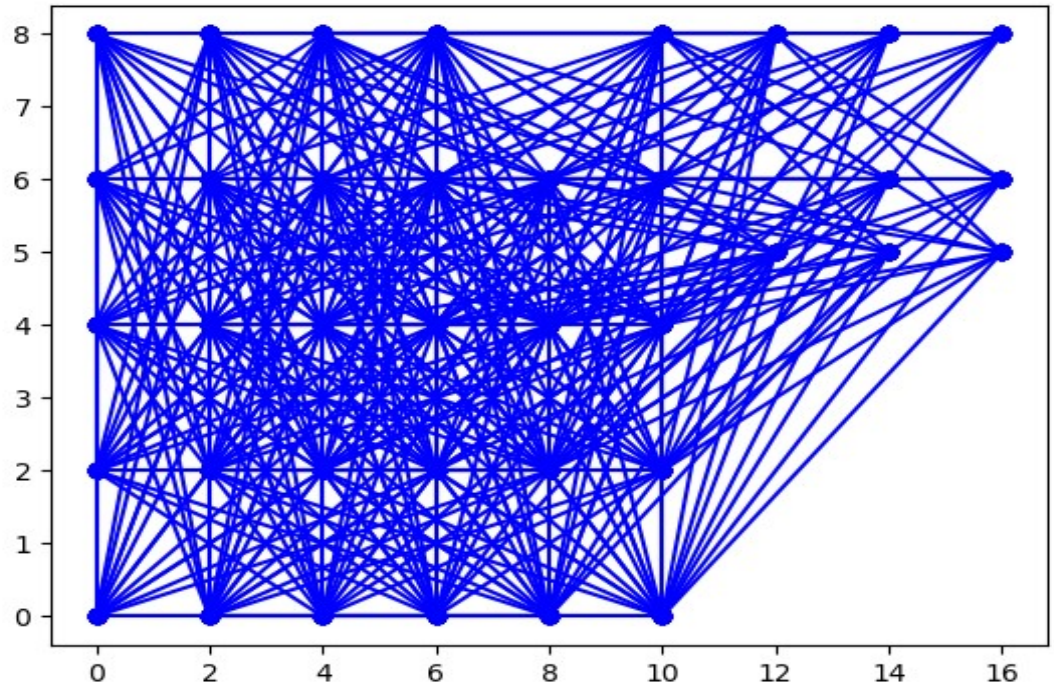
Σχήμα 14: SPG της τοπολογίας των 37 κόμβων με r_s=1.5 και α=2



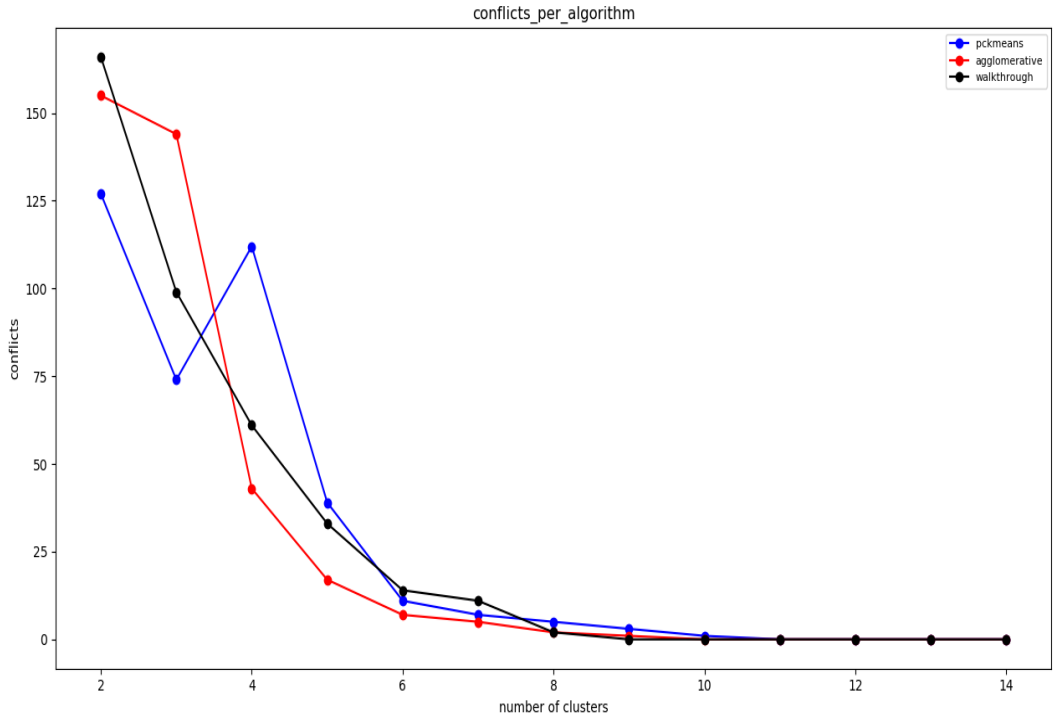
Σχήμα 15: Στατική ανάλυση τοπολογίας 37 κόμβων με r_s=1.5 και α=2

Παρατηρείται ότι ο pckmeans εμφανίζει λιγότερα conflicts έως και k=3 ενώ ο agglomerative έχει καλύτερο ρυθμό σύγκλισης από τον walkthrough. Και οι 3 αλγόριθμοι συγκλίνουν στο ίδιο σημείο k=6 ενώ ο pckmeans από k=7 και μετά γίνεται ασταθής. Το σημείο σύγκλισης των αλγορίθμων παρέμεινε σταθερό αφού δεν άλλαξε δραστικά η πυκνότητα του γράφου.

μέσος όρος γειτόνων=17.8,ακτίνα ανίχνευσης r_s=4.5,παράγοντας ασφάλειας α=4.7



Σχήμα 16: SPG της τοπολογίας των 37 κόμβων με r_s=4.5 και α=4.7

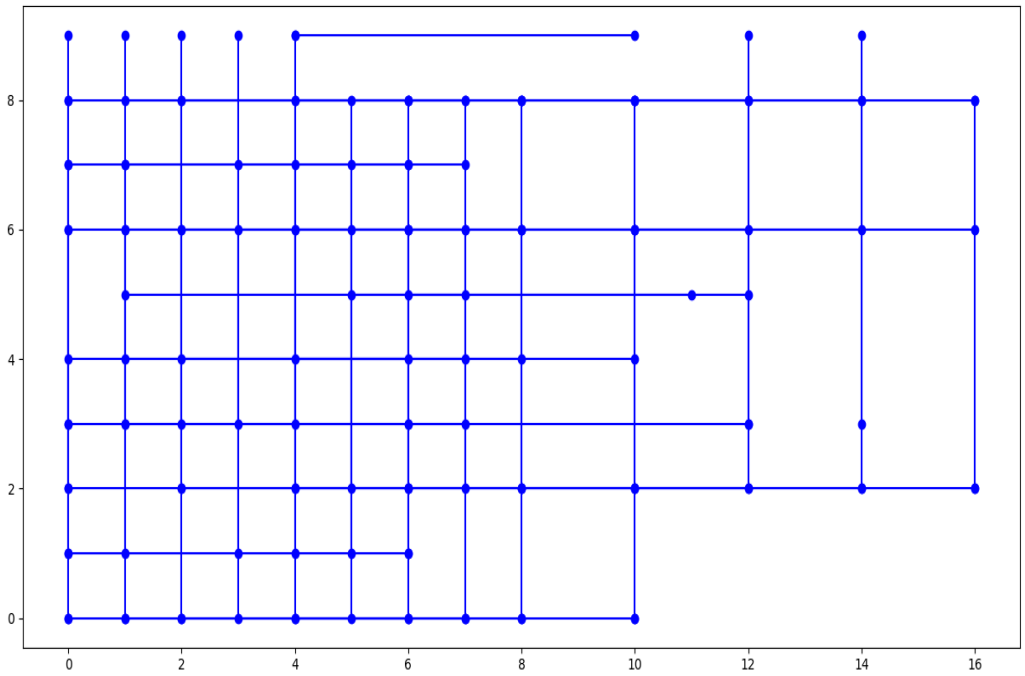


Σχήμα 17: Στατική ανάλυση τοπολογίας 37 κόμβων με r_s=4.5 και α=4.7

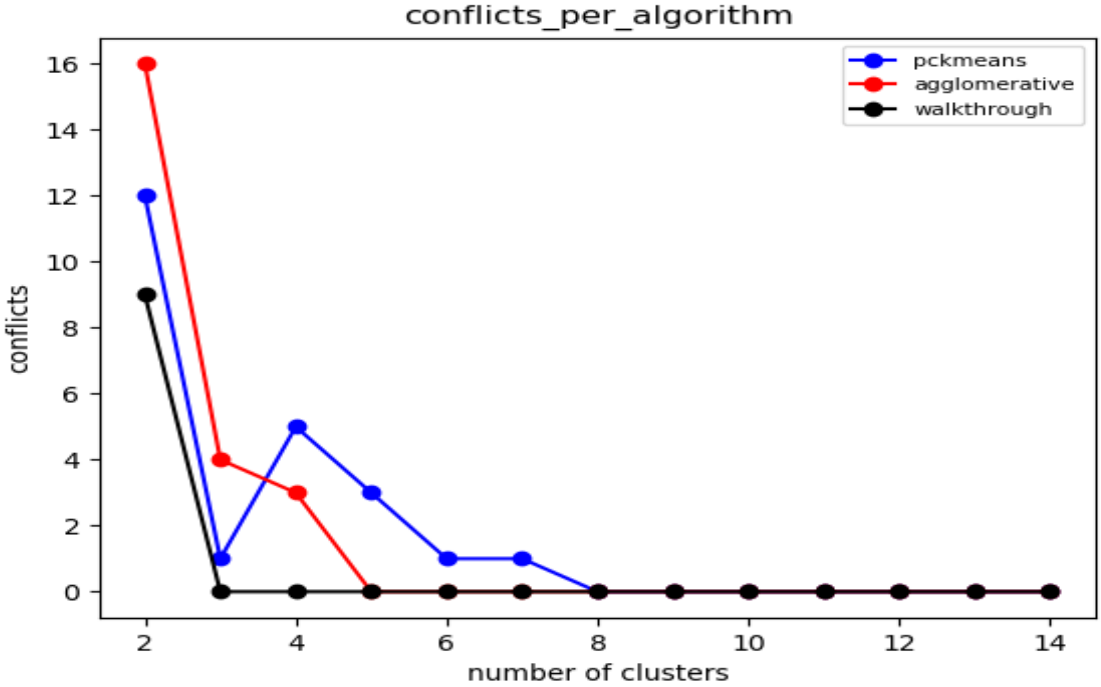
Παρατηρείται ότι ο pckmeans εμφανίζει αισθητά λιγότερα conflicts από τους άλλους 2 αλγορίθμους έως και το σημείο k=3 ενώ από εκεί και πέρα παρουσιάζει αστάθεια. Ο τρόπος που διεξάγει την αρχικοποίηση του, τον βοηθά ακόμα και σε γράφους με μεγαλύτερη πυκνότητα. Ο agglomerative έχει καλύτερο ρυθμό σύγκλισης από τον walkthrough. Εντούτοις, ο walkthrough συγκλίνει στο σημείο k=8, ο agglomerative στο k=9 και ο pckmeans στο k=11. Τα σημεία σύγκλισης των αλγορίθμων αυξήθηκαν καθώς αυξήθηκε δραστικά η πυκνότητα του γράφου και επομένως σχηματίστηκαν περισσότερες συστάδες.

100 κόμβοι

μέσος όρος γειτόνων=1.4, ακτίνα ανίχνευσης r_s=3, παράγοντας ασφάλειας α=0



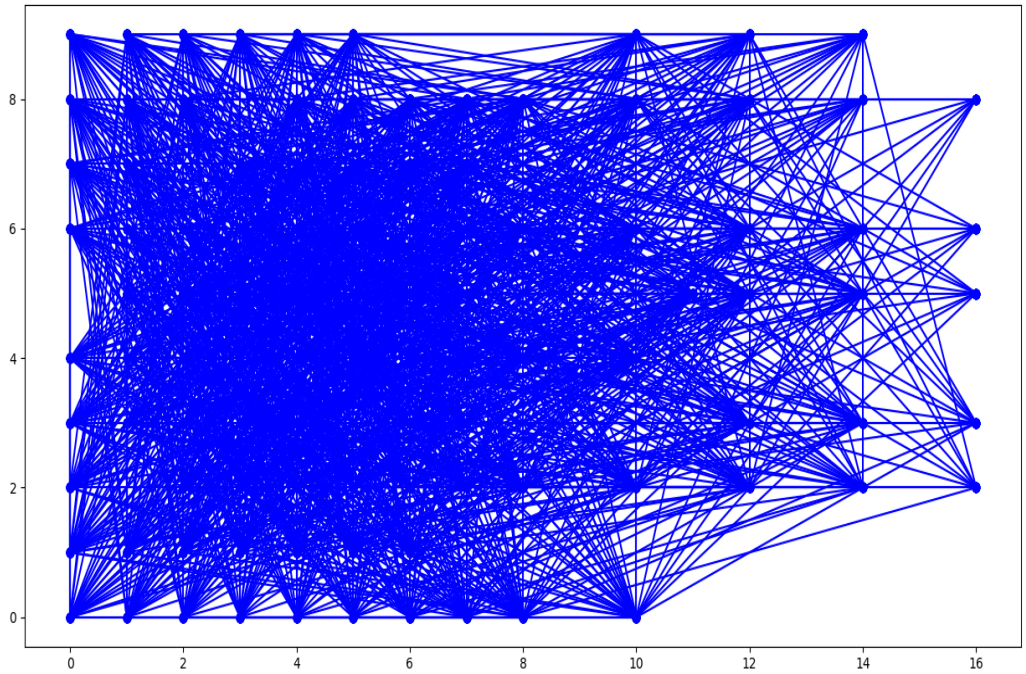
Σχήμα 18: SPG της τοπολογίας των 100 κόμβων με r_s=3 και α=0



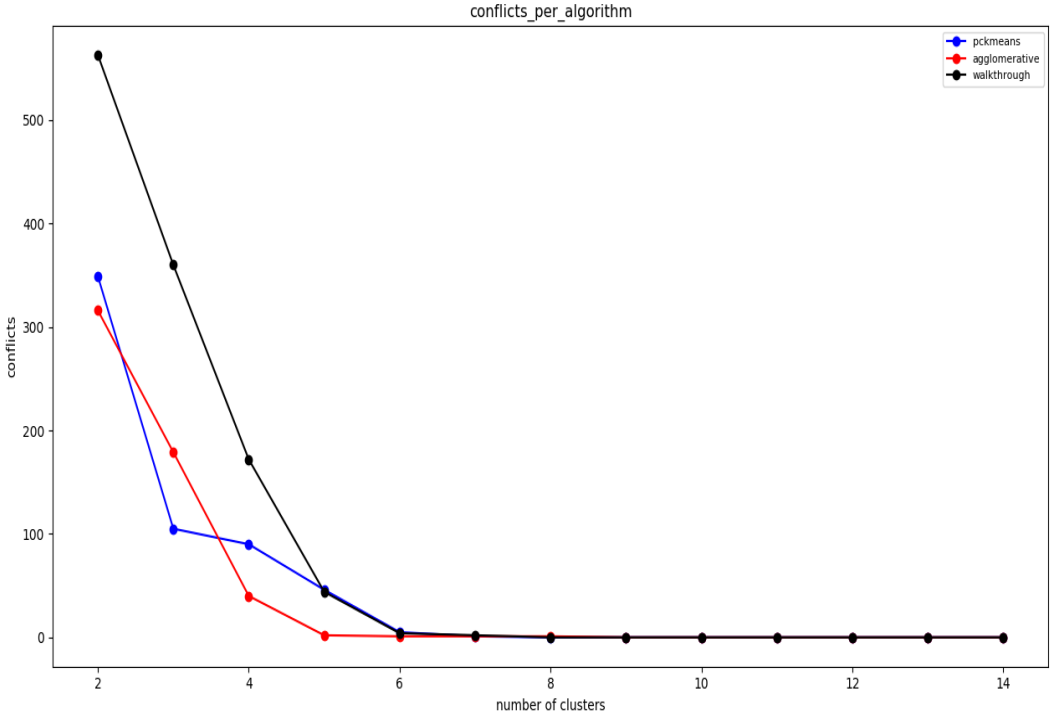
Σχήμα 19: Στατική ανάλυση τοπολογίας 100 κόμβων με r_s=3 και α=0

Παρατηρείται ότι ο αλγόριθμος walkthrough εμφανίζει λιγότερα conflicts από τους άλλους δύο. Εντούτοις, η διαφορά που προκύπτει μεταξύ του walkthrough και του pckmeans έως και k=3 είναι ιδιαιτέρως μικρή. Ο pckmeans από το σημείο k=3 και μετά γίνεται ασταθής. Ο walkthrough συγκλίνει στο k=3, ο agglomerative στο k=5 και ο pckmeans στο k=8. Το αποτέλεσμα αυτό είναι αναμενόμενο καθώς οι γειτονιές πλέον έχουν γίνει αρκετά αραιές και ο walkthrough τις διαχειρίζεται καλύτερα.

μέσος όρος γειτόνων=22.6,ακτίνα ανίχνευσης r_s=4.5,παράγοντας ασφάλειας α=2.6



Σχήμα 20: SPG της τοπολογίας των 100 κόμβων με r_s=4.5 και α=2.6

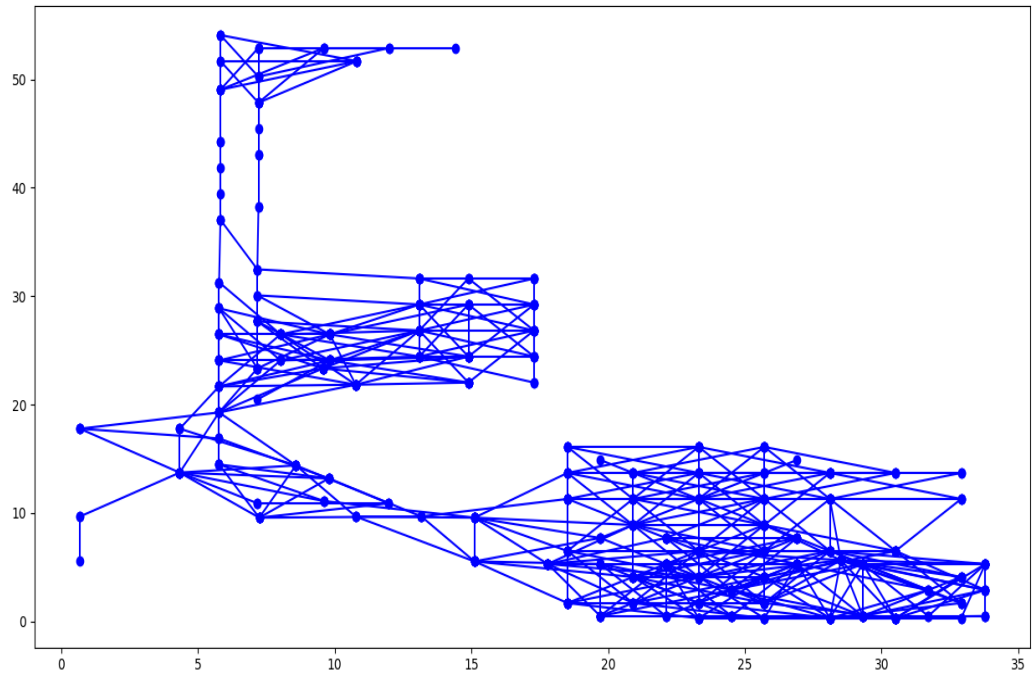


Σχήμα 21: Στατική ανάλυση τοπολογίας 100 κόμβων με r_s=4.5 και α=2.6

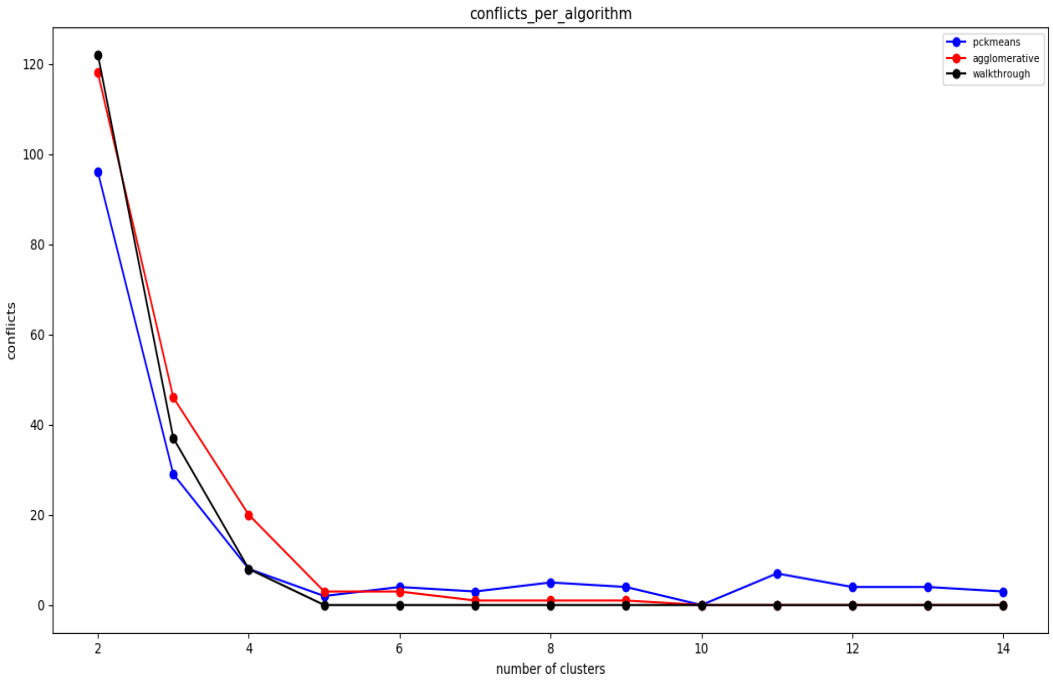
Παρατηρείται ότι ο agglomerative παρουσιάζει καλύτερη σύγκλιση από τους άλλους δύο αλγορίθμους ενώ μόνο ο pckmeans επιτυγχάνει λιγότερα conflicts για k=3. Το σημείο σύγκλισης του agglomerative είναι το k=6, του walkthrough το k=8 ενώ και του pckmeans το k=8. Ο γράφος έχει αρκετά μεγάλη πυκνότητα κάτι που οδηγεί σε χειροτέρευση της επίδοσης του walkthrough και αύξηση του σημείου σύγκλισης των αλγορίθμων. Ενδιαφέρον παρουσιάζει το γεγονός ότι ο pckmeans δεν έχει καθόλου αστάθεια.

123 κομβοι

μέσος όρος γειτόνων=5.2,ακτίνα ανίχνευσης r_s=3,παράγοντας ασφάλειας α=2



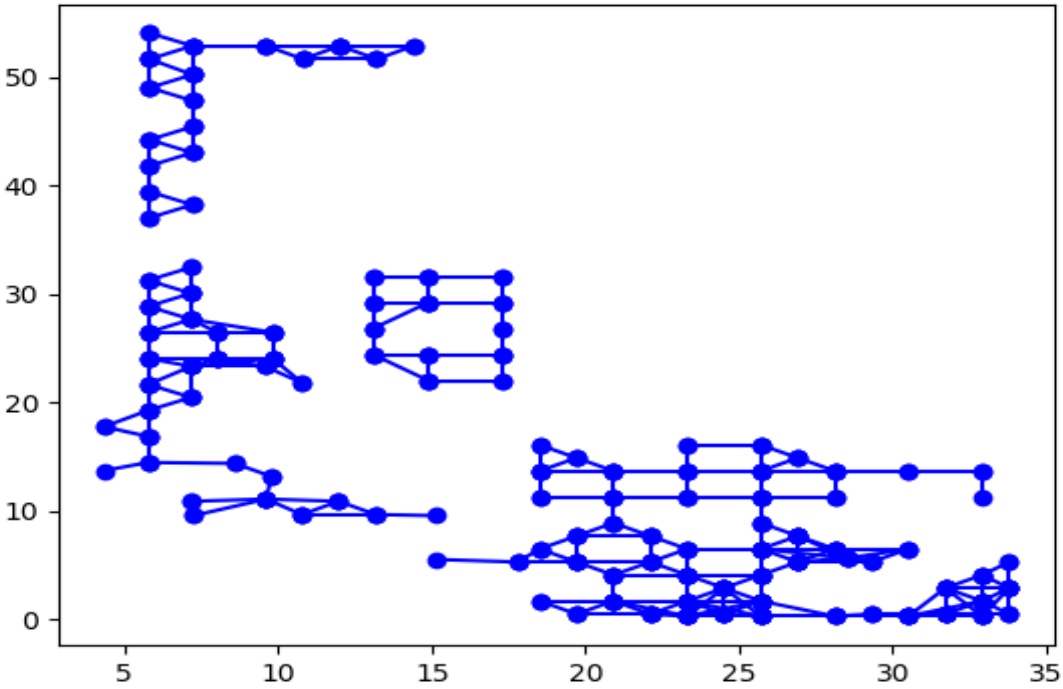
Σχήμα 22: SPG της τοπολογίας των 123 κόμβων με r_s=3 και α=2



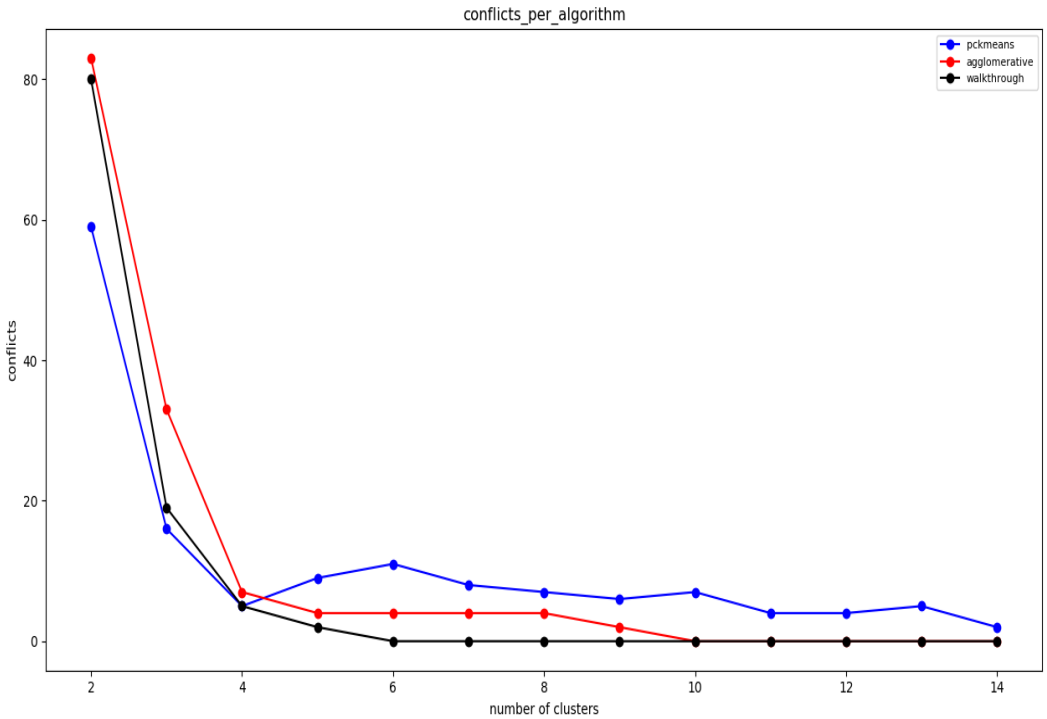
Σχήμα 23: Στατική ανάλυση τοπολογίας 123 κόμβων με r_s=3 και α=2

Ο αλγόριθμος pckmeans επιτυγχάνει λιγότερα σφάλματα έως και το σημείο k=4 ενώ ο walkthrough παρουσιάζει καλύτερη σύγκλιση σε σχέση με τον agglomerative. Ο walkthrough συγκλίνει στο k=5 ενώ ο agglomerative στο k=10 όπως και ο pckmeans ο οποίος παρουσιάζει αστάθεια από εκείνο το σημείο και έπειτα.

μέσος όρος γειτόνων=3.5,ακτίνα ανίχνευσης r_s=1.5,παράγοντας ασφάλειας α=2



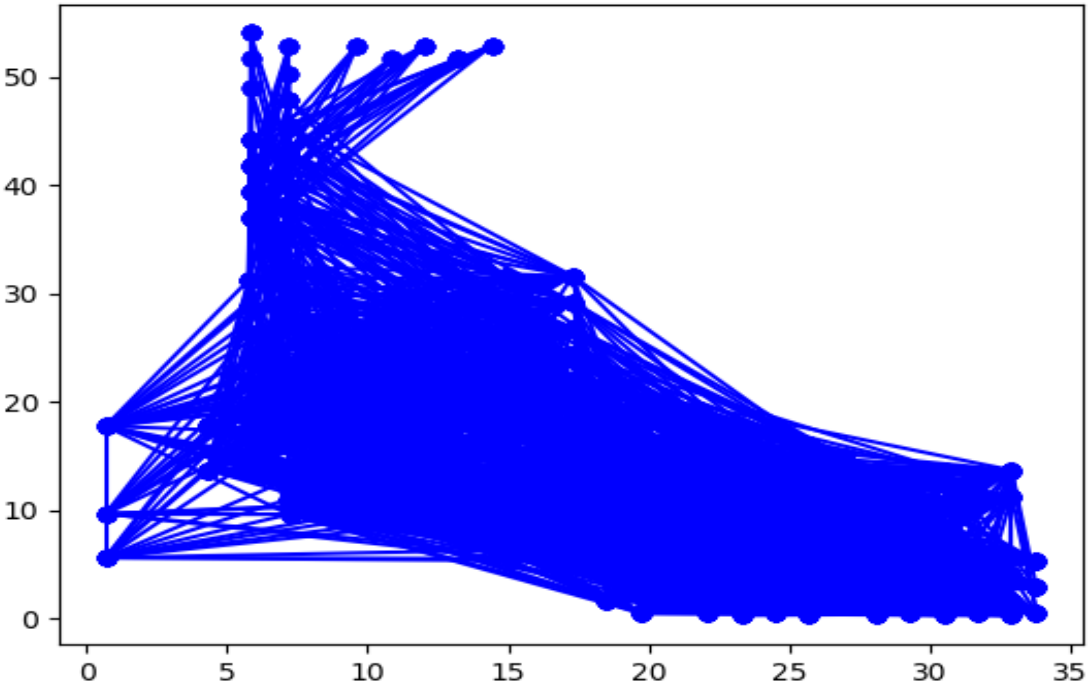
Σχήμα 24: SPG της τοπολογίας των 123 κόμβων με r_s=1.5 και α=2



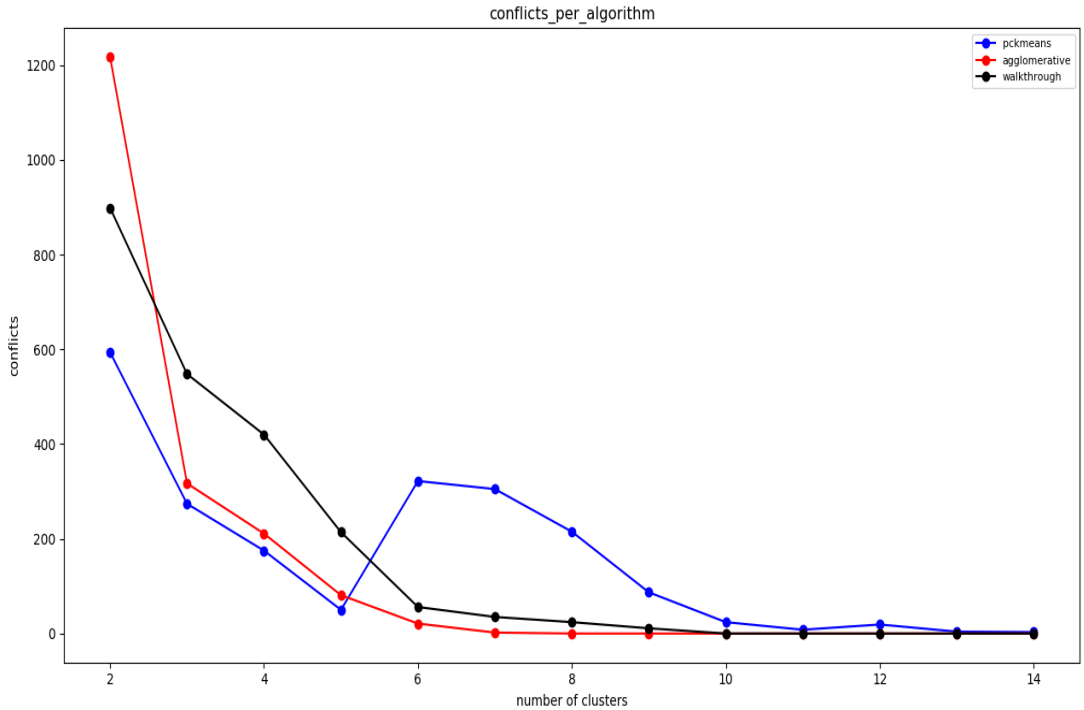
Σχήμα 25: Στατική ανάλυση τοπολογίας 123 κόμβων με r_s=1.5 και α=2

Παρατηρείται ότι ο pckmeans παρουσιάζει λιγότερα σφάλματα έως και το σημείο k=4. Ο walkthrough έχει καλύτερο ρυθμό σύγκλισης από τον agglomerative. Ο walkthrough συγκλίνει στο k=6 ενώ ο agglomerative στο k=10. Ο pckmeans δεν συγκλίνει έως και το σημείο k=14 αλλά ενδέχεται να το κάνει σε κάποιο μεταγενέστερο. Στην συγκεκριμένη τοπολογία βλέπουμε ότι ο agglomerative δυσκολεύεται να βρει το σημείο σύγκλισης όσο η πυκνότητα είναι αρκετά πιο μικρή από το πλήθος των servers.

μέσος όρος γειτόνων=29.9,ακτίνα ανίχνευσης r_s=9,παράγοντας ασφάλειας α=8.2



Σχήμα 26: SPG της τοπολογίας των 123 κόμβων με r_s=9 και α=8.2



Σχήμα 27: Στατική ανάλυση τοπολογίας 123 κόμβων με r_s=9 και α=8.2

Παρατηρείται ότι ο pckmeans επιτυγχάνει λιγότερα conflicts έως και το σημείο k=5 ενώ από εκεί και έπειτα παρουσιάζει αστάθεια. Ο ρυθμός σύγκλισης του agglomerative είναι πολύ ταχύτερος από εκείνον του walkthrough. Ο agglomerative συγκλίνει στο σημείο k=7, ο walkthrough στο σημείο k=10 και ο pckmeans στο k=14. Η απότομη αύξηση της πυκνότητας του γράφου περίπου στο 1/4 του πλήθους των servers οδήγησε σε θεαματική βελτίωση του agglomerative έναντι του walkthrough ο οποίος ιδιαίτερα έως και k=5 επιτύγχανε αρκετά περισσότερα conflicts. Ιδιαίτερα σε τόσο πυκνές τοπολογίες φαίνονται τα πλεονεκτήματα της αρχικοποίησης του pckmeans. Για k=2 πετυχαίνει έως και 600 conflicts λιγότερα από τον agglomerative ενώ έως k=5 έχει 300 λιγότερα conflicts από τον walkthrough.

Συμπεράσματα

Ο κάθε αλγόριθμος παρουσιάζει πλεονεκτήματα και μειονεκτήματα ανάλογα με τον αριθμό των κόμβων αποθήκευσης, την πυκνότητα του γράφου, που ισοδυναμεί με το μέσο πλήθος γειτόνων ενός κόμβου αποθήκευσης, και την φύση της τοπολογίας. Σημαντική παρατήρηση αποτελεί όχι μόνο το σημείο στο οποίο οι αλγόριθμοι επιλύουν πλήρως το πρόβλημα αλλά και η ταχύτητα σύγκλισης τους μέχρι το σημείο αυτό. Όσο λιγότερα conflicts δημιουργούνται στο δίκτυο, τόσο περισσότερο αυξάνεται και το επίπεδο της ασφάλειας.

Κατόπιν της ανάλυσης των αλγορίθμων που διεξήχθη, καταλήξαμε σε ορισμένα συμπεράσματα. Αρχικά, ο αλγόριθμος pckmeans λειτουργεί, σχεδόν σε όλες τις περιπτώσεις, καλύτερα από τους άλλους δύο όταν το δίκτυο αποτελείται από μικρό πλήθος κόμβων αποθήκευσης και κυρίως για k<4. Αυτό είναι πολύ σημαντικό ιδιαίτερα όσον αφορά το κόστος καθώς η παράταξη νέων κόμβων αποθήκευσης επιβαρύνει τον προϋπολογισμό ενός δικτύου. Εντούτοις, λόγω της τυχαιότητας που χρησιμοποιεί στην αρχικοποίηση των κέντρων του, παρουσιάζει μεγάλη αστάθεια από εκείνο το σημείο και έπειτα κρίνοντας τον μη αξιόπιστο όσο μεγαλώνει το πλήθος των κόμβων αποθήκευσης. Ο αλγόριθμος walkthrough, για δίκτυα που έχουν σχετικά μικρό μέσο όρο πλήθος γειτόνων κόμβων ανίχνευσης σε σχέση με το πλήθος τους, παρουσιάζει καλύτερη σύγκλιση από τους άλλους δύο επιτυγχάνοντας λιγότερα conflict στην πλειοψηφία των περιπτώσεων. Ωστόσο ιδιαίτερα σε περιπτώσεις που ο γράφος γίνεται πολύ πυκνός και αυξάνει ο μέσος όρος γειτόνων, η επίδοση αυτού του αλγορίθμου γίνεται αρκετά χειρότερη σε σχέση με τους άλλους δύο. Επιπλέον, το επίπεδο τυχαιότητας είναι αρκετά υψηλό καθώς το πλήθος των conflict που δημιουργεί μπορεί να ποικίλει σημαντικά ανάμεσα σε διαδοχικές εκτελέσεις του αλγορίθμου ακόμα και αν διατηρείται σταθερό το πλήθος των κόμβων αποθήκευσης. Ο αλγόριθμος constrained agglomerative clustering δεν παρουσιάζει κάποιο συγκεκριμένο πλεονέκτημα έναντι των άλλων δύο, ωστόσο αποτελεί έναν σταθερό αλγόριθμο δίχως τυχαιότητα που δημιουργεί είτε καλύτερο αριθμό conflict από τους άλλους δύο είτε το διατηρεί σε ένα αποδεκτό επίπεδο.

Τέλος, ο αριθμός των conflict, που δημιουργούν οι αλγόριθμοι, αποτελεί πολύ σημαντικός παράγοντας καθώς έχει μεγάλη επίπτωση στα επίπεδα κατανάλωσης ενέργειας του δικτύου. Όσο μικρότερος είναι ο αριθμός των conflict, τόσο μικρότερος είναι ο αριθμός των κόμβων ανίχνευσης που επισυνάπτονται στο χρώμα n+1 και άρα τόσο μικρότερος είναι ο αριθμός των μηνυμάτων “εκδήλωσης ενδιαφέροντος” που ανταλλάσσονται μεταξύ τους.

6.5 Πειραματική αξιολόγηση δυναμικής ανάλυσης

Για την πειραματική αξιολόγηση της δυναμικής ανάλυσης χρησιμοποιήθηκαν οι ίδιες τοπολογίες, που παραθέσαμε στο κεφάλαιο 6.4, εισάγοντας όμως και κινητούς στόχους οι οποίοι ακολουθούσαν προκαθορισμένες τροχιές που παρέχονταν από το FIT IoT LAB. Για την αξιολόγηση της δυναμικής ανάλυσης έγινε σύγκριση του αλγορίθμου Δυναμικού Χρωματισμού και του walkthrough μέσω γραφήματος P-t όπου στον άξονα y φαίνεται το P(t) (privacy), δηλαδή το επίπεδο ασφάλειας τη χρονική στιγμή t και στον άξονα x η αντίστοιχη χρονική στιγμή t. Μέσω του FIT IoT LAB λαμβάνουμε μια δειγματοληψία της τροχιάς του κινητού στόχου ανά 0,1 sec και τις αντίστοιχες x,y συντεταγμένες. Μια επιπλέον μετρική που χρησιμοποιείται για την σύγκριση μεταξύ των δύο αλγορίθμων είναι η εξής:

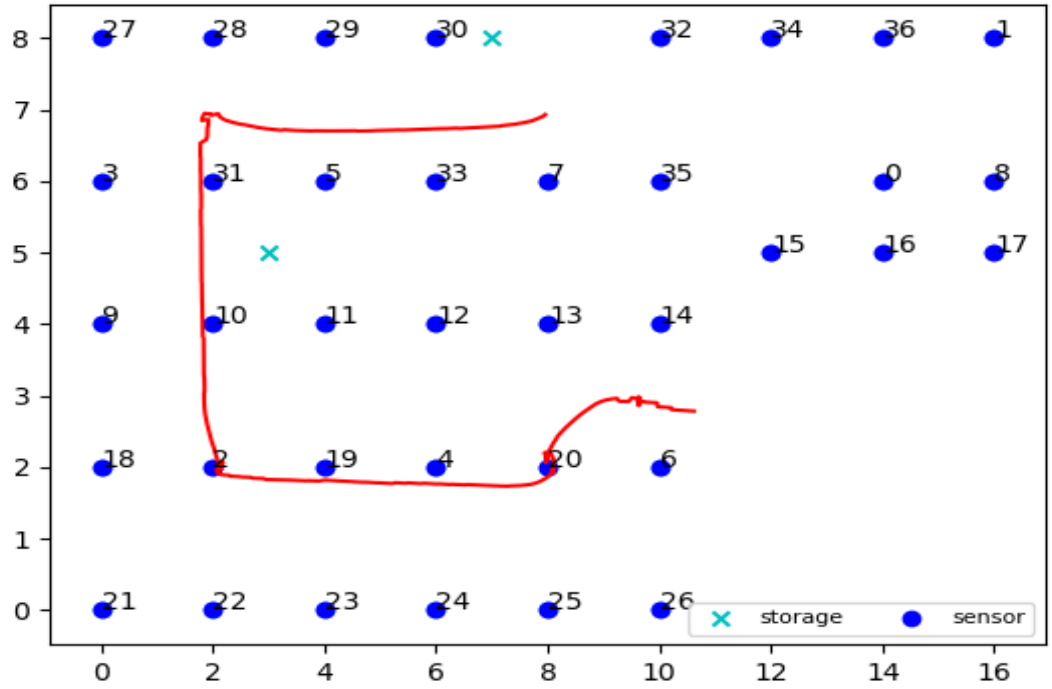
$$Privacy_t = \sum_{t_1=0}^{t_2=T} \frac{P_{t_1}}{t_2 - t_1} \qquad \text{Σχέση 6.α}$$

Όπου Pt_l η στιγμιαία ασφάλεια που δείχνει το επίπεδο της ιδιωτικότητας τη συγκεκριμένη χρονική στιγμή t_l όπως ορίστηκε και από την Σχέση 4.α.

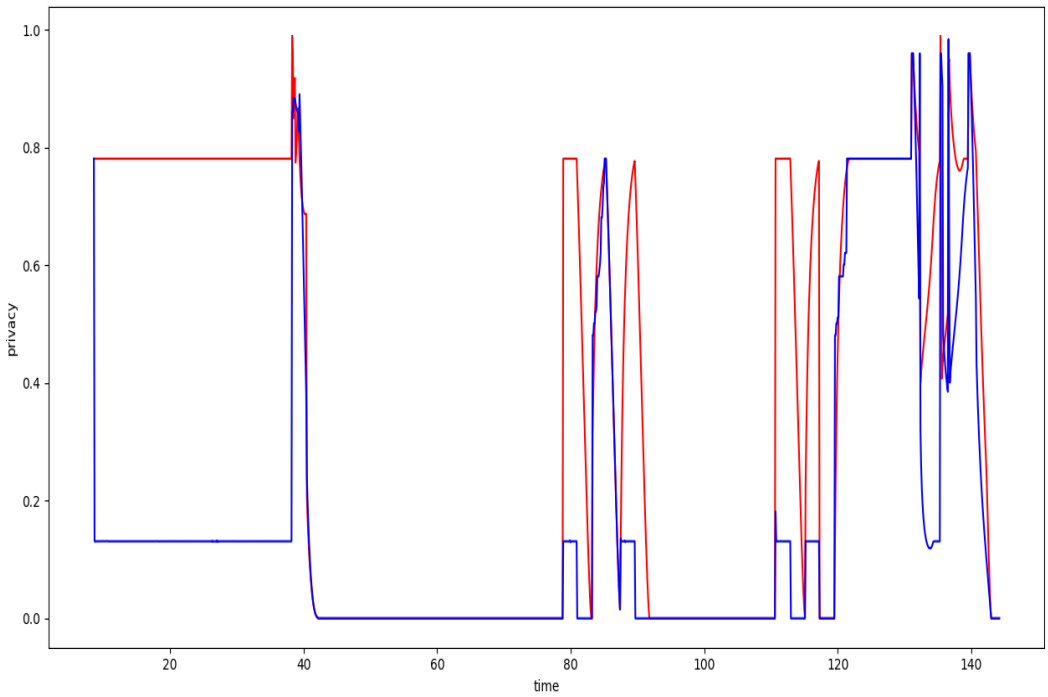
Η παράθεση των αποτελεσμάτων γίνεται κατηγοριοποιημένη ανά τοπολογία. Διεξήχθησαν 12 πειράματα δυναμικής ανάλυσης των τοπολογιών. Σε κάθε πείραμα έχουμε τροποποίηση είτε του αριθμού των κόμβων αποθήκευσης (servers) είτε της ακτίνας ανίχνευσης r_s και του παράγοντα ασφάλειας α. Οι παραμετροποιήσεις αυτές ήταν απαραίτητες προκειμένου να διαπιστωθεί πως αντιδρά ο Αλγόριθμος Δυναμικού Χρωματισμού όσο αυξάνονται τα σημεία αποθήκευσης του δικτύου καθώς και η πυκνότητά του. Έτσι εξετάστηκε ο προτεινόμενος αλγόριθμος σε περιβάλλοντα με διαφορετικές συνθήκες. Η κόκκινη γραμμή των γραφημάτων δείχνει τον Αλγόριθμο Δυναμικού Χρωματισμού ενώ η μπλε γραμμή τον αλγόριθμο walkthrough.

37 κόμβοι

αριθμός server=2,ακτίνα ανίχνευσης r_s=1.5,παράγοντας ασφάλειας α=2,τροχιά=square_1



Σχήμα 28: 37 κόμβοι με 2 servers και τροχιά square_1

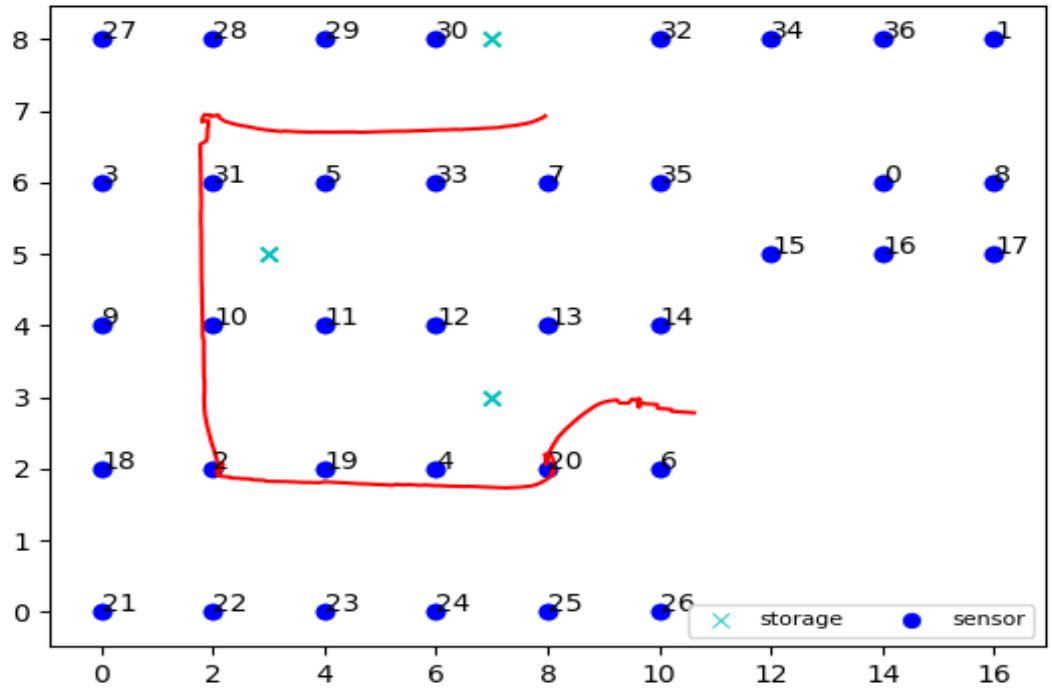


Σχήμα 29: Δυναμική ανάλυση τοπολογίας 37 κόμβων με 2 servers και τροχιά square_1

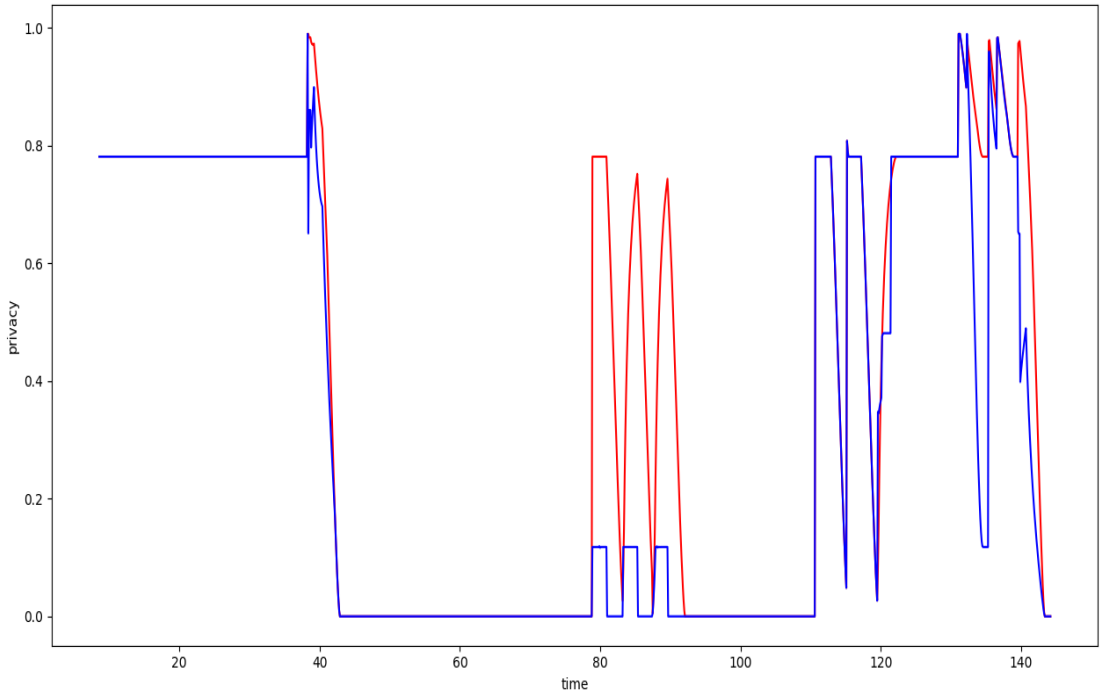
Παρατηρείται ότι καθ' όλη την χρονική διάρκεια του πειράματος ο ΑΔΧ (κόκκινη γραμμή) επιτυγχάνει μεγαλύτερο privacy στο δίκτυο από ότι ο walkthrough (μπλε γραμμή). Σε αρκετά χρονικά διαστήματα μάλιστα όπως πχ μεταξύ 10-40 sec, 80-90sec και 110-120 η αύξηση ξεπερνά

το 60%. Κάθε χρονική στιγμή το privacy του ΑΔΧ είναι καλύτερο από του walkthrough καθώς ο ΑΔΧ χρωμάτισε τους κόμβους που απέχουν λιγότερο από $2r_s$-a με αποδοτικό τρόπο.

αριθμός server=3,ακτίνα ανίχνευσης r_s=1.5,παράγοντας ασφάλειας α=2,τροχιά=square_1



Σχήμα 30: 37 κόμβοι με 3 servers και τροχιά square_1

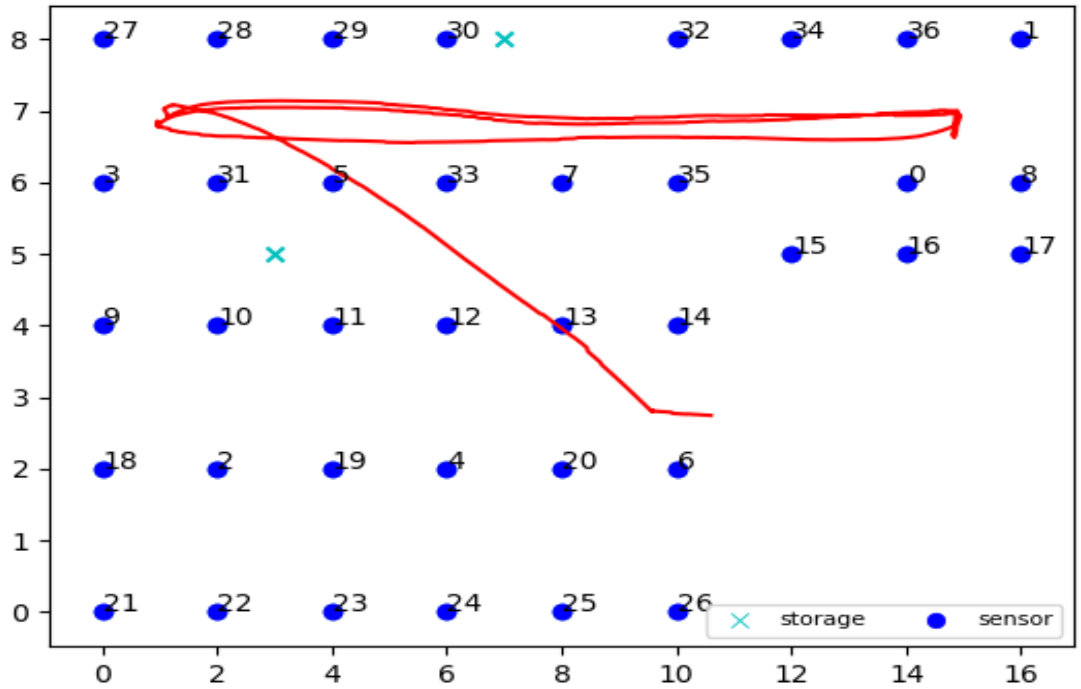


Σχήμα 31: Δυναμική Ανάλυση τοπολογίας 37 κόμβων με 3 servers και τροχιά square_1

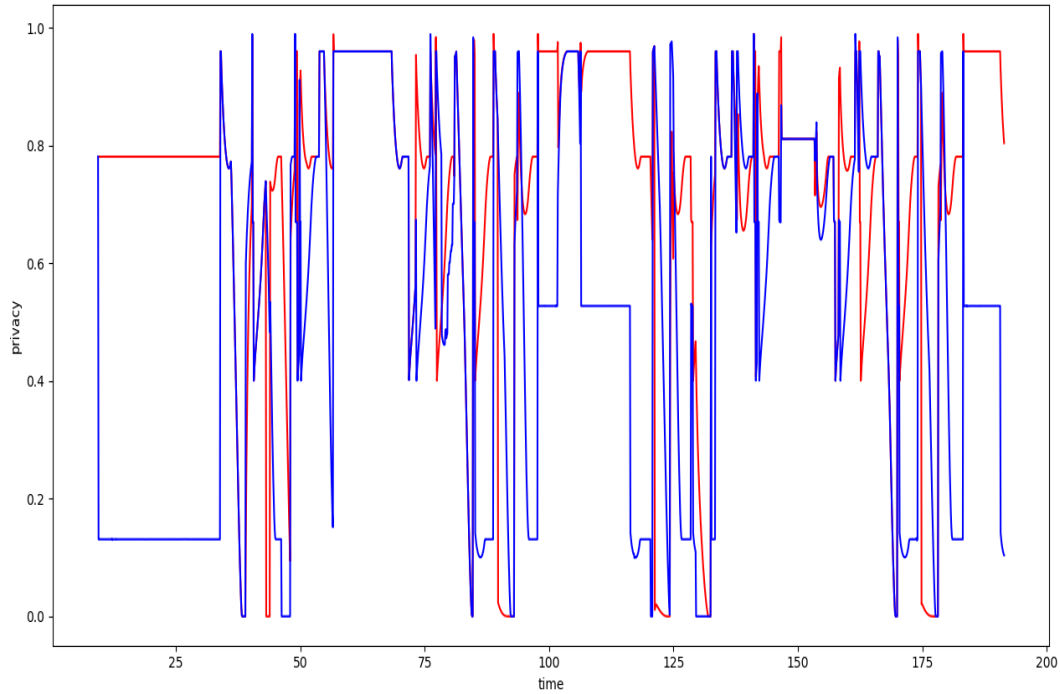
Παρατηρείται ότι ο ΑΔΧ (κόκκινη γραμμή) είναι καλύτερος καθ' όλη τη χρονική διάρκεια του πειράματος από τον wakthrough (μπλε γραμμή). Μάλιστα υπάρχουν διαστήματα που η αύξηση του privacy αγγίζει το 70% όπως για παράδειγμα μεταξύ 80-90 sec. Παρατηρείται ότι με την

αύξηση των servers ο walkthrough σε μεγάλα διαστήματα βελτίωσε την απόδοση του όμως και πάλι ο ΑΔΧ επιτυγχάνει ίσο η μεγαλύτερο αποτέλεσμα.

αριθμός server=2,ακτίνα ανίχνευσης r_s=1.5,παράγοντας ασφάλειας α=2,τροχιά=h_line_2



Σχήμα 32: 37 κόμβοι με 2 servers και τροχιά h_line_2

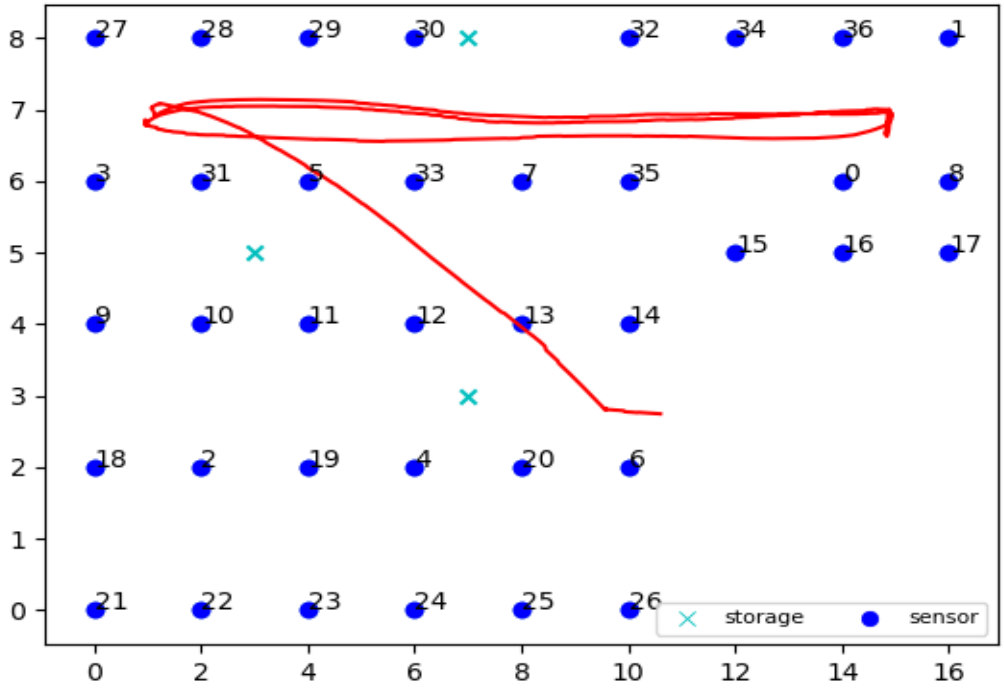


Σχήμα 33: Δυναμική Ανάλυση τοπολογίας 37 κόμβων με 2 servers και τροχιά h_line_2

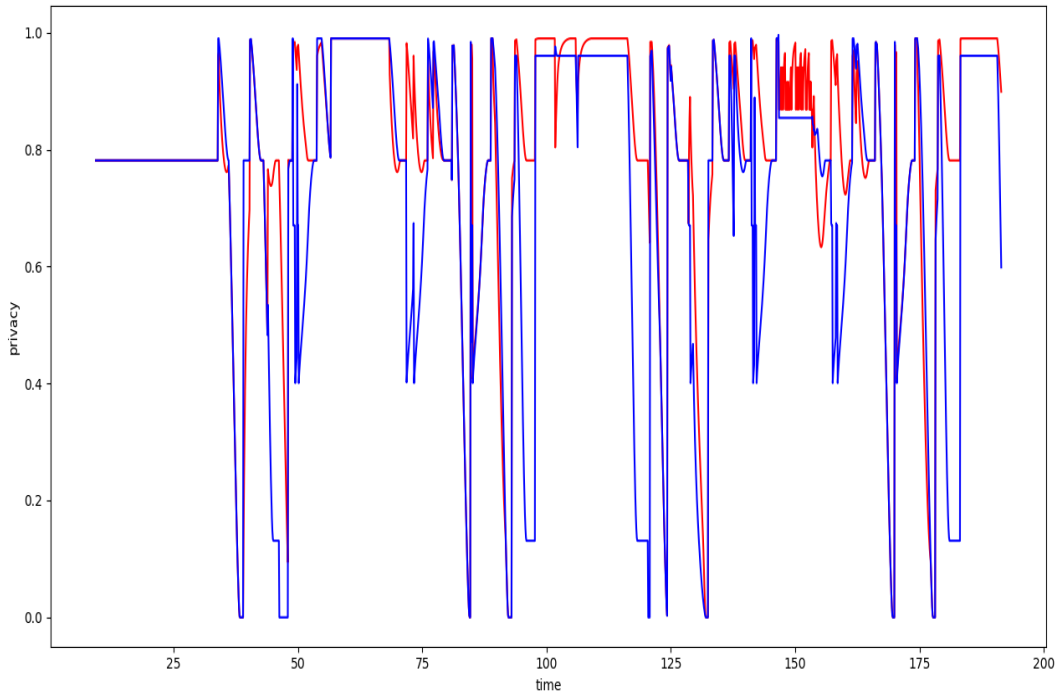
Παρατηρείται ότι ο ΑΔΧ (κόκκινη γραμμή) επιτυγχάνει υψηλότερο privacy από τον walkthrough (μπλε γραμμή) σε όλη την χρονική διάρκεια εκτός από ορισμένα διαστήματα ολίγων δευτερολέπτων. Στο διάστημα από 10-30 sec παρατηρείται αύξηση περίπου 65% στο privacy ενώ

ενδιαφέρον παρουσιάζει το γεγονός ότι στα διαστήματα 105-125sec και 180-195sec έχουμε αύξηση περίπου 50% με το privacy για αρκετό διάστημα να αγγίζει και το 96%. Αυτό συμβαίνει διότι στις συγκεκριμένες περιοχές ο walkthrough, με τον τυχαίο χρωματισμό που κάνει όταν δεν υπάρχουν ελεύθερα χρώματα, επισύναψε γειτονικούς κόμβους στο ίδιο χρώμα.

αριθμός server=3,ακτίνα ανίχνευσης r_s=1.5,παράγοντας ασφάλειας α=2,τροχιά=h_line_2



Σχήμα 34: 37 κόμβοι με 3 servers και τροχιά h_line_2



Σχήμα 35: Δυναμική Ανάλυση τοπολογίας 37 κόμβων με 3 servers και τροχιά h_line_2

Παρατηρείται ότι παρόλο που ο αλγόριθμος walkthrough (μπλε γραμμή) παράγει ένα πολύ καλό μέσο privacy περίπου 0,7 , ο ΑΔΧ (κόκκινη γραμμή) το βελτιώνει κι άλλο φτάνοντας το μέσο privacy περίπου στο 0,79. Αυτό το αποτέλεσμα επιτυγχάνεται με την βελτίωση του privacy ακόμα

και στα διαστήματα που ο walkthrough παρουσιάζει την μέγιστη τιμή του αλλά κυρίως εξομάλυνση των διαστημάτων που το privacy του walkthrough φτάνει χαμηλότερα του 0,35. Η τεχνική του χρωματισμού των κόμβων που δημιουργούν τα conflicts από τον ΑΔΧ εξομάλυνε τις απότομες μεταπτώσεις στην απόδοση του walkthrough.

Privacy	square_1	h_line_2
2 storage nodes	0.1643	0.5108
3 storage nodes	0.3489	0.7018

Πίνακας 1: Privacy walkthrough

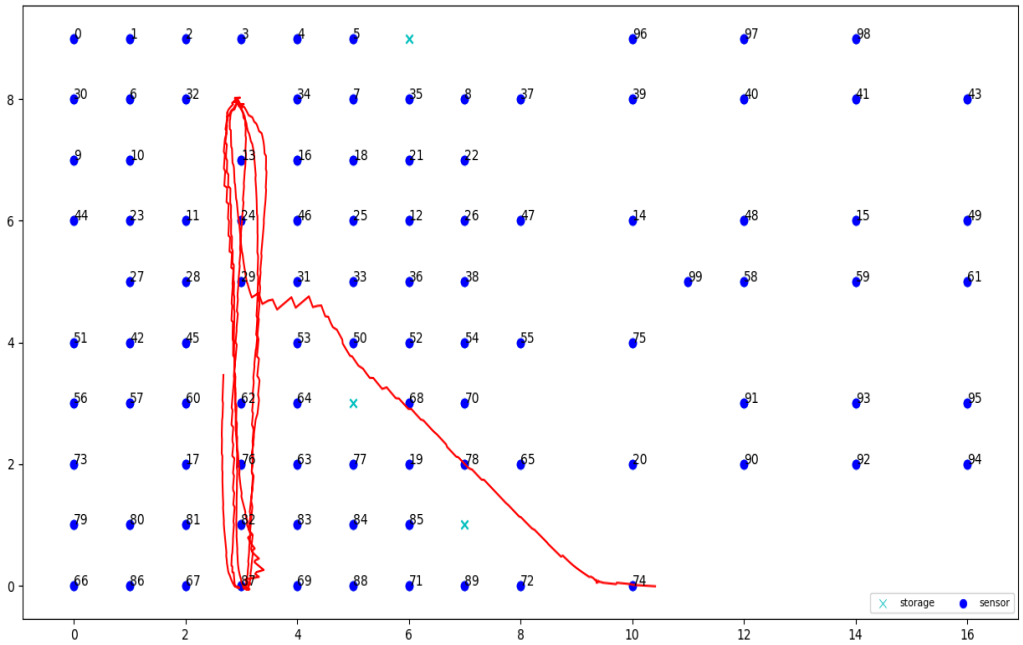
Privacy	square_1	h_line_2
2 storage nodes	0.3830	0.7180
3 storage nodes	0.4108	0.7898

Πίνακας 2: Privacy Αλγορίθμου Δυναμικού Χρωματισμού

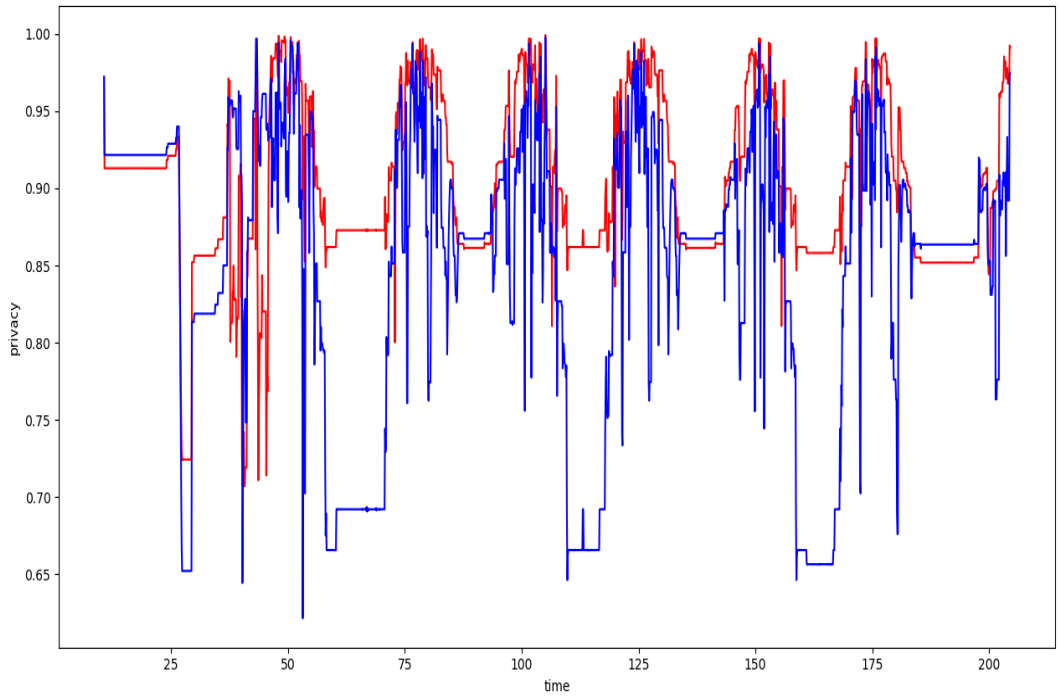
Στους παραπάνω πίνακες φαίνονται τα επίπεδα privacy που επιτυγχάνουν οι δύο αλγόριθμοι στα σε πειράματα με ίδιες συνθήκες στην τοπολογία των 37 κόμβων ανίχνευσης. Στις γραμμές δηλώνεται το πλήθος των servers του πειράματος ενώ στις στήλες η τροχιά του κινητού στόχου. Γίνεται έτσι σαφής η βελτίωση που επιτυγχάνεται με τον ΑΔΧ έναντι του walkthrough. Στο μεγαλύτερο μέρος των πειραμάτων έχουμε μία σταθερή αύξηση του privacy από 8-20%. **Πολύ σημαντικό αποτελεί το γεγονός ότι όσο αυξάνονται οι servers, η απόδοση του ΑΔΧ γίνεται ακόμα καλύτερη καθώς επιτυγχάνεται ολοένα και μεγαλύτερο privacy. Επίσης, βασική βελτίωση αποτελεί το αυξημένο επίπεδο privacy ακόμα και όταν υπάρχουν αρκετά λίγοι servers στο δίκτυο, διατηρώντας το περίπου στο 38% έναντι του 16% που επιτυγχάνει ο walkthrough**. Ακόμα και με ελάχιστο πλήθος servers, μπορούμε να διατηρήσουμε ένα καλό επίπεδο ασφάλειας μειώνοντας έτσι το κόστος της υποδομής.

100 κόμβοι

αριθμός server=3,ακτίνα ανίχνευσης r_s=4.5,παράγοντας ασφάλειας α=2.6,τροχιά=v_line_3



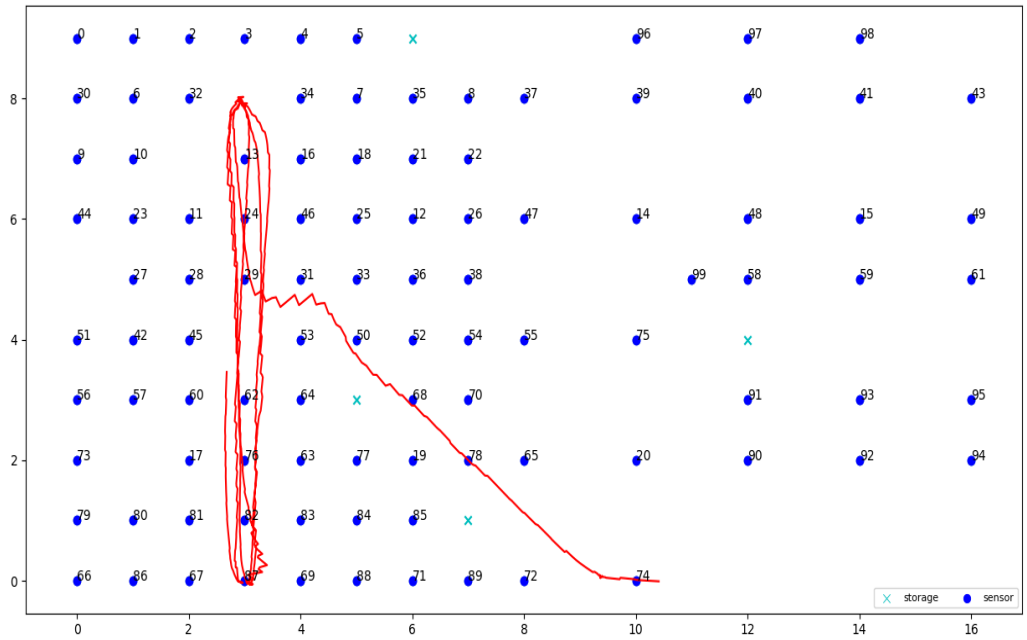
Σχήμα 36: 100 κόμβοι με 3 servers και τροχιά v_line_3



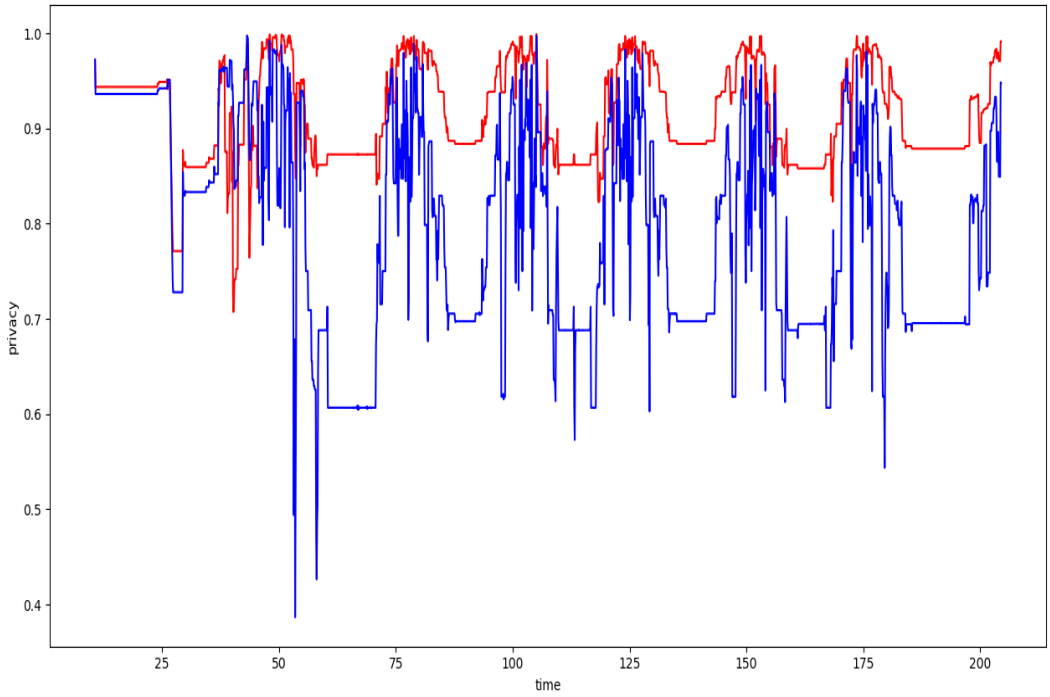
Σχήμα 37: Δυναμική Ανάλυση τοπολογίας 100 κόμβων με 3 servers και τροχιά v_line_3

Παρατηρείται ότι ο ΑΔΧ (κόκκινη γραμμή) επιτυγχάνει σε αρκετά μεγάλα διαστήματα όπως από 55-75sec, 110-125sec, 155-175sec αύξηση του privacy περίπου κατά 30%. Φαίνεται ότι υπάρχουν ορισμένα μικρά διαστήματα όπως από 80-85sec, 130-135sec, 190-200sec που το privacy του ΑΔΧ είναι κατά 1-3% μικρότερο. Αυτό οφείλεται αφενός στην ακρίβεια της διαδρομής του κινητού στόχου ανάμεσα σε διαδοχικές εκτελέσεις της διαδρομής και αφετέρου στο γεγονός ότι ο walkthrough (μπλε γραμμή) στην συγκεκριμένη εκτέλεση πέτυχε καλύτερο χρωματισμό των κόμβων που απέχουν λιγότερο από το εύρος που ορίζει ο SPG.

αριθμός server=4,ακτίνα ανίχνευσης r_s=4.5,παράγοντας ασφάλειας α=2.6,τροχιά=v_line_3



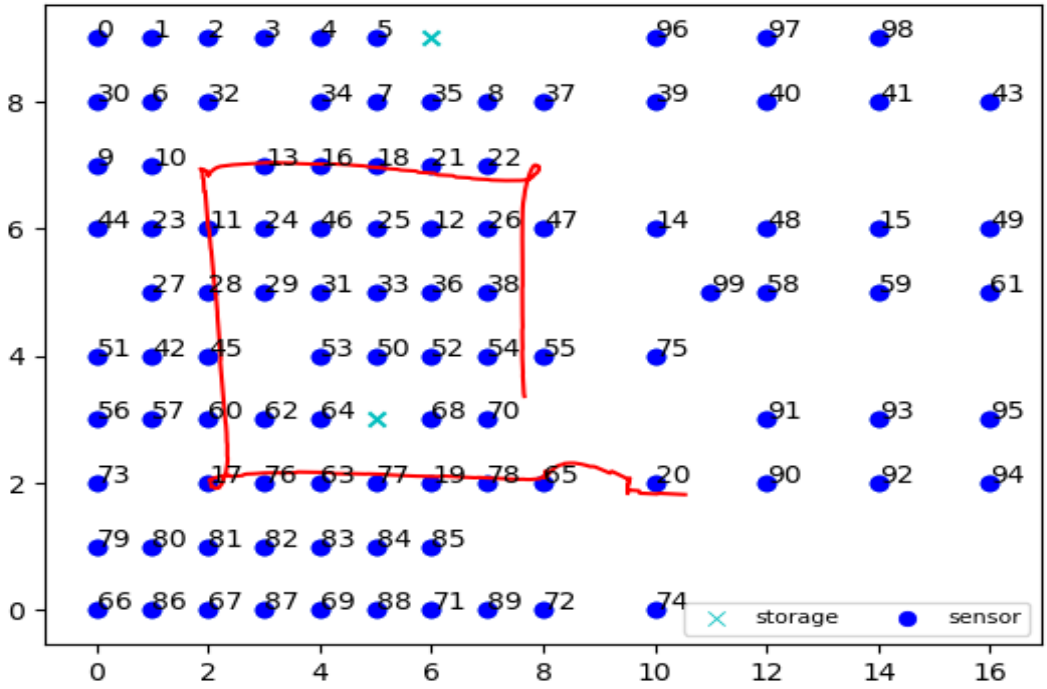
Σχήμα 38: 100 κόμβοι με 4 servers και τροχιά v_line_3



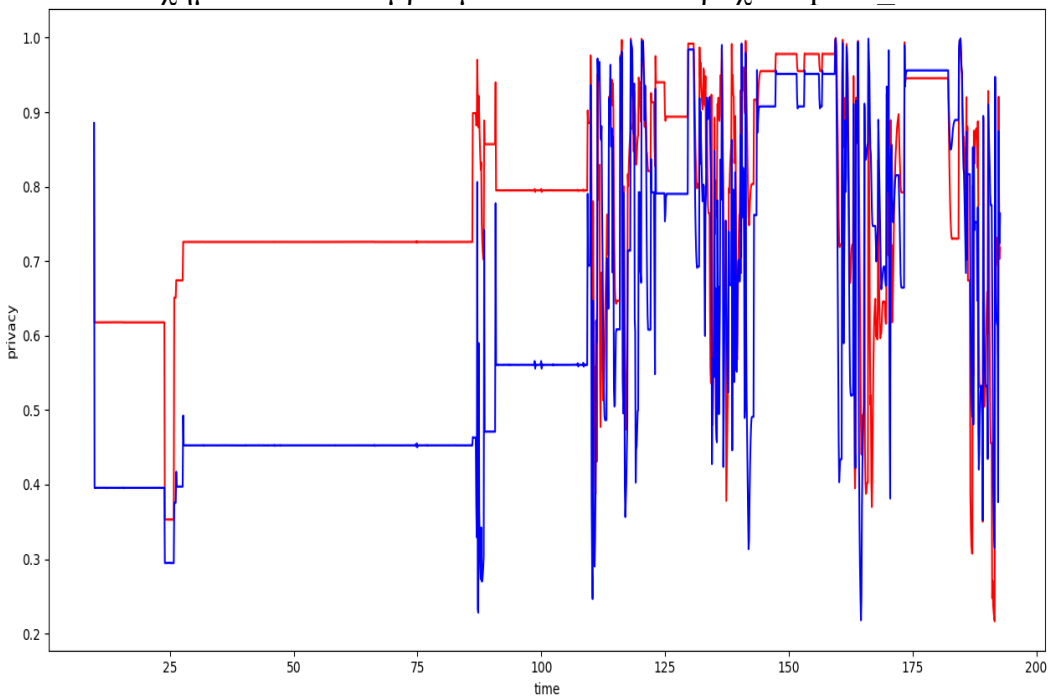
Σχήμα 39:Δυναμική Ανάλυση τοπολογίας 100 κόμβων με 4 servers και τροχιά v_line_3

Παρατηρείται ότι με την αύξηση των servers έχουμε θεαματική βελτίωση του privacy που δημιουργεί ο ΑΔΧ (κόκκινη γραμμή) σε σχέση με τον walkthrough (μπλε γραμμή). Από τη χρονική στιγμή 60sec και μετά υπάρχει αύξηση κατά 25-30% στο μεγαλύτερο διάστημα έως το τέλος του πειράματος εκτοξεύοντας το privacy από 0,79 σε 0,92. Παρατηρείται έτσι η αναλογική σχέση μεταξύ του privacy και του πλήθους των servers.

αριθμός server=2,ακτίνα ανίχνευσης r_s=3,παράγοντας ασφάλειας α=0,τροχιά=square_1



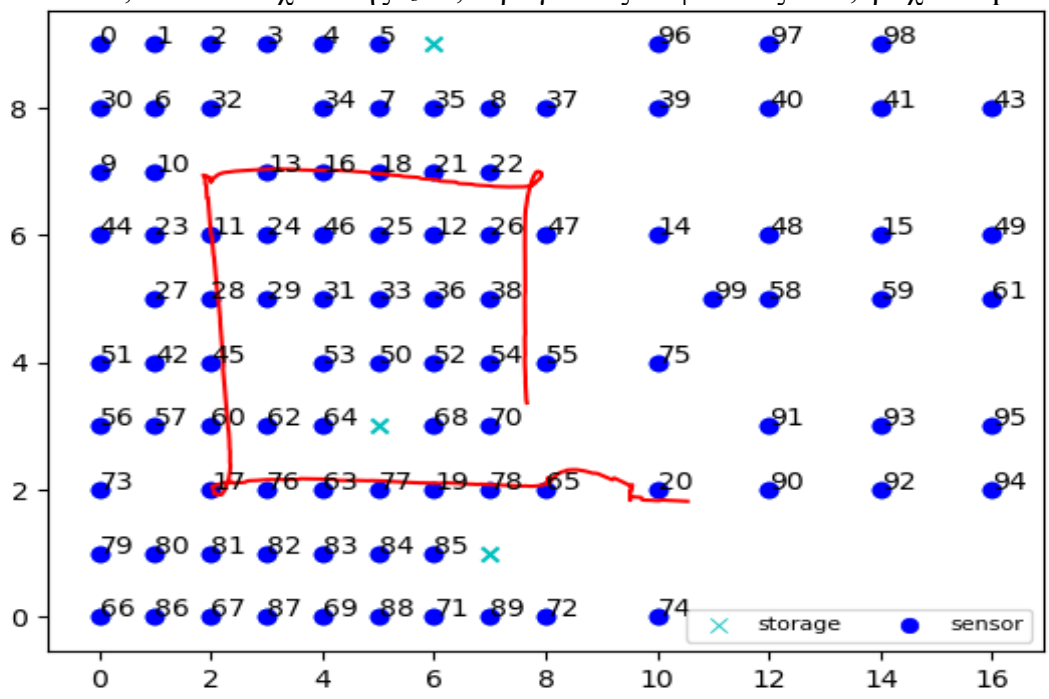
Σχήμα 40: 100 κόμβοι με 2 servers και τροχιά square_1



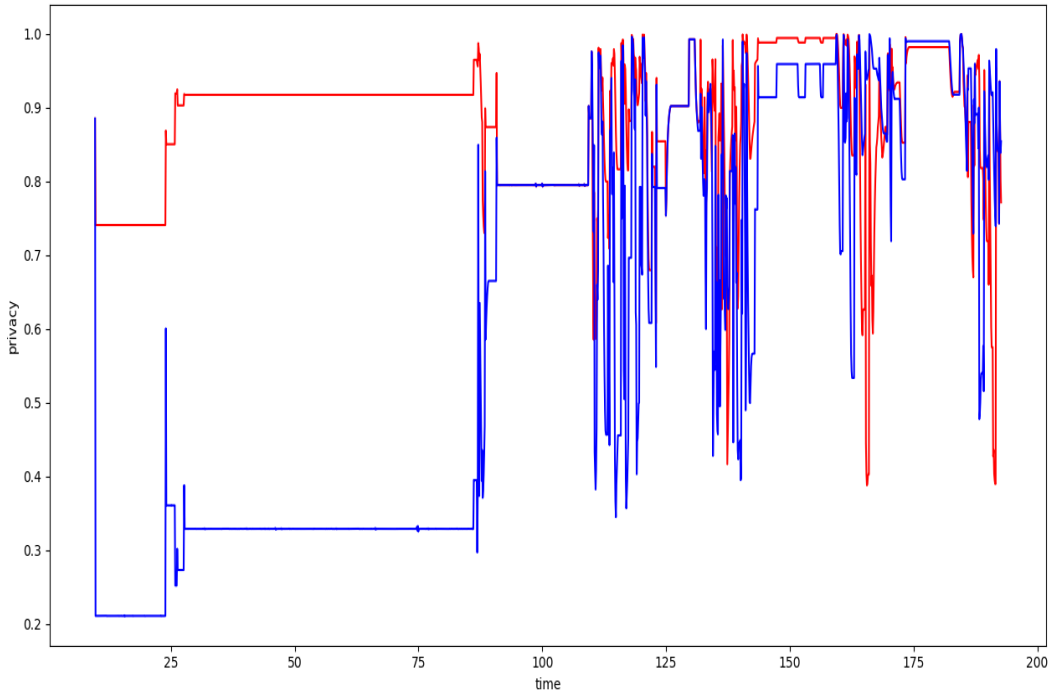
Σχήμα 41: Δυναμική Ανάλυση τοπολογίας 100 κόμβων με 2 servers και τροχιά square_1

Παρατηρείται σε αυτήν την περίπτωση μια αύξηση του privacy της τάξης του 20-25% για σχεδόν τη μισή χρονική διάρκεια του πειράματος από 10-110sec. Επιπλέον για το δεύτερο μισό του πειράματος το privacy διατηρείται στα ίδια και ακόμα υψηλότερα επίπεδα με εξαίρεση ορισμένα διαστήματα ολίγων δευτερολέπτων. Αυτό οφείλεται στην ακρίβεια του κινητού στόχου ανάμεσα σε διαδοχικές εκτελέσεις της ίδιας διαδρομής και τον χρωματισμό των κόμβων που απέχουν λιγότερο από το διάστημα που ορίζει ο SPG.

αριθμός server=3,ακτίνα ανίχνευσης r_s=3,παράγοντας ασφάλειας α=0,τροχιά=square_1



Σχήμα 42: 100 κόμβοι με 3 servers και τροχιά square_1



Σχήμα 43: Δυναμική Ανάλυση τοπολογίας 100 κόμβων με 3 servers και τροχιά square_1

Παρατηρείται ότι ο ΑΔΧ (κόκκινη γραμμή) επιτυγχάνει μια αύξηση του privacy κατά 65% για τη χρονική διάρκεια από 10-90sec. Στο δεύτερο μισό του πειράματος το privacy διατηρείται στα ίδια υψηλά επίπεδα εκτός από ορισμένα διαστήματα λίγων δευτερολέπτων. Έτσι, η μέση τιμή του privacy εκτοξεύεται από το 0,60 σε 0,88. Παρατηρούμε ότι ενώ ο walkthrough επιτυγχάνει μηδενικά conflicts το privacy ιδιαίτερα έως και τη χρονική στιγμή το privacy του είναι σε ιδιαίτερα χαμηλά επίπεδα. Αυτό οφείλεται στους κόμβου που συνανιχνεύουν και απέχουν λιγότερο από $2r_s$-a τους οποίους ο ΑΔΧ διαχειρίστηκε καλύτερα με το επιπλέον χρώμα 4 (n+1).

Privacy	v_line_3	square_1
2 storage nodes	-	0.6088
3 storage nodes	0.8477	0.6084
4 storage nodes	0.7917	-

Πίνακας 3: Walkthrough

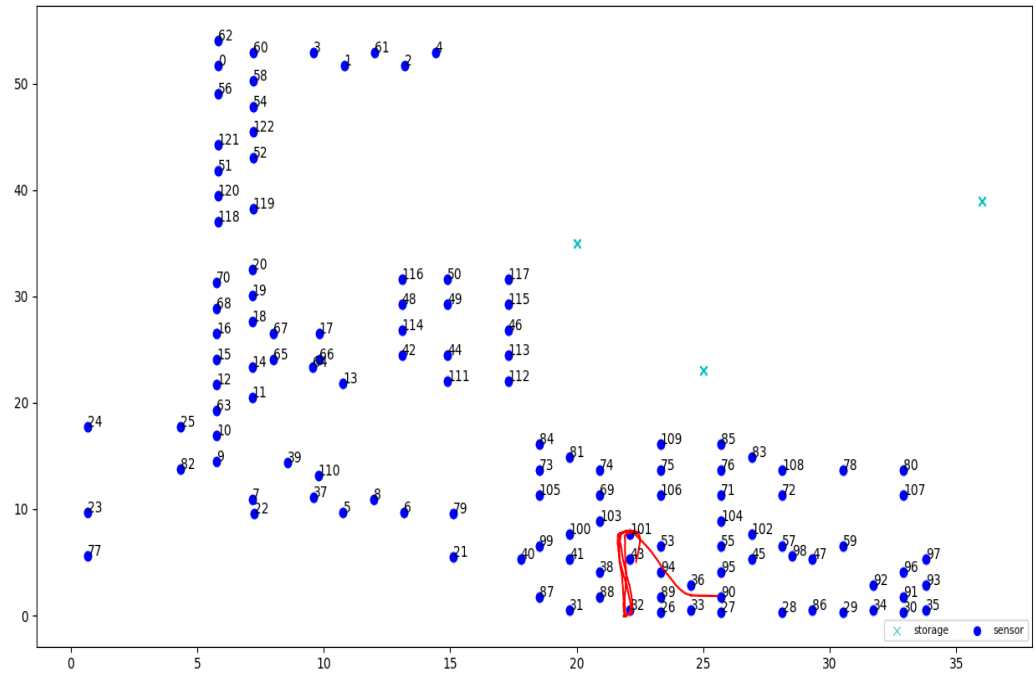
Privacy	v_line_3	square_1
2 storage nodes	-	0.7707
3 storage nodes	0.9135	0.8817
4 storage nodes	0.9248	-

Πίνακας 4: Αλγόριθμος Δυναμικού Χρωματισμού

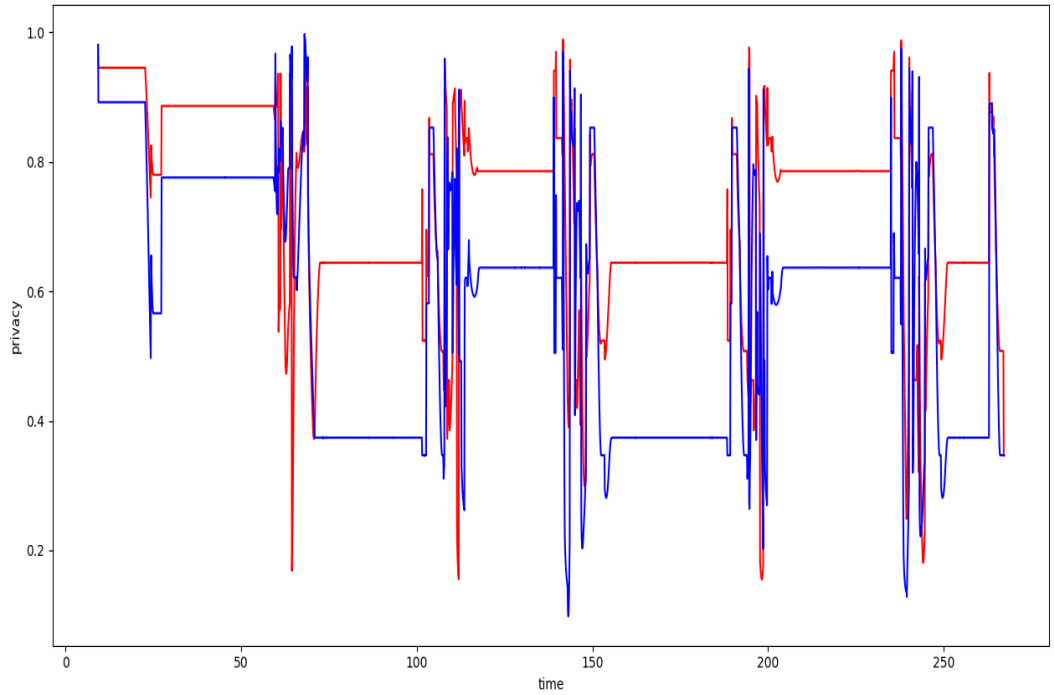
Στους παραπάνω πίνακες φαίνονται τα αποτελέσματα που επιτυγχάνουν οι δύο αλγόριθμοι στα διαφορετικά πειράματα που διεξήχθησαν στην τοπολογία των 100 κόμβων ανίχνευσης. Στις γραμμές δηλώνεται το πλήθος των servers και στις στήλες η τροχιά του κινητού αντικειμένου. Οι παύλες που φαίνονται στους πίνακες δηλώνουν ότι δεν διεξήχθη πείραμα υπό αυτές τις συνθήκες. Η επιλογή του εύρους των servers έγινε εμπειρικά ανάλογα με την πυκνότητα της τοπολογίας. Βλέπουμε την θεαματική βελτίωση του privacy που κυμαίνεται από 10-30%. Παρατηρούνται επίσης τα αποτελέσματα που επιτεύχθηκαν και στην προηγούμενη τοπολογία όπως το υψηλό επίπεδο privacy με λίγους servers και την αναλογική αύξηση του όσο αυξάνεται το πλήθος των servers. **Πολύ σημαντικό γεγονός όμως αποτελεί ότι στο πείραμα που είχαμε ως τροχιά square 1, ενώ ο walkthrough επιτυγχάνει μηδενικά conflicts από τους 2 servers και μετά, εντούτοις επιτυγχάνει privacy μόλις 60%. Αντιθέτως ο ΑΔΧ σε αυτές τις περιπτώσεις εκτοξεύει το privacy κοντά στο 80% και 90% αντίστοιχα.**

123 κόμβοι

αριθμός server=3,ακτίνα ανίχνευσης r_s=3,παράγοντας ασφάλειας α=2,τροχιά=v_line_5



Σχήμα 44: 123 κόμβοι με 3 servers και τροχιά v_line_5

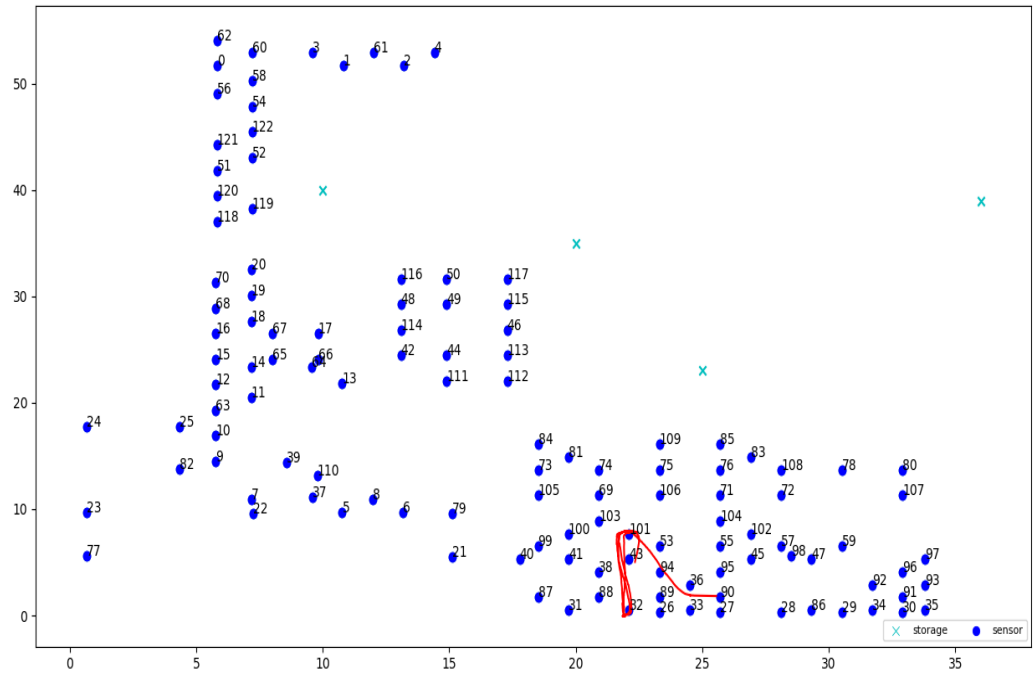


Σχήμα 45: Δυναμική Ανάλυση τοπολογίας 123 κόμβων με 3 servers και τροχιά v_line_5

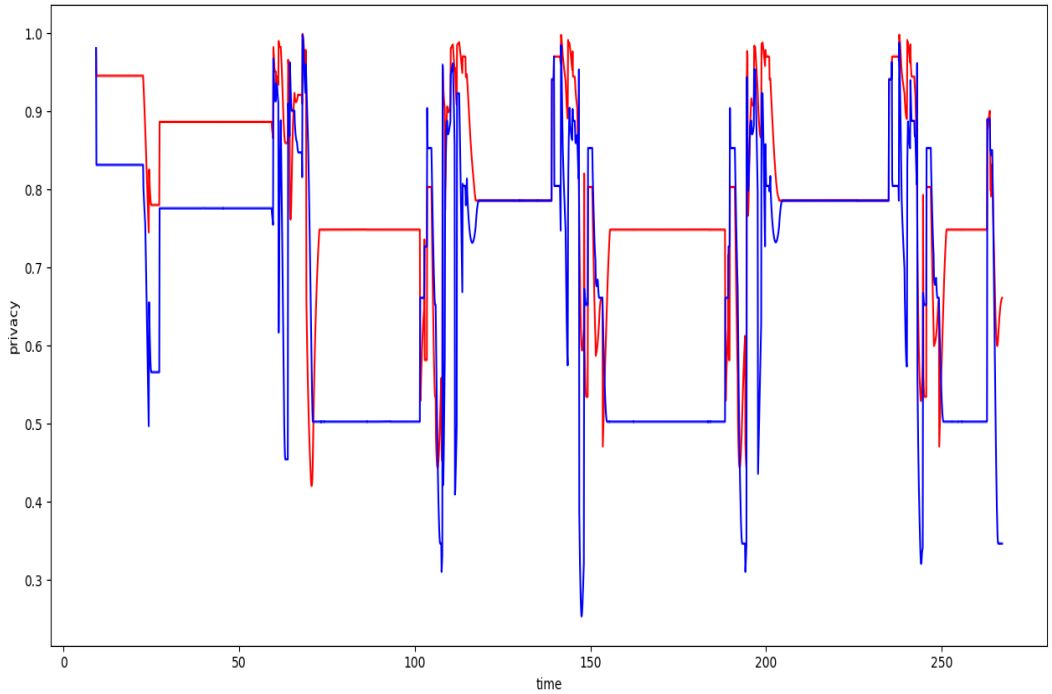
Παρατηρείται ότι ο ΑΔΧ (κόκκινη γραμμή) επιτυγχάνει μεγαλύτερο privacy σε σχέση με τον walkthrough (μπλε γραμμή) κατά 20-25% σχεδόν σε όλο το εύρος της διαδρομής εκτός από μικρά χρονικά διαστήματα μικρότερα των 10sec. Ακόμα και σε αυτά τα διαστήματα ο ΑΔΧ καταφέρνει να επιτύχει ίδιο ή μεγαλύτερο privacy με εξαίρεση πολύ μικρές περιόδους των 2-3sec. Όλα αυτά τα γεγονότα συντελούν σε μια αύξηση του μέσου privacy από 0,57 σε 0,72. Η τοπολογία είναι αρκετά αραιή αλλά παρόλα αυτά επιτυγχάνει καλύτερο αποτέλεσμα από τον walkthrough. **Το πλεονέκτημα walkthrough της επίτευξης του μικρότερου πλήθους conflicts σε αραιές γειτονιές**

εξαλείφεται. Ο ΑΔΧ εκμεταλλεύεται το μικρότερο πλήθος ακμών του γράφου για να επισυνάψει με ευκολότερο τρόπο το επιπλέον χρώμα στους κόμβους που δημιουργούν πρόβλημα.

αριθμός server=4,ακτίνα ανίχνευσης r_s=3,παράγοντας ασφάλειας α=2,τροχιά=v_line_5



Σχήμα 46: 123 κόμβοι με 4 servers και τροχιά v_line_5

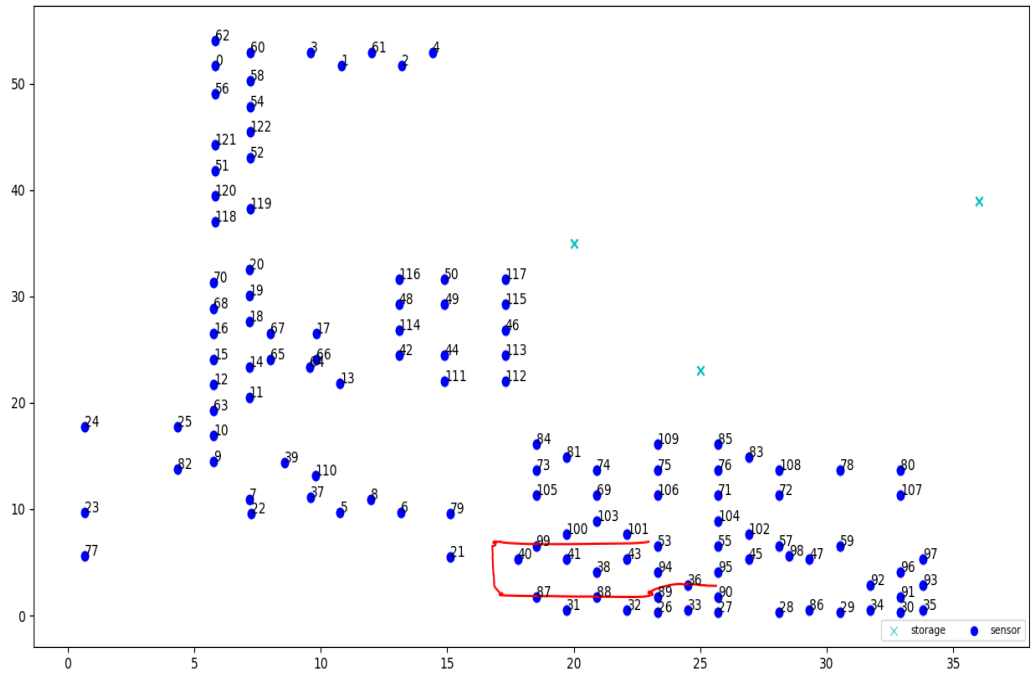


Σχήμα 47: Δυναμική Ανάλυση τοπολογίας 123 κόμβων με 4 servers και τροχιά v_line_5

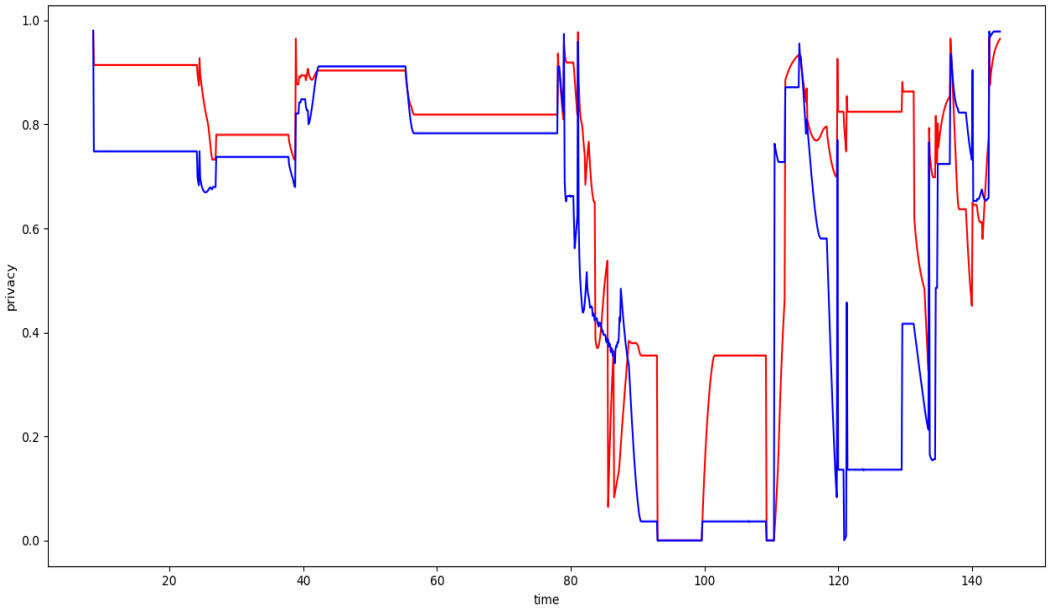
Παρατηρούνται μεγάλα διαστήματα στα οποία έχουμε αύξηση του privacy από 15-25% όπως για παράδειγμα από 25-60sec, 70-100sec, 160-190sec, 250-270sec. Το μέσο privacy αυξάνεται από 0,68 σε 0,79. Παρόλο που αυξήθηκε ο αριθμός των servers ο walkhrough δεν μπόρεσε παρά ελάχιστα διαστήματα δευτερολέπτων να επιτύχει καλύτερο αποτέλεσμα από τον

ΑΔΧ καθώς εκείνος αύξησε αναλογικά την επίδοση του. **Παρατηρούμε ότι σε ιδιαίτερα αραιές τοπολογίες το πλεονέκτημα που είχε ο walkthrough, δηλαδή την επίτευξη μικρότερου αριθμού conflicts, εξαλείφεται και μάλιστα ο ΑΔΧ γίνεται αισθητά καλύτερος. Αυτό συμβαίνει διότι οι πιο αραιές γειτονιές διευκολύνουν τον ΑΔΧ να επισυνάψει το επιπλέον χρώμα στους κόμβους που δημιουργούν πρόβλημα**.

αριθμός server=3,ακτίνα ανίχνευσης r_s=3,παράγοντας ασφάλειας α=2, τροχιά=square_1



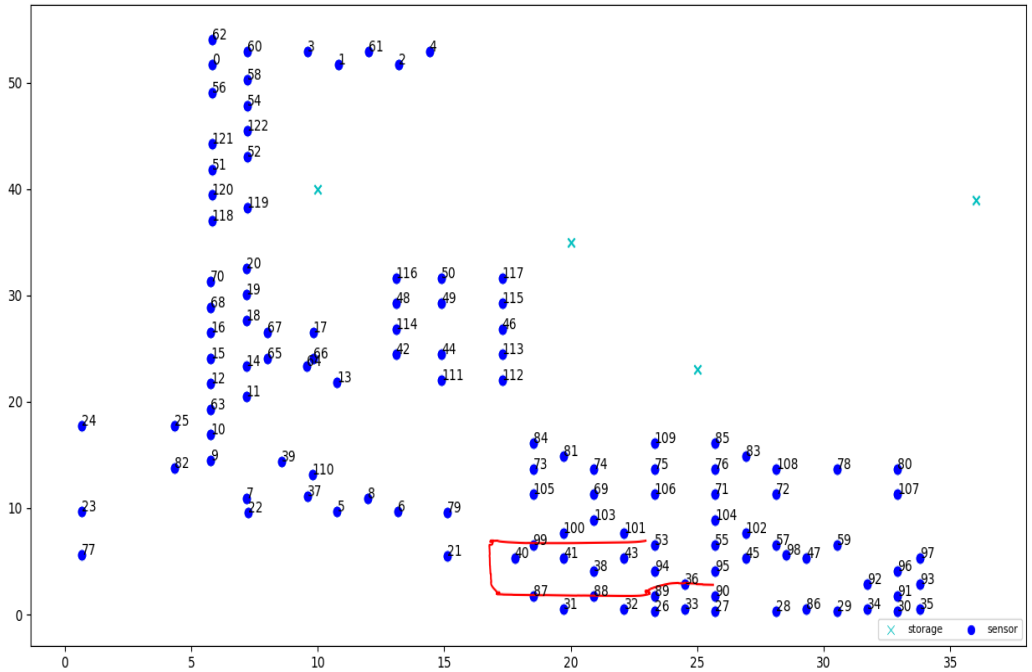
Σχήμα 48: 123 κόμβοι με 3 servers και τροχιά square_1



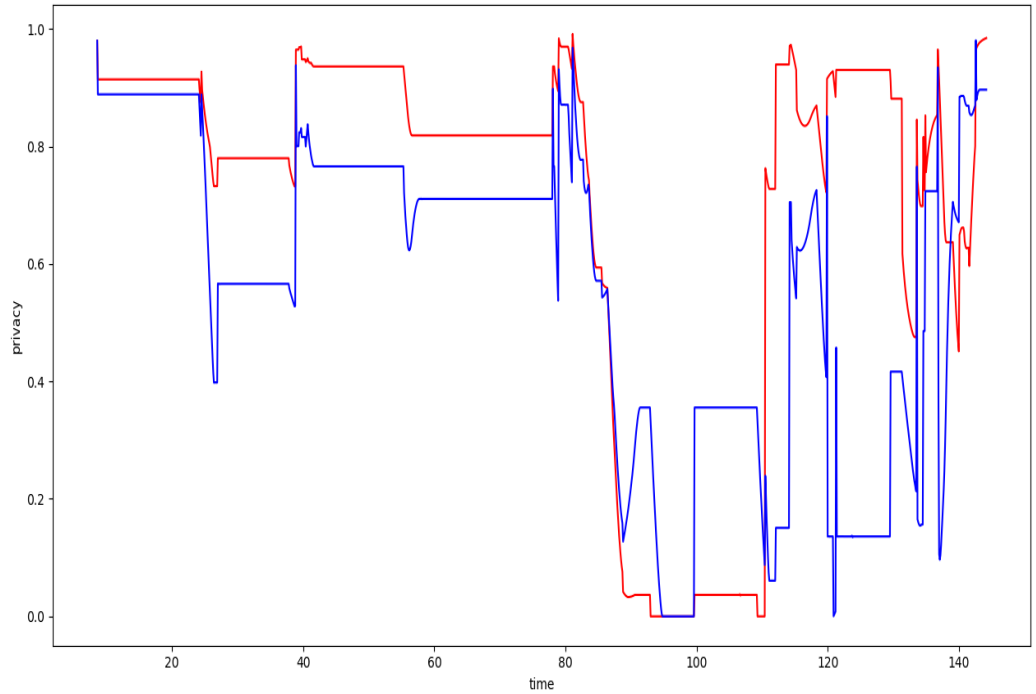
Σχήμα 49: Δυναμική Ανάλυση τοπολογίας 123 κόμβων με 3 servers και τροχιά square_1

Παρατηρείται ότι ο ΑΔΧ (κόκκινη γραμμή) παρουσιάζει καλύτερο privacy καθ' όλη την διάρκεια του πειράματος. Στο διάστημα μεταξύ 90-110 sec σημειώνεται μια αύξηση του privacy περίπου 30% οδηγώντας έτσι στη βελτίωση του μέσου όρου από 0,56 σε 0,70.

αριθμός server=4,ακτίνα ανίχνευσης r_s=3,παράγοντας ασφάλειας α=2,τροχιά=square_1



Σχήμα 50: 123 κόμβοι με 4 servers και τροχιά square_1



Σχήμα 51: Δυναμική Ανάλυση τοπολογίας 123 κόμβων με 3 servers και τροχιά square_1

Παρατηρείται ότι ο ΑΔΧ (κόκκινη γραμμή) στο μεγαλύτερο διάστημα του πειράματος βελτιώνει το privacy του δικτύου. Είναι ενδεικτικό ότι στο διάστημα από 110-115 sec και 120-130 sec υπάρχει βελτίωση έως και 80% ενώ ο συνολικός μέσος έχει αυξηθεί από 0,59 σε 0,70. Εντούτοις, στο διάστημα από 90-110 sec παρατηρείται μια πτώση έως και 30% στο privacy που επιτυγχάνει ο ΑΔΧ. Το γεγονός αυτό οφείλεται στον χρωματισμό των κόμβων που απέχουν απόσταση λιγότερη από αυτή που ορίζεται στον SPG δηλαδή λιγότερη από $2r_s$-a. Σε αυτήν την εκτέλεση ο αλγόριθμος walkthrough επέτυχε καλύτερο χρωματισμό των κόμβων αυτό οδηγώντας σε αύξηση του privacy για το χρονικό αυτό διάστημα.

Privacy	square_1	v_line_5
3 storage nodes	0.5608	0.5708
4 storage nodes	0.5978	0.6877

Πίνακας 5: Walkthrough

Privacy	square_1	v_line_5
3 storage nodes	0.7039	0.7289
4 storage nodes	0.7046	0.7989

Πίνακας 6: Αλγόριθμος Δυναμικού Χρωματισμού

Στους παραπάνω πίνακες φαίνονται τα αποτελέσματα που επιτυγχάνουν οι δύο αλγόριθμοι στην τοπολογία με 123 κόμβους ανίχνευσης. Στις γραμμές δηλώνεται το πλήθος των κόμβων ανίχνευσης ενώ στις στήλες η τροχιά του κινητού κόμβου. Γίνεται εμφανής η βελτίωση του privacy που επιτυγχάνει ο ΑΔΧ οδηγώντας σε μία αύξηση της τάξης του 10-15%. Παρατηρούνται και πάλι τα ίδια αποτελέσματα με εκείνα που που επιτεύχθηκαν στις προηγούμενες τοπολογίες **όπως το αυξημένο επίπεδο ασφάλειας σε μικρό πλήθος servers και η αναλογική βελτίωση του privacy με την αύξηση τους**.

Συμπεράσματα

Ο αλγόριθμος Δυναμικού Χρωματισμού αποτελεί ουσιαστικά έναν αλγόριθμο constrained clustering 2 επιπέδων. Στο πρώτο επίπεδο έχουμε έναν αρχικό διαχωρισμό των δεδομένων σε k ομάδες, όπου το k αποτελεί ένα δοθέν αριθμό, ενώ στο δεύτερο επίπεδο έχουμε μία εξομάλυνση των όποιων constraints έχουν παραβιασθεί κατά τη δημιουργία των αρχικών ομάδων. Έτσι τα δεδομένα αλλάζουν δυναμικά την ομάδα στην οποία ανήκουν, ανάλογα με τις συνθήκες που επικρατούν την εκάστοτε χρονική στιγμή.

Με την εφαρμογή του Δυναμικού Χρωματισμού καταφέραμε να **αυξήσουμε το επίπεδο ασφάλειας του δικτύου σε σχέση με τον προτεινόμενο αλγόριθμο walkthrough**. Βλέπουμε λοιπόν, σύμφωνα με τα ευρήματα που παραθέσαμε πιο πάνω, ότι ανεξαρτήτου διαδρομής κινητού στόχου, πυκνότητας γράφου, ακτίνα ανίχνευσης και τοπολογίας, το δίκτυο παρουσιάζει μεγαλύτερο privacy. Πολύ σημαντική παρατήρηση αποτελεί το γεγονός ότι ο τρόπος με τον οποίων γίνεται η επιλογή των κόμβων, που θα αλλάζουν δυναμικά το χρώμα τους, έχει πολύ μεγάλη επίδραση στην αποτελεσματικότητα του αλγορίθμου. Βάσει εμπειρίας και πειραματικών αποτελεσμάτων, η μέθοδος selective αποδείχθηκε η πιο κατάλληλη μιας και οι κόμβοι που επιλέγονται είναι πιο διάσπαρτοι μέσα στο δίκτυο και όχι εντοπισμένοι σε ένα συγκεκριμένο μέρος του. Έτσι είναι πιο εύκολα διαχειρίσιμοι και δεν δημιουργούνται γειτονιές με μεγάλο πλήθος κόμβων, που αλλάζουν δυναμικά το χρώμα τους, οι οποίες οδηγούν σε μεγαλύτερα προβλήματα συγχρονισμού, κατανάλωσης ενέργειας και υπολογιστικής ισχύς καθώς ανταλλάζονται διαρκώς μηνύματα μεταξύ τους.

Η επιλογή του αλγορίθμου που χρησιμοποιείται στο πρώτο στάδιο του Δυναμικού Χρωματισμού είναι καθοριστική για το αποτέλεσμα του προτεινόμενου αλγορίθμου. **Η τυχαιότητα**

του αλγορίθμου walkthrough μπορεί να οδηγήσει σε μεγάλες μεταβολές στην μετρική privacy, και κατά συνέπεια στο επίπεδο της ασφάλειας. Ενώ ο Δυναμικός Χρωματισμός μπορεί να διαχειριστεί αυτήν την τυχαιότητα και να οδηγήσει σε ένα βελτιωμένο αποτέλεσμα, εντούτοις προτιμήθηκε να χρησιμοποιείται ο constrained agglomerative clustering, ο οποίος ενώ μπορεί να δημιουργεί σε διάφορες περιπτώσεις περισσότερα conflicts από τον walkthrough, εντούτοις αυτά διατηρούνται σε ένα αποδεκτό επίπεδο. Επιπλέον, τα conflicts που δημιουργούνται με αυτόν τον αλγόριθμο είναι σταθερά σε κάθε διαδοχική του εκτέλεση, κάνοντας τον Δυναμικό Χρωματισμό να παράγει ένα σταθερό επίπεδο privacy.

Όπως γίνεται φανερό από τις γραφικές που παρατέθηκαν πιο πάνω, υπάρχουν ορισμένα μικρά διαστήματα κατά τη διάρκεια του πειράματος που ο αλγόριθμος walkthrough επιτυγχάνει καλύτερο privacy σε σχέση με τον αλγόριθμο Δυναμικού Χρωματισμού. Αυτό το φαινόμενο σχετίζεται με το γεγονός ότι πρόβλημα στο επίπεδο ασφάλειας ενός δικτύου ενδέχεται να προκαλέσουν και οι κόμβοι που απέχουν απόσταση μεταξύ τους μικρότερη του 2rs-α. Σύμφωνα με την υλοποίηση των δύο αλγορίθμων, αυτοί οι κόμβοι δεν διαχειρίζονται με κάποιον τρόπο καθώς στον walkthrough ο SPG κατασκευάζεται βάση αυτών που απέχουν απόσταση μεταξύ [2rs-a,2rs], ενώ στον constrained agglomerative clustering εκχωρείται το max (A)+100 στις θέσεις του πίνακα όπου οι κόμβοι απέχουν και πάλι εντός αυτού του διαστήματος. Έτσι, οι αλγόριθμοι μπορούν να χρωματίσουν αυτούς τους κόμβους με διαφορετικό τρόπο, που σε ορισμένες περιπτώσεις εκείνος του walkthrough να είναι καλύτερος. Ιδιαίτερα αν ο Δυναμικός Χρωματισμός δεν αναθέσει σε κάποιον από αυτούς το μη υπαρκτό χρώμα n+1, ο χρωματισμός αυτός θα μείνει έτσι μέχρι το τέλος του πειράματος. Ωστόσο, στον προτεινόμενο αλγόριθμο της εργασίας, μπορεί να γίνει αυτή η βελτίωση στο στάδιο του preprocessing, ώστε να γίνεται διαχείριση σε γειτονιές κόμβων που είναι πολύ κοντά μεταξύ τους και έχουν χρωματιστεί με τρόπο που δημιουργεί πρόβλημα ύστερα και από την ολοκλήρωση του πρώτου επιπέδου συσταδοποίησης. **Ωστόσο ακόμα και δίχως αυτήν την βελτίωση, ο αλγόριθμος Δυναμικού Χρωματισμού επιτυγχάνει καλύτερο privacy από τον walkthrough στο συντριπτικά μεγαλύτερο κομμάτι της διάρκειας του πειράματος.**

ΚΕΦΑΛΑΙΟ 7 -Συμπεράσματα και μελλοντική μελέτη

7.1 Εποπτική θεώρηση της προτεινόμενης μεθόδου

Τα βήματα 2,3,4 που παρουσιάστηκαν στο κεφάλαιο 5.1 μπορούν να εκτελεστούν είτε κατανεμημένα από κάθε κόμβο ανίχνευσης ξεχωριστά είτε από έναν κεντρικό σύστημα διαχείρισης που θα έχει την εποπτεία του δικτύου, θα είναι ενήμερο για τις γειτονιές που υπάρχουν καθώς και τα χρώματα των κόμβων και θα τους διαχειρίζεται ανάλογα με την κατάσταση στην οποία έχει επέλθει το δίκτυο. Ωστόσο, είναι σημαντικό οι αλγόριθμοι αυτοί να μπορούν να τρέξουν κατανεμημένα διότι έτσι αυξάνεται ραγδαία η επίδοση του συστήματος και μειώνεται αισθητά ο χρόνος απόκρισης. Το γεγονός αυτό είναι πολύ σημαντικό καθώς το σύστημα αυτό αναμένεται να χρησιμοποιηθεί σε real-time εφαρμογές. Όμως οι κατανεμημένοι αλγόριθμοι έχουν ορισμένες δυσκολίες και ένα πολύ κοινό πρόβλημα αποτελεί το deadlock. Σε γειτονιές του δικτύου που θα υπάρχουν περισσότεροι του ενός κόμβου ανίχνευσης με χρώμα n+1, υπάρχει η περίπτωση να σταματήσει η εκτέλεση καθώς θα περιμένει ο ένας κόμβος τον άλλον να αποκτήσει υπαρκτό χρώμα. Έτσι θα πρέπει να υπάρχει ένας μηχανισμός η αλλιώς ένα πρωτόκολλο σύμφωνα με το οποίο να αποφεύγεται αυτή η δυσάρεστη κατάσταση. Μια απλή λύση θα ήταν μετά την φάση του preprocessing στους κόμβους αυτούς να υπάρχει ένας δείκτης προτεραιότητας ώστε να γνωρίζουν οι κόμβοι πότε να περιμένουν την εκτέλεση ενός άλλου και πότε όχι.

Ένα άλλο σημαντικό πρόβλημα που δημιουργείται είναι η αύξηση στην κατανάλωση της ενέργειας των κόμβων από τη στιγμή που θα πρέπει να στέλνουν μήνυμα στους κόμβους με χρώμα n+1 κάθε χρονική στιγμή t_i. Είναι σημαντικό να διακρίνουμε ότι δεν χρειάζεται να στέλνονται μηνύματα κάθε χρονική στιγμή αλλά όταν γίνεται μια ανανέωση των κόμβων που συνανιχνεύουν το κινητό. Έτσι θα μπορούσε ο κεντρικός μηχανισμός που αναφέραμε πριν, ο οποίος θα μπορούσε να ήταν ένας επιπλέον κινητός κόμβος με μεγαλύτερη υπολογιστική ισχύ και αποθηκευτικό χώρο να ενημερώνει τους κόμβους μιας γειτονιάς πότε να στείλουν μήνυμα η όχι. Μια δεύτερη λύση ώστε να μην αυξηθεί το κόστος του δικτύου εξαιτίας του επιπλέον κόμβου θα ήταν να υπάρχουν δύο ειδών μηνύματα, το ένα θα είναι την χρονική στιγμή t_i που ο κόμβος χρώματος <n+1 θα ξεκινάει να ανιχνεύει το κινητό και το δεύτερο τη χρονική στιγμή t_j που σταματά. Έτσι καθ όλη τη διάρκεια t_j- t_i κόμβος που δεν έχει υπαρκτό χρώμα θα τον λαμβάνει υπόψιν του και μετά θα τον αγνοεί. Επιπλέον και για τους κόμβους που δεν έχουν κάποιο υπαρκτό χρώμα θα ήταν αναγκαίο να υπάρχουν δύο ειδών μηνύματα, ένα που θα δείχνει την έναρξη της δικής του περιόδου ανίχνευσης και ένα την παύση αυτής προκειμένου οι γείτονες του να μην του στέλνουν περιττά μηνύματα όταν εκείνος δεν ανιχνεύει το κινητό αντικείμενο.

Ένα τρίτο σημαντικό ζήτημα είναι ότι θα πρέπει το σύστημα να είναι γρήγορα αποκρίσιμο ειδικότερα όταν υπάρχει ανάγκη για real-time monitoring. Οι γραμμές 7-14 ίσως δημιουργήσουν πρόβλημα επίδοσης ειδικά αν το δίκτυο είναι πολύ πυκνό και οι γειτονιές πολύ μεγάλες. Μια απλή λύση σε αυτήν την περίπτωση θα ήταν οι κόμβοι να διαθέτουν μια μικρή cache η ένα buffer στους κόμβους ανίχνευσης ώστε να βλέπουν τα ενεργά χρώματα και να διαλέγουν αμέσως χωρίς καθυστέρηση.

7.2 Μελλοντικές επεκτάσεις

Ο αλγόριθμος Δυναμικού Χρωματισμού μπορεί να εξελιχθεί και να γίνει ακόμα πιο αποτελεσματικός στην μετάδοση των δεδομένων σε περιβάλλοντα IoT στα οποία διακινούνται ευαίσθητες πληροφορίες. Αρχικά το στάδιο του preprocessing μπορεί να εξελιχθεί για να εξομαλύνει τα όποια προβλήματα δημιουργούνται μεταξύ των δύο επιπέδων συσταδοποίησης και να έχει καλύτερη διαχείριση στις γειτονιές των κόμβων όπως προαναφέρθηκε και στο κεφάλαιο 6.

Ένας πολύ σημαντικός παράγοντας που δεν εξετάστηκε ενδελεχώς, στην συγκεκριμένη εργασία, αποτελεί η κατανάλωση ενέργειας των κόμβων καθώς αποτελούν συσκευές με πολύ συγκεκριμένους περιορισμούς ενέργειας. Είναι πολύ σημαντικό να μειώσουμε το πλήθος των μηνυμάτων που ανταλλάζονται στο στάδιο της δυναμικής ανάθεσης των χρωμάτων και κατά επέκταση το πλήθος των κόμβων που επισυνάπτονται στο μη υπαρκτό χρώμα. Επιπλέον, είναι πολύ σημαντικό να γίνεται καλύτερη αντιστοιχία των κόμβων ανίχνευσης στους κόμβους αποθήκευσης λαμβάνοντας υπ’ όψιν και τον παράγοντα της ενέργειας.

Ως μελλοντική επέκταση, θα ήταν πολύ σημαντικό να προσπαθήσουμε να εφαρμόσουμε αυτόν τον αλγόριθμο σε μεγαλύτερης κλίμακας τοπολογίες και να εξετάσουμε την επίδοση και την αποτελεσματικότητά του. Η γενίκευση αυτού του αλγορίθμου σε άλλου τύπου επιβάλλοντα IoT ή ακόμα και σε συμβατικά δίκτυα υπολογιστών αρχιτεκτονικής client-server στα οποία ανταλλάζονται ευαίσθητες πληροφορίες, αποτελεί τον απώτερο σκοπό μας.

ΚΕΦΑΛΑΙΟ 8 -Επίλογος

Στο πλαίσιο της παρούσας διπλωματικής εξετάσθηκε και υλοποιήθηκε ένα νέο πρωτόκολλο μετάδοσης δεδομένων με σκοπό την προστασία ευαίσθητων δεδομένων από επίδοξους εισβολείς. Έτσι, σχεδιάστηκε ένας νέος αλγόριθμος constrained clustering δύο επιπέδων με σκοπό την βελτίωση του επιπέδου ιδιωτικότητας του IoT περιβάλλοντος. Ο αλγόριθμος αυτός στην ουσία χωρίζεται σε 3 στάδια, το στάδιο εφαρμογής ήδη υπαρχόντων αλγορίθμων για έναν πρώτο διαχωρισμό των κόμβων και επιλογή των κόμβων που θα επισυναφθούν το ανύπαρκτο χρώμα, το ενδιάμεσο στάδιο του preprocessing το οποίο προσπαθεί να επεξεργαστεί και να βελτιώσει αυτόν τον αρχικό διαχωρισμό και το τρίτο στάδιο της Δυναμικής ανάθεσης που διαχειρίζεται τους κόμβους που αλλάζουν δυναμικά το χρώμα τους.

Ξεκινήσαμε με την περιγραφή της φύσης του δικτύου, των κόμβων που το απαρτίζουν αλλά και ο τρόπος με τον οποίο επεξεργάζονται και μεταδίδονται τα δεδομένα. Επιπλέον, παρουσιάσαμε μια σειρά δημοφιλών αλγορίθμων constrained clustering που χρησιμοποιήθηκαν για την στατική ανάλυση του δικτύου όπως ο pck-means, constrained agglomerative clustering, walkthrough. Στη συνέχεια αναλύθηκε ο σχεδιασμός του νέου προτεινόμενου αλγορίθμου.

Τέλος, παρουσιάστηκαν τα πειραματικά αποτελέσματα της στατικής και δυναμικής ανάλυσης πάνω στις αντίστοιχες τοπολογίες και αναλύθηκαν λεπτομερώς τα σημαντικά πλεονεκτήματα που φέρει ο νέος αυτός αλγόριθμος καθώς και την θεαματική αύξηση της μετρικής privacy των δικτύων. Επιπλέον, επεξηγήθηκαν διάφορες βελτιώσεις του αλγορίθμου Δυναμικού Χρωματισμού προκειμένου να επιτύχουμε το μέγιστο δυνατό privacy. Τα θετικά αποτελέσματα του αλγορίθμου αυτού μπορούν να ανοίξουν νέους δρόμους στη ανάπτυξη μιας νέας γενιάς αλγορίθμων constrained clustering.

ΚΕΦΑΛΑΙΟ 9 -Βιβλιογραφία

[1] Manuel Eduardo Ares Brea, Dr. Alvaro Barreiro Garcıa. Constrained Clustering Algorithms: Practical Issues and Applications-phd thesis, universidade da coruña, 2013.

[2] Eric Bair. Semi-supervised clustering methods, WIREs Comp Stat, Vol. 5, No. 5 pp. 349-361 2013.

[3] Miao Xu, Wenyuan Xu, JasonM. O'Kane. Privacy Preservation Data Dissemination, Security and Privacy in Internet of Things (IoTs), CRC Press, 2016.

[4] Hoffman's Bound for Vector Colouring, Available at
http://www.sfu.ca/~mdevos/notes/misc/vec-chrom-hoffman.pdf

[5] Eline Rietberg. Bachelor Thesis, The Shannon capacity of graphs, TuDelft, Netherlands, August 7, 2013.

[6] Lee In, Lee Kyoochun. The Internet of Things (IoT) : applications, investments, and challenges for enterprises, Business horizons. - Amsterdam : Elsevier, ISSN 0007-6813, ZDB-ID 222663-7. - Vol. 58, No. 4, p. 431-440, 2015.

[7] Richard Boddington. An Analysis of Triangulation Techniques for Radio-Telemetry, 2017. DOI: 10.13140/RG.2.2.17919.71849

[8] Vincent Delos, Denis Teissandier. Minkowski Sum of Polytopes Defined by Their Vertices. Journal of Applied Mathematics and Physics (JAMP), Scientific Research Publishing, Vol. 3, No. 1, pp.62-67, 2015.

[9] What is Chromatic number Available at
http://mathworld.wolfram.com/ChromaticNumber.html

[10] R. Agrawal and R. Srikant. Privacy-preserving data mining. In *Proc. of the ACMSIGMOD* Conference on Management of Data, pp. 439–450. ACM Press, 2000.

[11] U. Cetintemel, A. Flinders, and Y. Sun. Power-efficient data dissemination in wireless sensor networks. In Proceedings of Workshop on Data Engineering for Wireless and Mobile Access (MobiDe), pp. 1–8, 2003.

[12] Deng, J., Han, R., & Mishra, S. Intrusion tolerance and anti-traffic analysis strategies for wireless sensor networks. In International Conference on Dependable Systems and Networks, pp. 637-646, IEEE, 2014.

[13] R. Kotla, L. Alvisi, and M. Dahlin. Safestore: A durable and practical storage system. In USENIX Annual Technical Conference, pp. 07–20, 2007.

[14] C. K. Liew, U. J. Choi, and C. J. Liew. A data distortion by probability
distribution. ACM Transactions on Database Systems, Vol. 10, No.3, pp. 395–411, 1985.

[15] Y. Yang,M. Shao, S. Zhu, B. Urgaonkar, and G. Cao. Towards event source unobservability with minimum network traffic in sensor networks. In Proceedings of Conference on Wireless Network Security (WiSec), pp. 77-88, 2008.

[16] Mehrotra, Anuj, and Michael A. Trick. A column generation approach for graph coloring.informs Journal on Computing Vol. 8, No. 4, pp. 344-354, 1996.

[17] Sándor Szabó, What is Clique in graph theory, Searching Cliques in Graphs, LAP LAMBERT Academic Publishing, 2015.

[18] What is Brooks theorem, Notes on Brooks' Theorem, Available at
https://www.math.brown.edu/~res/M123/brooks.pdf

[19] Robin J. Wilson, Introduction to Graph Theory (5th Edition), Amazon, 2012.

[20] Maarten van Steen, Graph Theory and Complex Networks: An Introduction, Amazon, 2010.

[21] G. Ganger, P. Khosla, M. Bakkaloglu, M. Bigrigg, G. Goodson, S. Oguz, V. Pandurangan, C. Soules, J. Strunk, and J. Wylie. Survivable storage systems. DARPA Information Survivability Conference and Exposition, Vol. 2, pp. 184–195, 2001.

[22] Kamat, P., Zhang, Y., Trappe, W., & Ozturk, C. Enhancing source-location privacy in sensor network routing. In 25th IEEE international conference on distributed computing systems (ICDCS'05), pp. 599-608, IEEE, 2005

[23] K. Mehta, D. Liu, and M. Wright. Location privacy in sensor networks against a global eavesdropper. In Proceedings of Conference on Network Protocols (ICNP), pp. 314–323, 2007.

[24] N.Minsky. Intentional resolution of privacy protection in database systems. *Commun. ACM*, Vol. 19, No.3, pp. 148–159, 1976.

[25] What is Fractional_coloring Available at
http://mathworld.wolfram.com/FractionalColoring.html

[27] M. Shao, S. Zhu, W. Zhang, G. Cao, and Y. Yang. pDCS: Security and privacy support for data-centric sensor networks. *IEEE Transactions on Mobile Computing*, Vol. 8, No.8, pp. 1023–1038, 2009.

[28] FIT-iot LAB Available at
https://www.iot-lab.info/

[29] Github Available at
https://github.com/iot-lab/iot-lab/wiki/

[30] What is ROS system Available at

[31] Charu C. Aggarwal, Chandan K. Reddy. Data Clustering: Algorithms and Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series)1st Edition, Amazon, 2014

[32] Steinley, Douglas. K-means clustering: a half-century synthesis. British Journal of Mathematical and Statistical Psychology, Vol.59, No.1 pp.1-34, 2006

[33] Gan, Guojun, Chaoqun Ma, and Jianhong Wu. Data clustering: theory, algorithms, and applications. Vol. 20. Siam, 2007.

[34] Rui Xu,Don Wunsch. Clustering, Amazon, 2008.

[35] Murtagh, Fionn, and Pedro Contreras. Algorithms for hierarchical clustering: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery Vol.2, No.1 pp. 86-97,2012.

[36] Nasraoui, Olfa, and Chiheb-Eddine Ben N'Cir, eds.Clustering Methods for Big Data Analytics: Techniques, Toolboxes and Applications. Springer, 2018.

[37] Basu, Sugato, Ian Davidson, and Kiri Wagstaff, eds. Constrained clustering: Advances in algorithms, theory, and applications. CRC Press, 2008.

[38] Klein, Dan, Sepandar D. Kamvar, and Christopher D. Manning. From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. Stanford, 2002.

[39] Kamvar, K., Sepandar, S., Klein, K., Dan, D., Manning, M., & Christopher, C. Spectral learning. In International Joint Conference of Artificial Intelligence. Stanford InfoLab. 2003.

[40] Daniel, Larry. Cell Phone Location Evidence for Legal Professionals: Understanding Cell Phone Location Evidence from the Warrant to the Courtroom. Academic Press, 2017.

[41] Search Engine Land Available at
https://searchengineland.com/cell-phone-triangulation-accuracy-is-all-over-the-map-14790

[42] Steven G. Krantz, Convex Analysis (Textbooks in Mathematics), Amazon, 2014

[43] Andreas Brandstädt, Klaus Jansen. Graph-Theoretic Concepts in Computer Science, Springer, 39th International Workshop, Lübeck, Germany, 2013.

[44] Sivaraman, Vaidy. A unified proof of Brooks' theorem and Catlin's theorem. Discrete Mathematics, Vol. 338, No.2, pp.272-273, 2015.

[45] What is ZMTP protocol Available at
https://rfc.zeromq.org/spec:23/ZMTP/

[46] Michael Soltys-Kulinicz, Introduction To The Analysis Of Algorithms, An (3rd Edition), Amazon, 2018.

[47] Karumanchi, Narasimha. Algorithm Design Techniques: Recursion, Backtracking, Greedy, Divide and Conquer, and Dynamic Programming. 2018.