



Πολυτεχνείο Κρήτης

Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Η/Υ

Θέμα Διπλωματικής Εργασίας

*Κατανεμημένη Παρακολούθηση Παραβίασης Κατωφλίου σε Δίκτυα
Αισθητήρων*

Ζωγράφου Ιακωβίνα

Εξεταστική Επιτροπή:

Αντώνιος Δεληγιαννάκης (επιβλέπων)

Μίνως Γαροφαλάκης

Βασίλειος Σαμολαδάς

ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη	3
1. Εισαγωγή	4
1.1 Σκοπός εργασίας	5
1.2 Δομή εργασίας	5
2. Δομή δικτύου και Επερωτήσεις	7
2.1 Κλιμακωτά δίκτυα αισθητήρων	7
2.2 Συνεχείς επερωτήσεις	9
2.3 Επερωτήσεις συνάθροισης	10
3. Προηγούμενες εργασίες	12
4. Γεωμετρική προσέγγιση	14
4.1 Υπολογιστικό μοντέλο	14
4.2 Λειτουργία δικτύου ενός cluster	15
4.3 Χαλάρωση απαιτήσεων	19
5. Εργασία	21
5.1 Συνάρτηση συνάθροισης	21
5.2 Εφαρμογή γεωμετρικής προσέγγισης στην αποστολή δεδομένων μεταξύ των κόμβων του δέντρου	22
5.3 Ανίχνευση παραβίασης	23
5.4 Η προς μετάδοση τιμή	23
5.5 Λειτουργία ρίζας δέντρου και συγχρονισμός	24
5.5.1 Πλήρης συγχρονισμός	24
5.5.2 Μερικός συγχρονισμός	25
6. Πειράματα	26
6.1 Συνάρτηση average	26
6.2 Συνάρτηση variance	30
7. Συμπεράσματα και μελλοντική εργασία	35
Βιβλιογραφία	36

Περίληψη:

Τα δίκτυα αισθητήρων (Wireless Sensor Networks) αποτελούν μία ραγδαία ανερχόμενη δικτυακή τεχνολογία με πολυάριθμες, σημαντικές εφαρμογές: περιβαλλοντικές, κατά την απομακρυσμένη παρακολούθηση μιας περιοχής ενδιαφέροντος (π.χ. εντοπισμός πυρκαγιών, πλημμύρων, σεισμικής και ηφαιστειακής δραστηριότητας), εφαρμογές στην οικονομία και τη βιομηχανία (π.χ. στη γεωργία ακριβείας, στη χρόνια παρακολούθηση της επιβάρυνσης και φθοράς ενός κτιρίου), εφαρμογές στην περιβάλλουσα νοημοσύνη (π.χ. στα λεγόμενα «έξυπνα σπίτια», στα «έξυπνα» ρούχα), στην υγεία (τηλε-παρακολούθηση ασθενών, βοήθεια σε άτομα με ειδικές ανάγκες), εφαρμογές ασφάλειας και άλλες πολλές εφαρμογές.

Οι περιορισμένες δυνατότητες των αισθητήρων σε αποθέματα ενέργειας καθώς και η ανάγκη για λήψη αξιόπιστων ειδοποιήσεων σε περιπτώσεις ανίχνευσης παραβιάσεων καθιστούν απαραίτητη την ανάπτυξη τεχνικών που θα εξασφαλίζουν στο δίκτυο μεγαλύτερη διάρκεια ζωής ενώ θα μπορούν παράλληλα να αναγνωρίσουν περιπτώσεις άμεσης ανάγκης και επέμβασης. Πολλά πρωτόκολλα και αλγόριθμοι έχουν αναπτυχθεί προς αυτή την κατεύθυνση, με έμφαση όμως στην αποτίμηση απλών συναρτήσεων, όπως ο υπολογισμός του αθροίσματος/μέσου όρου τιμών των αισθητήρων κτλ. Πρόσφατα προτάθηκε η γεωμετρική μέθοδος, η οποία επιτρέπει σε ένα δίκτυο κόμβων να ελέγχει με καταναμημένο τρόπο αν η τιμή μίας σύνθετης συνάρτησης, υπολογισμένη πάνω σε δεδομένα των κόμβων, βρίσκεται πάνω ή κάτω από ένα κατώφλι.

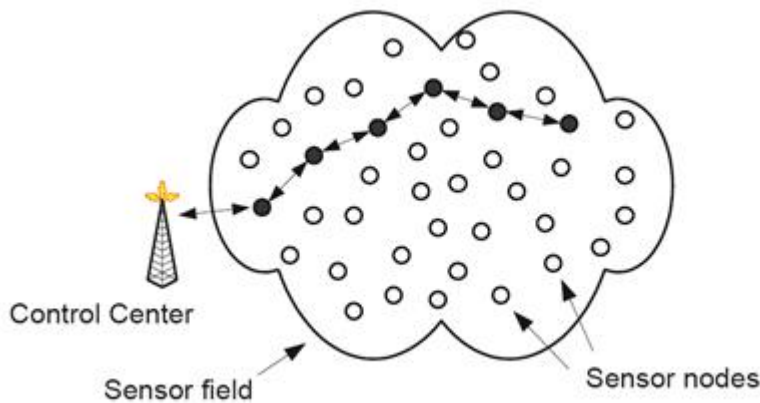
Στην εργασία μας αντιμετωπίσαμε το παραπάνω πρόβλημα σε ιεραρχικές τοπολογίες, δίνοντας έμφαση σε δίκτυα αισθητήρων, τα οποία οργανώνονται με φυσικό τρόπο σε δέντρα συλλογής δεδομένων. Σε κάθε τέτοιο δέντρο δημιουργούνται ιεραρχικές σχέσεις μεταξύ κάθε κόμβου με κόμβους πατέρα/παιδιών. Η δυνατότητα όμως των κόμβων για broadcast επικοινωνία επιτρέπει σε κάθε κόμβο να συνομιλεί με περισσότερους του ενός εν-δυνάμει πατεράδες. Χρησιμοποιώντας συναρτήσεις ορισμένες πάνω στο μέσο όρο των διανυσμάτων μέτρησης που συλλέγουν οι κόμβοι, καθώς και με την επιβολή περιορισμών (constraints) σε αυτούς και της παρακολούθησής τους μέσω της γεωμετρικής προσέγγισης, δώσαμε στους κόμβους τη δυνατότητα να μπορούν να ελέγχουν σε μία ευρύτερη περιοχή αν υπάρχει παραβίαση και να στέλνουν ειδοποίηση στην περίπτωση που οι καθορισμένοι περιορισμοί δεν ικανοποιούνται. Με αυτόν τον τρόπο κατορθώσαμε να μειώσουμε σημαντικά τον όγκο της μεταδιδόμενης πληροφορίας που στέλνεται συνολικά στο δίκτυο.

1. Εισαγωγή

Τα δίκτυα αισθητήρων αποτελούν μία ραγδαία ανερχόμενη δικτυακή τεχνολογία. Αποτελούνται από μεγάλο πλήθος μικροσκοπικών συσκευών που ενσωματώνουν δυνατότητες υπολογισμού, ασύρματης επικοινωνίας, περιορισμένη μνήμη, ενώ λειτουργούν με μικρή μπαταρία και φέρουν ποικιλία αισθητήρων για τη μέτρηση της θερμοκρασίας, του επιπέδου του φωτός, της υγρασίας, της κίνησης κλπ.

Ενώ κάθε μια συσκευή αυτή καθαυτή έχει μικρές δυνατότητες και υπόκειται σε αυστηρούς περιορισμούς (π.χ. περιορισμένη ενέργεια, τοπική γνώση), η κατανεμημένη αυτό-οργάνωση των συσκευών σε ένα αδόμητο δίκτυο και η αποδοτική και ευσταθής συνεργατική τους λειτουργία μπορεί να πραγματοποιήσει μεγάλης κλίμακας ανιχνεύσεις.

Η τυπική αποστολή ενός δικτύου αισθητήρων είναι ο εντοπισμός ενός τοπικού φαινομένου: ένας αισθητήρας ανιχνεύει τοπικά μια μεταβολή (π.χ. μια πυρκαγιά σε ένα δάσος). Στη συνέχεια, μεταδίδει σχετική πληροφορία προς ένα κέντρο αναφοράς με τη συνεργασία των υπόλοιπων αισθητήρων, όπως φαίνεται στο σχήμα που ακολουθεί.



Εικόνα 1

1.1 Σκοπός εργασίας

Στην παρούσα διπλωματική εργασία, ερευνήσαμε τη δυνατότητα σχεδίασης πρωτοκόλλων για την παρακολούθηση πολύπλοκων, μη γραμμικών συναρτήσεων εκφρασμένων πάνω σε δεδομένα που συλλέγουν οι αισθητήρες κόμβοι. Το παραπάνω πρόβλημα παρέχει την ανάγκη επέκτασης της γεωμετρικής μεθόδου ώστε να μπορεί να εφαρμοστεί σε ιεραρχικές τοπολογίες, με έμφαση σε δίκτυα αισθητήρων. Εκμεταλλευόμαστε την ιεραρχική τοπολογία που δημιουργείται κατά τη διάδοση επερωτήσεων σε δίκτυα αισθητήρων. Σε κάθε τέτοιο δέντρο δημιουργούνται ιεραρχικές σχέσεις μεταξύ κάθε κόμβου με κόμβους πατέρα/παιδιών. Η δυνατότητα όμως των κόμβων για broadcast επικοινωνία επιτρέπει σε κάθε κόμβο να συνομιλεί με περισσότερους του ενός εν-δυνάμει πατεράδες. Χρησιμοποιώντας συναρτήσεις ορισμένες πάνω στο μέσο όρο των διανυσμάτων μέτρησης που συλλέγουν οι κόμβοι, καθώς και με την επιβολή περιορισμών (constraints) σε αυτούς και της παρακολούθησής τους μέσω της γεωμετρικής προσέγγισης, δώσαμε στους κόμβους τη δυνατότητα να μπορούν να ελέγχουν σε μία ευρύτερη περιοχή αν υπάρχει παραβίαση και να στέλνουν ειδοποίηση στην περίπτωση που οι καθορισμένοι περιορισμοί δεν ικανοποιούνται. Με αυτόν τον τρόπο κατορθώσαμε να μειώσουμε σημαντικά τον όγκο της μεταδιδόμενης πληροφορίας που στέλνεται συνολικά στο δίκτυο.

Στην υλοποίησή μας, κάθε κόμβος ελέγχει αν ένας τοπικός περιορισμός έχει παραβιαστεί. Σε μία τέτοια περίπτωση, ο κόμβος προσπαθεί να ισορροπήσει την παραβίαση αυτή, μεταφέροντάς την αυτόματα σε έναν ή περισσότερους πιθανούς γονείς του στην ιεραρχική τοπολογία. Η υλοποίησή μας διαφοροποιείται από ήδη υπάρχουσες εργασίες τόσο ως προς τη μέθοδο αυτόματης εξισορρόπησης των τοπικών παραβιάσεων, όσο και ως προς τη λειτουργία του οργανωτή (ρίζας του δέντρου) στην περίπτωση που εντοπίσει μια πιθανή ολική παραβίαση.

1.2 Δομή εργασίας

Στην ενότητα 2 που ακολουθεί θα αναφερθούμε στη δομή των δικτύων αισθητήρων, στα μέρη από τα οποία αποτελούνται και τα χαρακτηριστικά τους. Θα μιλήσουμε επίσης για την εφαρμογή καθολικών συναρτήσεων πάνω στις μετρήσεις των αισθητήρων που βοηθούν στην απόκτηση μίας γενικότερης εικόνας της κατάστασης του δικτύου κάνοντας ιδιαίτερη αναφορά στις συναθροιστικές επερωτήσεις κατωφλίου (aggregate threshold queries).

Στην ενότητα 3 θα ασχοληθούμε με προηγούμενες εργασίες που έχουν γίνει πάνω στην εφαρμογή συναθροιστικών επερωτήσεων σε δίκτυα αισθητήρων, ενώ στην ενότητα 4 θα περιγράψουμε αναλυτικά τη γεωμετρική προσέγγιση και την εφαρμογή τοπικών περιορισμών (constraints) στους κόμβους του δικτύου που έχουν ως σκοπό τη μείωση της επικοινωνίας εντός αυτού.

Στην ενότητα 5 θα μιλήσουμε με λεπτομέρεια για το δίκτυο αισθητήρων που υποστηρίζει η παρούσα διπλωματική εργασία, ενώ παράλληλα θα περιγράψουμε το πώς εφαρμόζεται η γεωμετρική προσέγγιση σε συνδυασμό με νέες τεχνικές που έχουμε αναπτύξει στο δίκτυο αυτό.

Η ενότητα 6 περιλαμβάνει τα πειράματα που πραγματοποιήθηκαν για τις ανάγκες της συγκεκριμένης εργασίας, ενώ στην ενότητα 7 περιέχουμε τα συμπεράσματα αυτής καθώς και ενδεχόμενες μελλοντικές εργασίες και εφαρμογές της.

2. Δομή δικτύου αισθητήρων και Επερωτήσεις

Τα δίκτυα αισθητήρων έχουν πολυάριθμες, σημαντικές εφαρμογές: περιβαλλοντικές, κατά την απομακρυσμένη παρακολούθηση μιας περιοχής ενδιαφέροντος (π.χ. εντοπισμός πυρκαγιών, πλημμύρων, σεισμικής και ηφαιστειακής δραστηριότητας), εφαρμογές στην οικονομία και τη βιομηχανία (π.χ. στη γεωργία ακριβείας, στη χρόνια παρακολούθηση της επιβάρυνσης και φθοράς ενός κτιρίου), εφαρμογές στην περιβάλλουσα νοημοσύνη (π.χ. στα λεγόμενα «έξυπνα σπίτια», στα «έξυπνα» ρούχα), στην υγεία (τηλε-παρακολούθηση ασθενών, βοήθεια σε άτομα με ειδικές ανάγκες), εφαρμογές ασφάλειας και άλλες πολλές εφαρμογές.

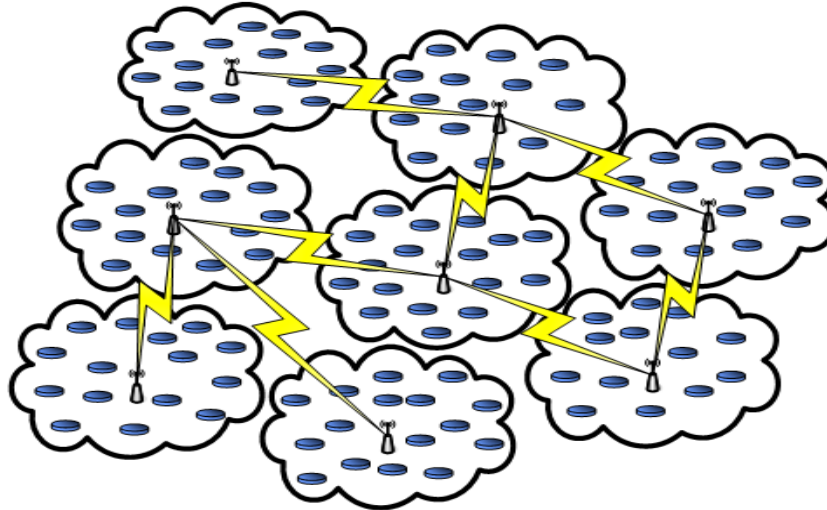
Πολλά πρωτόκολλα και αλγόριθμοι έχουν αναπτυχθεί για τη λειτουργία και την επικοινωνία των αισθητήρων μεταξύ τους. Οι αλγόριθμοι αυτοί στοχεύουν κυρίως στη μείωση της ποσότητας ενέργειας που δαπανά ένας αισθητήρας για τη λειτουργία του, κάτι το οποίο οδηγεί σε αύξηση του χρόνου ζωής του δικτύου. Επίσης, στοχεύουν στην αποδοτική και αξιόπιστη δρομολόγηση των πακέτων στο δίκτυο αισθητήρων καθώς και στην ευρωστία και αυτονομία αυτού.

Στις πλείστες των ρεαλιστικών εφαρμογών τα δεδομένα που συλλέγονται από το ασύρματο δίκτυο αισθητήρων αποθηκεύονται στη μορφή αριθμητικών τιμών σε κάποιο κεντρικό σταθμό βάσης. Ο σταθμός αυτός είναι κάποια υπολογιστική μηχανή μεγαλύτερης ισχύος με ικανότητες επεξεργασίας και αποθήκευσης μεγάλης ποσότητας δεδομένων και φυσικά επικοινωνίας με τον έξω κόσμο.

2.1 Κλιμακωτά δίκτυα αισθητήρων

Σε πολλές υλοποιήσεις των δικτύων αισθητήρων, το δίκτυο αποτελείται από δύο μονάδες: πολλές φθηνές ασύρματες συσκευές ανίχνευσης, γνωστές ως κόμβοι (nodes), και λιγότερες, πιο ισχυρές μονάδες ελέγχου, γνωστές ως μάκρο – κόμβοι (macro-nodes). Τα nodes είναι απλές, ασύρματες μονάδες με βάση απλούς 8-bit επεξεργαστές. Ωστόσο, οι δυνατότητές τους για επεξεργασία και αποθήκευση δεδομένων καθώς και σε αποθέματα ενέργειας, είναι περιορισμένες. Αντίθετα, οι μάκρο – κόμβοι είναι πιο πλούσιοι σε πόρους όσον αφορά την επεξεργασία δεδομένων, τη μνήμη καθώς και τις ασύρματες επικοινωνίες. Βασίζονται συνήθως σε 32-bit επεξεργαστές ενώ έχουν μεγάλα αποθέματα ενέργειας σε σχέση με τα nodes.

Το δίκτυο χωρίζεται σε μονάδες (clusters), όπως φαίνεται στην Εικόνα 2 που ακολουθεί. Κάθε cluster αποτελείται από ένα σύνολο από motes, που αποτελούν με τη σειρά τους τα μέλη του cluster, και ένα μοναδικό μάκρο – κόμβο, που ονομάζεται αρχηγός του cluster (cluster head).



Εικόνα 2

Ο ρόλος των motes περιορίζεται στο να ανιχνεύουν το περιβάλλον με βάση τις οδηγίες που έχουν λάβει από τον cluster head, καθώς και να υποβάλλουν σε αυτόν τις μετρήσεις τους. Επιπλέον, ένα mote είναι δυνατόν να συμμετέχει στη multi – hop δρομολόγηση δεδομένων που προέρχονται από άλλο mote του cluster και έχουν ως τελικό αποδέκτη τον cluster head. Οι μάκρο – κόμβοι είναι υπεύθυνοι για την εκτέλεση της λειτουργικότητας σε επίπεδο εφαρμογής. Χρησιμοποιούν τις ανώτερες ικανότητες επικοινωνίας που διαθέτουν ώστε να υλοποιήσουν έναν ευρύ δικτυακό κορμό επικοινωνιών. Η δρομολόγηση (routing) μεταξύ των μάκρο – κόμβων γίνεται μόνο από τους μάκρο – κόμβους, δηλαδή τα motes δεν συμμετέχουν στην ανταλλαγή μηνυμάτων μεταξύ των μάκρο – κόμβων.

Αυτή η κλιμακωτή αρχιτεκτονική έχει πολλά πλεονεκτήματα. Εργασίες [3,6,14,16] έδειξαν ότι αυτό το μοντέλο έχει καλύτερες επιδόσεις σε σχέση με την περίπτωση που όλοι οι κόμβοι βρίσκονται σε ένα ενιαίο cluster. Δεδομένου ότι το κόστος των motes είναι χαμηλό και ότι συνήθως ο αριθμός τους είναι μία ή δύο τάξεις μεγέθους μεγαλύτερος από τον αριθμό των μάκρο – κόμβων, η ανάπτυξη ενός κλιμακωτού δικτύου είναι σχετικά ανέξοδη.

2.2 Συνεχείς επερωτήσεις

Η παρακολούθηση ροών δεδομένων (data stream) σε καταναμημένα συστήματα αποτελεί το επίκεντρο των ερευνών για πολλά χρόνια. Τα περισσότερα σενάρια ασχολούνται με την παρακολούθηση απλών συναθροιστικών τιμών, όπως για παράδειγμα η συχνότητα εμφάνισης ενός αντικειμένου σε μία ροή. Κοινή απαίτηση πολλών εφαρμογών είναι η επεξεργασία σε πραγματικό χρόνο συνεχών ροών δεδομένων μεγάλου όγκου, όπως για παράδειγμα η ανάλυση οικονομικών δεδομένων ή η ανίχνευση εισβολής. Η –πραγματικού χρόνου – φύση των συστημάτων ροής δεδομένων, καθώς και ο τεράστιος όγκος δεδομένων που καλούνται να επεξεργαστούν, εισάγουν θεμελιώδη προβλήματα που δεν καλύπτονται από τα παραδοσιακά Συστήματα Διαχείρισης Βάσεων Δεδομένων (Database Management Systems – DBMS, ΣΔΒΔ). Τα παραδοσιακά ΣΔΒΔ επεξεργάζονται επερωτήσεις που θέτει ο χρήστης και επιστρέφουν το αποτέλεσμα. Στα συστήματα ροής δεδομένων, οι χρήστες θέτουν συνεχείς επερωτήσεις στο σύστημα, αυτό τις επεξεργάζεται κατά την άφιξή τους και παρέχει στο χρήστη τα εκάστοτε ανανεωμένα αποτελέσματα.

Μία χρήσιμη κατηγορία επερωτήσεων στο πλαίσιο των καταναμημένων ροών δεδομένων είναι οι επερωτήσεις παρακολούθησης (monitoring queries). Ένας πιο γενικός τύπος επερωτήσεων παρακολούθησης είναι ο ακόλουθος:

Έστω X_1, X_2, \dots, X_n η συχνότητα που μετράται για n στοιχεία σε ένα σύνολο ροών δεδομένων και $f(X_1, X_2, \dots, X_n)$ μία αυθαίρετη συνάρτηση πάνω στη μετρούμενη συχνότητα. Μας ενδιαφέρει να καθορίσουμε πότε η τιμή της συνάρτησης $f(X_1, X_2, \dots, X_n)$ αυξάνεται περισσότερο ή πέφτει κάτω από ένα προκαθορισμένο όριο - κατώφλι. Αυτού του είδους οι επερωτήσεις είναι γνωστές ως επερωτήσεις συνάρτησης κατωφλίου (threshold function queries).

Υπάρχει μία θεμελιώδης διαφορά μεταξύ γραμμικών και μη γραμμικών συναρτήσεων f , που μπορεί να αποδειχθεί ακόμα και για την περίπτωση μονοδιάστατων δεδομένων. Έστω x_1 και x_2 τιμές που έχουν αποθηκευτεί σε δύο διαφορετικούς κόμβους, και έστω $f(x) = 6x - x^2$. Ας υποθέσουμε ότι χρειάζεται να καθορίσουμε εάν $f(\frac{x_1 + x_2}{2}) > 1$. Αν η f ήταν γραμμική, η λύση θα ήταν απλή από τη στιγμή που ισχύει ότι $f(\frac{x_1 + x_2}{2}) = \frac{f(x_1) + f(x_2)}{2}$. Αν υποθέσουμε πως αρχικά η τιμή σε κάθε κόμβο είναι < 1 , τότε ένας απλός καταναμημένος αλγόριθμος που θα παρακολουθεί αν $f(\frac{x_1 + x_2}{2}) > 1$,

είναι για κάθε κόμβο i να μη στέλνει ειδοποίηση όσο $f(x_i) < 1$. Όμως, ακόμα και για την παραπάνω απλή μη γραμμική συνάρτηση, είναι αδύνατο να καθοριστεί από τις τιμές της f στους κόμβους αν η τιμή της στο μέσο όρο είναι μεγαλύτερη από 1 ή όχι. Για παράδειγμα, αν $x_1=0$ και $x_2=6$, τότε η τιμή της f για κάθε κόμβο είναι <1 , ενώ η τιμή της στο μέσο όρο των x_1, x_2 είναι 9. Αν όμως $x_1=10$ και $x_2=20$, τότε και η τιμή της συνάρτησης για κάθε κόμβο αλλά και η τιμή της πάνω στο μέσο όρο των x_1, x_2 είναι <1 .

Σε αυτό το παράδειγμα, το κόστος αποστολής δεδομένων που είναι αποθηκευμένα στους κόμβους είναι το ίδιο με το κόστος αποστολής μίας τιμής, αλλά σε data mining εφαρμογές, τα δεδομένα είναι δυνατόν να έχουν πολλές διαστάσεις. Αυτό καθιστά απαραίτητη την ανάπτυξη καταναμημένου αλγορίθμου που θα καθορίζει τοπικά αν η τιμή της συνάρτησης f πάνω στο average data vector ξεπερνά ένα προκαθορισμένο κατώφλι.

2.3 Επερωτήσεις συνάθροισης

Σκοπός ενός δικτύου αισθητήρων είναι να παρέχει μία γενική εικόνα για την κατάσταση του περιβάλλοντος που παρακολουθείται και όχι η παρουσίαση μιας σειράς μετρήσεων που παίρνουν οι αισθητήρες. Μία μέθοδος για την απόκτηση καθολικής εικόνας της κατάστασης του δικτύου είναι η εφαρμογή καθολικών συναθροίσεων πάνω στις μετρήσεις των αισθητήρων: ας πάρουμε για παράδειγμα ένα δίκτυο αισθητήρων που παρακολουθεί ένα οικοσύστημα και στο οποίο τίθενται επερωτήσεις με σκοπό να καθοριστεί η μέση θερμοκρασία στην υπό μελέτη περιοχή.

Τα τελευταία χρόνια έχουν προταθεί πολλά μοντέλα για τη συνάθροιση δεδομένων σε δίκτυα αισθητήρων. Αν και τα σενάρια αυτά επιτρέπουν τον αποδοτικό υπολογισμό συναθροιστικών τιμών, το αποτέλεσμά τους υπολογίζεται βάσει των δεδομένων που έχουν αποθηκεύσει οι αισθητήρες μία συγκεκριμένη χρονική στιγμή. Υπάρχουν ωστόσο αλγόριθμοι που επιτρέπουν την εφαρμογή συνεχών επερωτήσεων στο δίκτυο και στους οποίους η υπολογιζόμενη συναθροιστική τιμή (aggregate) ανανεώνεται συνεχώς με σκοπό να αντικατοπτρίζει τις τρέχουσες μετρήσεις των αισθητήρων, ελαχιστοποιώντας παράλληλα τον όγκο των δεδομένων που θα χρειαστεί να μεταδώσει ένας αισθητήρας. Αν και η συνεχής παροχή μιας καθολικής ενημερωμένης τιμής μπορεί να είναι χρήσιμη, υπάρχουν περιπτώσεις όπου απαιτούμε από το δίκτυο να αναφέρει μόνο συνθήκες συναγερού, οι οποίες χαρακτηρίζονται από μία καθολική συναθροιστική τιμή που παραβιάζει κάποιο κατώφλι. Αυτό το είδος των συνεχών επερωτήσεων χαρακτηρίζονται ως συναθροιστικές επερωτήσεις κατωφλίου (aggregate threshold queries).

Σκοπός των συναθροιστικών επερωτήσεων κατώφλιου είναι να παρακολουθούν το δίκτυο και να ειδοποιούν κάθε φορά που μία συναθροιστική τιμή ξεπερνά ένα προκαθορισμένο κατώφλι. Είναι ιδιαίτερα σημαντικές καθώς επιτρέπουν την ανίχνευση γεγονότων, όπως ο υπολογισμός της διακύμανσης στις μετρήσεις των αισθητήρων που υπερβαίνουν ένα συγκεκριμένο όριο, ελαχιστοποιώντας τον όγκο της πληροφορίας που θα χρειαστεί να μεταδώσει ένας αισθητήρας και κατ' επέκταση και την ποσότητα της ενέργειας που χρειάζεται να δαπανήσει ο αισθητήρας για να μεταδώσει.

Μία συνεχής συναθροιστική επερώτηση κατώφλιου μπορεί να οριστεί με τον ακόλουθο τρόπο: Ας υποθέσουμε ότι ένα σύνολο μετρήσεων που παίρνει ένα mote, αναπαριστάται από ένα vector με πραγματικές τιμές. Έστω μία αυθαίρετη συναθροιστική συνάρτηση, η οποία λαμβάνει το vector με τις μετρήσεις (measurement vector) και επιστρέφει μία πραγματική τιμή. Θέλουμε για οποιαδήποτε χρονική στιγμή να καθορίσουμε εάν η τιμή της συναθροιστικής συνάρτησης πάνω στο μέσο όρο των measurement vectors που κρατά κάθε mote, ξεπερνά ένα προκαθορισμένο κατώφλι.

3. Προηγούμενες εργασίες

Οι συναθροιστικές επερωτήσεις έχουν λάβει ιδιαίτερη προσοχή τα τελευταία χρόνια [2,4,5,6,7,9,10,13,17]. Το [7] προτείνει αλγόριθμο, στον οποίο οι κόμβοι διατάσσονται σε δέντρο, για την εκτέλεση επερωτήσεων συνάθροισης. Ο αλγόριθμος αυτός αποτελείται από δύο φάσεις: στην πρώτη φάση, η επερώτηση διαδίδεται σε όλο το δέντρο από πάνω προς τα κάτω. Όταν τη λάβουν τα φύλλα του δέντρου, αυτά θα στείλουν στους γονείς τους τη μέτρησή τους. Όταν ένας αισθητήρας λάβει τα δεδομένα που του έχουν στείλει όλα τα παιδιά του, τότε θα συναθροίσει τη μέτρηση του με τη συνολική τιμή που έχει λάβει από τα παιδιά του και θα στείλει με τη σειρά του το νέο aggregate στον πατέρα του. Στο [13] παρουσιάζεται ένας αλγόριθμος για προσεγγιστικά quantiles σε δίκτυα αισθητήρων, όπου κάθε αισθητήρας μεταδίδει ένα σταθερό αριθμό από bits. Οι αλγόριθμοι που βασίζονται σε δέντρο υπολογίζουν ένα aggregate που βασίζεται σε ένα στιγμιότυπο των δεδομένων.

Εναλλακτική προσέγγιση στη δενδρική δομή που περιγράψαμε παραπάνω, αποτελεί η multi – path προσέγγιση. Στην περίπτωση αυτή, οι αισθητήρες χωρίζονται σε ένα σύνολο από δακτύλιους. Ο i -οστός δακτύλιος αποτελείται από όλους τους αισθητήρες που απέχουν i hops από τον σταθμό βάσης. Η συνάθροιση εκτελείται από τον πιο μακρινό δακτύλιο με κατεύθυνση τον σταθμό βάσης. Σε αντίθεση με την δεντρική προσέγγιση όπου κάθε αισθητήρας στέλνει το aggregate που υπολογίζει στον πατέρα του, στη multi – path προσέγγιση κάθε κόμβος κάνει broadcast το aggregate που υπολογίζει. Στη συνέχεια, οι αισθητήρες του επόμενου δακτυλίου που λαμβάνουν το aggregate, το επεξεργάζονται. Τα ενδιάμεσα aggregates αναπαριστούνται από δομές σκίτσων ανθεκτικές σε διπλά αποτελέσματα που μπορεί να έχουν προκύψει από το γεγονός ότι το aggregate επεξεργάζεται από πολλαπλούς αισθητήρες στον επόμενο δακτύλιο. Η προσέγγιση αυτή είναι ανθεκτική σε απώλεια πακέτων, όμως το ότι η συνάθροιση των δεδομένων εκτελείται πάνω σε σκίτσα των τιμών και όχι στις ίδιες τις τιμές, παρέχει προσεγγιστικά aggregates, ενώ η δενδρική προσέγγιση παρέχει ακριβή αποτελέσματα όταν η παράδοση των δεδομένων είναι αξιόπιστη.

Το [9] παρουσιάζει έναν αλγόριθμο που συνδυάζει τα πλεονεκτήματα των δύο παραπάνω προσεγγίσεων. Ο αλγόριθμος αυτός χρησιμοποιεί τη δενδρική προσέγγιση για να συναθροίσει τα δεδομένα των αισθητήρων που βρίσκονται μακριά από τον σταθμό βάσης. Επιπρόσθετα, εφαρμόζει τη multi – path προσέγγιση για να συναθροίσει τα δεδομένα όσων κόμβων βρίσκονται κοντά στο σταθμό βάσης. Τέλος, καθορίζει δυναμικά πότε θα κάνει αλλαγή από τη μία προσέγγιση στην άλλη ανάλογα με τις συνθήκες που

επικρατούν στο δίκτυο. Σε αντίθεση με τους αλγόριθμους που κάνουν χρήση της δενδρικής και της multi – path προσέγγισης, και οι οποίοι, όπως αναφέραμε παραπάνω, υπολογίζουν aggregates με βάση στιγμιότυπα δεδομένων, ο αλγόριθμος του [9] μπορεί να χειριστεί συνεχή επερωτήσεις.

Στο [17] παρουσιάζεται αλγόριθμος για συνεχή εκτίμηση απλών συναθροιστικών τιμών, όπως sum, count, average, σε ένα δίκτυο αισθητήρων. Ο αλγόριθμος αυτός παρέχει συνεχώς μία εκτίμηση της συναθροιστικής τιμής. Ο αλγόριθμος του [1], πάνω στον οποίο βασίστηκε η υλοποίησή μας, διαφέρει από αυτόν του [17] ως προς το ότι χειρίζεται συναθροιστικές επερωτήσεις κατωφλίου, αλλά και στο ότι υποστηρίζει aggregates που μπορούν να εκφραστούν από αυθαίρετες συναρτήσεις και όχι απαραίτητα μόνο από απλές συναθροιστικές συναρτήσεις.

Αλγόριθμοι για την παρακολούθηση αυθαίρετων συναρτήσεων κατωφλίου σε κατανεμημένες ροές έχουν παρουσιαστεί στο [11]. Το [11] παρουσιάζει δύο αλγόριθμους παρακολούθησης για δύο διαφορετικά υπολογιστικά περιβάλλοντα. Ο πρώτος αλγόριθμος υποθέτει ότι κάθε μήνυμα που στέλνει ένας κόμβος, λαμβάνεται από όλους τους υπόλοιπους κόμβους. Ο δεύτερος αλγόριθμος προσδιορίζει έναν από τους κόμβους ως τον συντονιστή (coordinator) και υποθέτει ότι κάθε κόμβος επικοινωνεί με τον συντονιστή και μόνο. Και οι δύο αυτοί αλγόριθμοι βασίζονται στη γεωμετρική προσέγγιση του προβλήματος παρακολούθησης. Κανένας ωστόσο από τους αλγόριθμους στο [11] δεν είναι κατάλληλος για δίκτυα αισθητήρων καθώς η υλοποίηση των υποθέσεών τους είναι πολύ δαπανηρή όσον αφορά την κατανάλωση ενέργειας.

Το [1] προσαρμόζει τη γεωμετρική προσέγγιση που παρουσιάζεται στο [11] αλλά προτείνει έναν αλγόριθμο που μπορεί να επιλύει τις συγκρούσεις τοπικά και είναι ευαίσθητος σε περιορισμούς κατανάλωσης ενέργειας. Ο αλγόριθμος αυτός είναι σχεδιασμένος για κλιμακωτά δίκτυα αισθητήρων.

Η υλοποίησή μας βασίστηκε στη γεωμετρική προσέγγιση του [11] και στον αλγόριθμο του [1], και εφαρμόστηκε σε δίκτυο αισθητήρων, στο οποίο όλοι οι κόμβοι συγκροτούν ένα ενιαίο cluster.

4. Γεωμετρική Προσέγγιση

Η γεωμετρική προσέγγιση μελετά προσεγγίσεις της εξής μορφής: Υποθέτουμε ότι ένας κόμβος κρατά ένα διάνυσμα μετρήσεων d -διαστάσεων, ενώ μας έχει δοθεί μία αυθαίρετη συνάρτηση ελέγχου, η οποία είναι ορισμένη πάνω στο μέσο όρο των διανυσμάτων μέτρησης που συλλέγουν όλοι οι κόμβοι του δικτύου. Μας ενδιαφέρει να ανιχνεύσουμε πότε η τιμή της συνάρτησης ξεπερνά ένα προκαθορισμένο όριο. Η προσέγγιση αυτή καθορίζει έναν τοπικό περιορισμό (constraint) σε κάθε κόμβο και εγγυάται πως όσο οι τοπικοί αυτοί περιορισμοί δεν παραβιάζονται, τότε η τιμή της ελεγχόμενης συνάρτησης δεν έχει ξεπεράσει το προκαθορισμένο όριο.

4.1 Υπολογιστικό Μοντέλο

Στην εργασία [1] οι κόμβοι του δικτύου οργανώνονται σε συστάδες (clusters). Έστω n ο αριθμός των clusters στο δίκτυο, N_i ο αριθμός των motes στο i -οστό cluster και N_{tot} ο συνολικός αριθμός των motes στο δίκτυο. Έστω ακόμα ότι $s_{i,1}, s_{i,2}, \dots, s_{i,N_i}$ τα motes και m_i ο αρχηγός του cluster στο i -οστό cluster. Υποθέτουμε πως όλα τα motes σε ένα cluster (ή τουλάχιστον η πλειοψηφία αυτών) επικοινωνούν απευθείας με τον αρχηγό του cluster και πως όλες οι ραδιοζεύξεις είναι πλήρως αμφίδρομες.

Ο αρχηγός κάθε cluster, εκτός από το να επικοινωνεί με τα motes που βρίσκονται στο ίδιο με αυτόν cluster, μπορεί και επικοινωνεί με τους αρχηγούς γειτονικών clusters. Οι ζεύξεις μεταξύ των αρχηγών των clusters μοντελοποιούνται από έναν μη κατευθυνόμενο συνδετικό γράφο, όπου κάθε cluster αντιπροσωπεύεται από μία κορυφή στο γράφο, ενώ η ακμή που συνδέει δύο κορυφές αντιπροσωπεύει τα γειτονικά clusters. Υποθέτουμε ότι ο συνδετικός γράφος αντιστοιχίζεται σε δέντρο συλλογής δεδομένων. Υποθέτουμε επίσης ότι μία επερώτηση διαχέεται στο δίκτυο από έναν από τους αρχηγούς των clusters, ο οποίος συμβολίζεται m_1 .

Μία συνεχής συναθροιστική επερώτηση κατωφλίου (aggregate threshold query) ορίζεται ως εξής: Κάθε mote $s_{i,j}$ κρατάει ένα vector μετρήσεων d -διάστασης, το οποίο ονομάζεται measurement vector και συμβολίζεται ως $\vec{v}_{i,j} = (v_{i,j}^{(1)}, v_{i,j}^{(2)}, \dots, v_{i,j}^{(d)})^T$.

Ορίζουμε το καθολικό (global) measurement vector, $\vec{v} = (v^{(1)}, v^{(2)}, \dots, v^{(d)})^T$ ως τον μέσο όρο των measurement vectors που έχουν κρατήσει όλα τα motes στο δίκτυο:

$$\vec{v} = \frac{1}{N_{tot}} \sum_{i=1}^n \sum_{j=1}^{N_i} \vec{v}_{i,j}$$

Έστω αυθαίρετη συναθροιστική συνάρτηση $f : \mathbb{R}^d \rightarrow \mathbb{R}$ και r ένα προκαθορισμένο κατώφλι. Μία συνεχής συναθροιστική επερώτηση κατωφλίου απαιτεί ότι οποιοσδήποτε αισθητήρας στο δίκτυο είναι ικανός να καθορίσει κάθε χρονική στιγμή αν $f(v) > r$ ή όχι.

4.2 Λειτουργία Δικτύου ενός cluster

Ανά διαστήματα ο αρχηγός του cluster συγκεντρώνει τα measurement vectors από όλα τα motes, υπολογίζει τον μέσο όρο τους και στη συνέχεια στέλνει σε αυτά το average vector που υπολόγισε. Αυτή η δράση αναφέρεται ως συγχρονισμός (synchronization).

Το average vector που υπολογίζει ο αρχηγός του cluster κατά τον τελευταίο συγχρονισμό ονομάζεται estimate vector και συμβολίζεται με \vec{e} . Κάθε mote θυμάται το measurement vector που είχε στείλει στον αρχηγό του cluster κατά τον τελευταίο συγχρονισμό. Η μέτρηση που παίρνει ο κόμβος s_i κατά τον τελευταίο συγχρονισμό αναφέρεται ως reference vector και συμβολίζεται ως \vec{v}_i' .

Σύμφωνα με τα παραπάνω θα ισχύει:

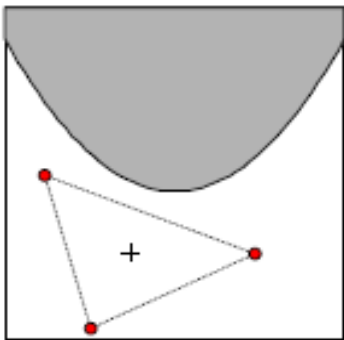
$$\vec{e} = \frac{1}{N} \sum_{i=1}^N \vec{v}_i'$$

Κάθε mote διατηρεί δύο επιπλέον μεταβλητές. Η πρώτη μεταβλητή ονομάζεται slack vector και συμβολίζεται με $\vec{\delta}_i$. Ο αλγόριθμος εγγυάται πως για οποιαδήποτε χρονική στιγμή το άθροισμα των slack vectors όλων των motes είναι μηδέν ($\sum_{i=1}^n \vec{\delta}_i = 0$). Μετά από κάθε συγχρονισμό, κάθε mote θέτει το slack vector του ίσο με μηδέν. Η δεύτερη μεταβλητή που κρατά κάθε κόμβος ονομάζεται drift vector και συμβολίζεται με \vec{u} . Το drift vector υπολογίζεται από την ακόλουθη σχέση:

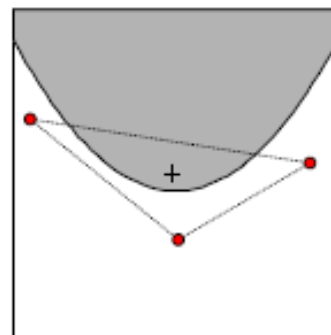
$$\vec{u}_i = \vec{e} + \vec{v}_i - \vec{v}_i' + \vec{\delta}_i$$

Κάθε χρονική στιγμή ο μέσος όρος των drift vectors που κρατούν τα motes ισούται με το καθολικό measurement vector. Επιπλέον, αμέσως μετά τον συγχρονισμό, για κάθε mote, το reference vector ισούται με το measurement vector, δηλαδή $\vec{v}_i = \vec{v}_i'$, ενώ το slack vector καθενός mote γίνεται μηδέν. Συνεπώς, αμέσως μετά τον συγχρονισμό, το drift vector κάθε κόμβου θα ισούται με το estimate vector.

Τα vectors που κρατούν τα motes συμβολίζονται ως σημεία στον \mathbb{R}^d . Ο συνδυασμός της συναθροιστικής συνάρτησης f , και της τιμής του κατωφλίου r , χρωματίζουν τον \mathbb{R}^d : κάθε σημείο $\vec{x} \in \mathbb{R}^d$ για το οποίο ισχύει ότι $f(\vec{x}) < r$ θα είναι λευκό, ενώ κάθε σημείο $\vec{y} \in \mathbb{R}^d$ για το οποίο $f(\vec{y}) \geq r$ θα είναι γκρι. Σκοπός είναι να καθοριστεί το χρώμα του σημείου που αντιπροσωπεύει το καθολικό measurement vector. Στις Εικόνες 3 και 4 που ακολουθούν, με κόκκινους κύκλους απεικονίζονται τα drift vectors που κρατούν τρεις κόμβοι του δικτύου ενώ ο σταυρός αναπαριστά το καθολικό measurement vector που ορίζουν, δηλαδή το μέσο όρο των measurement vectors τους. Και στις δύο εικόνες παρατηρούμε πως η περιοχή που βρίσκονται τα drift vectors (κόκκινοι κύκλοι) είναι λευκή, αλλά η περιοχή στην οποία βρίσκεται ο σταυρός στην Εικόνα 4 είναι γκρι. Από αυτό συμπεραίνουμε πως μία επερώτηση δεν μπορεί να απαντηθεί παρατηρώντας μόνο το χρώμα των drift vectors που κρατούν οι κόμβοι.



Εικόνα 3



Εικόνα 4

Κάθε φορά που αλλάζουν οι μετρήσεις που παίρνει ένας κόμβος, ο κόμβος ελέγχει αν το νέο vector μετρήσεων συμμορφώνεται με έναν τοπικό περιορισμό που έχει οριστεί. Αυτοί οι περιορισμοί κατασκευάζονται με τέτοιο τρόπο ώστε αν ικανοποιούνται για όλα τα motes, τότε εγγυημένα το convex hull που σχηματίζουν τα drift vectors θα είναι μονοχρωματικό. Δηλαδή όλα τα vectors που βρίσκονται μέσα στο convex hull θα έχουν

το ίδιο χρώμα. Από τη στιγμή που το καθολικό measurement vector υπολογίζεται ως ο μέσος όρος των drift vectors, θα βρίσκεται σίγουρα στο convex hull που ορίζουν τα drift vectors. Συνεπώς, αν το convex hull είναι μονοχρωματικό, το καθολικό measurement vector θα έχει το ίδιο χρώμα με το convex hull.

Για να καθορίσουμε αν το convex hull που ορίζουν τα drift vectors είναι μονοχρωματικό χρησιμοποιήσαμε το παρακάτω θεώρημα, του οποίου η απόδειξη εμπεριέχεται στο [13]:

Θεώρημα : Έστω $\vec{x}, \vec{y}_1, \vec{y}_2, \dots, \vec{y}_n \in \mathbb{R}^d$ ένα σύνολο από vectors στον \mathbb{R}^d . Έστω $\text{Conv}(\vec{x}, \vec{y}_1, \vec{y}_2, \dots, \vec{y}_n)$ το convex hull των $\vec{x}, \vec{y}_1, \vec{y}_2, \dots, \vec{y}_n$. Έστω $B(\vec{x}, \vec{y}_i)$ σφαίρα με κέντρο στο $\frac{1}{2}(\vec{x} + \vec{y}_i)$ και ακτίνα $\left\| \frac{1}{2}(\vec{x} - \vec{y}_i) \right\|_2$. Με άλλα λόγια, $B(\vec{x}, \vec{y}_i) = \{ \vec{z} \mid \left\| \vec{z} - \frac{1}{2}(\vec{x} + \vec{y}_i) \right\|_2 \leq \left\| \frac{1}{2}(\vec{x} - \vec{y}_i) \right\|_2 \}$. Έτσι, $\text{Conv}(\vec{x}, \vec{y}_1, \vec{y}_2, \dots, \vec{y}_n) \subset \bigcup_{i=1}^n B(\vec{x}, \vec{y}_i)$.

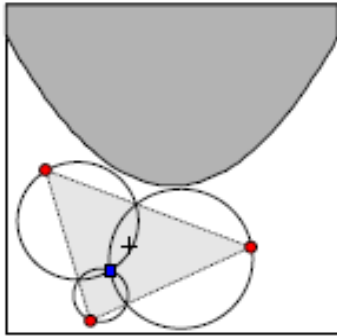
Το θεώρημα αυτό χρησιμοποιείται για να φράξει το convex hull n+1 vectors στον \mathbb{R}^d από την ένωση n d-διάστατων σφαιρών. Στην περίπτωση που μελετάμε, μας ενδιαφέρει να φράξουμε το convex hull που ορίζουν τα estimate vector και drift vectors, δηλαδή $\text{Conv}(\vec{e}, \vec{u}_1, \vec{u}_2, \dots, \vec{u}_n)$, από ένα σύνολο n σφαιρών, όπου κάθε σφαίρα κατασκευάζεται από ένα mote. Κάθε mote s_i κατασκευάζει μία σφαίρα $B(\vec{e}, \vec{u}_i)$, η οποία έχει κέντρο στο $\frac{\vec{e} + \vec{u}_i}{2}$ και ακτίνα ίση με $\left\| \frac{1}{2}(\vec{e} - \vec{u}_i) \right\|$. Αυτή η σφαίρα ονομάζεται drift σφαίρα.

Οποιαδήποτε χρονική στιγμή, κάθε κόμβος έχει όλη την απαραίτητη πληροφορία ώστε να μπορεί να κατασκευάσει τη drift σφαίρα του.

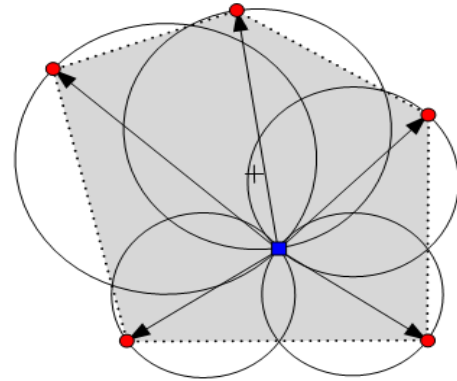
Το παραπάνω θεώρημα αναφέρει ότι το convex hull που ορίζουν estimate vector και drift vectors φράσσεται από την ένωση των drift σφαιρών που κατασκευάζουν οι κόμβοι. Συνεπώς, αν όλες οι σφαίρες είναι μονοχρωματικές, τότε το convex hull θα είναι σίγουρα μονοχρωματικό, και έτσι καθολικό measurement vector θα έχει το ίδιο χρώμα με το convex hull. Αφού το estimate vector αποτελεί μέρος του convex hull, αν οι drift σφαίρες είναι μονοχρωματικές, τότε το estimate vector θα έχει το ίδιο χρώμα με το καθολικό measurement vector. Για να ελέγξουμε αν μία σφαίρα είναι μονοχρωματική, υπολογίζουμε τη μέγιστη και την ελάχιστη τιμή της συνάρτησης f σε αυτή.

Οι Εικόνες 5 και 6 που ακολουθούν αποτελούν εφαρμογή του θεωρήματος που μελετήσαμε. Το μπλε τετράγωνο αντιπροσωπεύει το estimate vector και βάσει αυτού και του drift vector, κάθε mote κατασκευάζει τη σφαίρα του. Μπορεί εύκολα να

παρατηρήσει κανείς ότι το convex hull που ορίζουν τα drift vectors φράσσεται από την ένωση drift σφαιρών.



Εικόνα 5



Εικόνα 6

Στο σημείο αυτό θα επικεντρωθούμε στον τρόπο με τον οποίο επιλύονται τυχόν παραβιάσεις στους τοπικούς περιορισμούς που έχουν τεθεί σε κάθε κόμβο. Μία μέθοδος για να επιτευχθεί αυτό είναι η διαδικασία του συγχρονισμού. Όπως αναφέραμε παραπάνω στο κεφάλαιο, ο συγχρονισμός των motes παράγει ένα νέο estimate vector και οδηγεί στο να θέσουν οι κόμβοι τα drift vectors ίσα με το estimate vector. Αυτό έχει ως αποτέλεσμα, οι νέες drift σφαίρες που θα σχηματίσουν οι κόμβοι να έχουν μηδενική ακτίνα, κατ' επέκταση να είναι μονοχρωματικές και έτσι οι περιορισμοί γίνονται αποδεκτοί. Ωστόσο, ο συγχρονισμός των motes είναι σχετικά δαπανηρή διαδικασία, καθώς απαιτεί τη δράση όλων των κόμβων για να επιτευχθεί.

Μία άλλη μέθοδος, η οποία προσπαθεί να επιλύσει το ζήτημα των τοπικών περιορισμών με περισσότερο αποτελεσματικό τρόπο είναι το λεγόμενο balancing. Η φιλοσοφία του βασίζεται στον καθορισμό ενός συνόλου από motes, το balancing γκρουπ, που θα περιλαμβάνει τα motes των οποίων οι περιορισμοί έχουν παραβιαστεί, ενώ ο μέσος όρος των drift vectors των κόμβων του γκρουπ θα δημιουργεί μία μονοχρωματική drift σφαίρα. Αν το balancing γκρουπ έχει κατασκευαστεί επιτυχώς, τότε όλα τα motes στο γκρουπ θέτουν το drift vector τους ίσο με το μέσο όρο των drift vectors των κόμβων του γκρουπ μεταβάλλοντας τα slack vectors τους. Όπως έχουμε ήδη αναφέρει, το άθροισμα των slack vectors θα παραμείνει μηδέν, και ως εκ τούτου, το καθολικό measurement vector θα βρίσκεται μέσα στο convex hull των drift vectors. Το πλεονέκτημα του balancing σε σχέση με τον συγχρονισμό είναι ότι δεν απαιτεί τη δράση όλων των κόμβων παρά μόνο εκείνων που αποτελούν το balancing γκρουπ. Αν ένας περιορισμός που έχει

παραβιαστεί δεν επιλυθεί με τη διαδικασία του balancing, τότε θα γίνει συγχρονισμός στους κόμβους.

Η διαδικασία του balancing λειτουργεί ως εξής: Το mote, του οποίου ο περιορισμός έχει παραβιαστεί, στέλνει στον αρχηγό του cluster το drift vector του. Στη συνέχεια, το mote αυτό προστίθεται στο balancing γκρουπ, και ονομάζεται unbalanced mote. Ο αρχηγός του cluster με τη σειρά του, κατασκευάζει το γκρουπ προσθέτοντας επαναληπτικά νέα motes σε αυτό. Σε κάθε επανάληψη, επιλέγει με τυχαίο τρόπο έναν αριθμό από motes που δεν περιέχονται στο γκρουπ και ζητάει από αυτά να τους στείλουν το drift vector τους. Στην i -οστή επανάληψη, ο αρχηγός του cluster θα έχει επιλέξει 2^{i-1} νέα motes για να συμπεριλάβει στο γκρουπ. Ο μέσος όρος των drift vectors των κόμβων που συνθέτουν το balancing γκρουπ είναι γνωστό ως balanced vector.

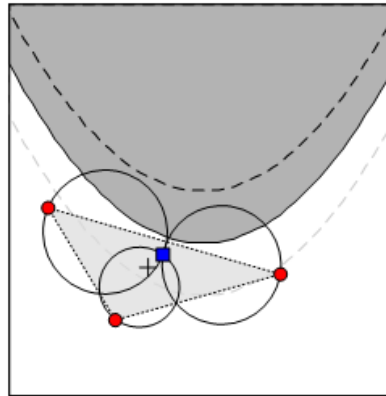
Από τη στιγμή που ο αρχηγός του cluster επικοινωνεί με τα motes χρησιμοποιώντας broadcast μηνύματα, σε κάθε επανάληψη θα ζητάει τα drift vectors από τα επιλεγμένα motes στέλνοντας μόνο ένα broadcast μήνυμα. Με τον διπλασιασμό των αριθμών των motes που προστίθενται στο balancing γκρουπ σε κάθε επανάληψη, ο αριθμός των μηνυμάτων που στέλνει ο αρχηγός του cluster κατά το balancing μεταβάλλεται λογαριθμικά σε σχέση με το μέγεθος του cluster.

Μετά από κάθε επανάληψη, ο αρχηγός του cluster ελέγχει αν το balanced vector δημιουργεί μονοχρωματική σφαίρα. Στην περίπτωση που δημιουργεί, θα έχουμε επιτυχές balancing. Αν το balancing γκρουπ περιέχει όλους τους κόμβους του δικτύου και το balanced vector δεν δημιουργεί μονοχρωματική σφαίρα, τότε το balancing θα είναι ανεπιτυχές. Στην περίπτωση πετυχημένου balancing, ο αρχηγός του cluster στέλνει το balanced vector στα motes του balancing γκρουπ. Στη συνέχεια, κάθε mote στο δίκτυο μεταβάλλει το slack vector του, ώστε το drift vector του να γίνει ίσο με το balanced vector.

4.3 Χαλάρωση απαιτήσεων

Στην Εικόνα 7 που ακολουθεί, παρατηρούμε ότι η τιμή μιας συνάρτησης είναι πολύ κοντά στο προκαθορισμένο όριο που ορίζει η γκρι περιοχή. Επιπλέον, οι drift σφαίρες που σχηματίζουν δύο από τα τρία motes είναι διχρωματικές. Ωστόσο, το convex hull, που εμπεριέχει το estimate vector (μπλε τετράγωνο) και το καθολικό measurement vector (σταυρός), είναι μονοχρωματικό. Για να αποφύγουμε τυχόν επικοινωνία των κόμβων από το γεγονός ότι ναι μεν η τιμή της συνάρτησης δεν παραβιάζεται, αλλά οι drift σφαίρες

είναι διχρωματικές, εισάγουμε περιθώρια λάθους (διακεκομμένες γραμμές) για τα οποία μιλούμε παρακάτω.



Εικόνα 7

Έστω \vec{v} το average measurement vector, f η συνάρτηση συνάθροισης, r ένα προκαθορισμένο κατώφλι και ε ένα προκαθορισμένο περιθώριο λάθους. Απαιτούμε πως αν $f(\vec{v}) < r + \varepsilon$, τότε κάθε mote θα μπορεί να καθορίσει ότι $f(\vec{v}) > r$ και αν $f(\vec{v}) \leq r - \varepsilon$, θα μπορεί να καθορίσει ότι $f(\vec{v}) \leq r$. Μία συναθροιστική επερώτηση που συμφωνεί με αυτές τις απαιτήσεις, λέμε ότι υποστηρίζει ένα περιθώριο λάθους ε .

Ο αλγόριθμος της γεωμετρικής προσέγγισης μπορεί εύκολα να προσαρμοστεί ώστε να υποστηρίζει περιθώριο λάθους. Ορίζουμε δύο χρωματισμούς, ο ένας εξ αυτών εισάγεται από τη συναθροιστική συνάρτηση f και την οριακή τιμή $r + \varepsilon$, ενώ ο δεύτερος από την f και την οριακή τιμή $r - \varepsilon$. Όταν ο αρχικός αλγόριθμος ελέγχει αν μία σφαίρα είναι μονοχρωματική, τότε αν $f(\vec{v}) \leq r$, ο τροποποιημένος αλγόριθμος θα ελέγξει αν η σφαίρα είναι μονοχρωματική ως προς τον πρώτο χρωματισμό (αυτόν που ορίζεται από την f και την τιμή $r + \varepsilon$). Αν $f(\vec{v}) > r$, ο τροποποιημένος αλγόριθμος θα ελέγξει αν η σφαίρα είναι μονοχρωματική ως προς τον δεύτερο χρωματισμό. Αυτό βεβαιώνει πως αν η τιμή της συνάρτησης πάνω στο estimate vector είναι κάτω από το όριο, ενώ η τιμή της πάνω στο average measurement vector είναι πάνω από το $r + \varepsilon$, τότε η επερώτηση θα υπολογιστεί ξανά. Επιπλέον, ο τροποποιημένος αλγόριθμος βεβαιώνει πως ανεξάρτητα από την τιμή του estimate vector, τα drift vectors έχουν ελάχιστα περιθώρια ελιγμών διατηρώντας έτσι τις σφαίρες τους μονοχρωματικές.

5. Εργασία

Στην παρούσα διπλωματική εργασία υποθέσαμε μια ιεραρχική οργάνωση των αισθητήρων σε ένα δέντρο συλλογής δεδομένων, κατασκευασμένο με βάση την απόσταση στην οποία οι αισθητήρες μπορούν να μεταδώσουν. Θεωρήσαμε ότι ένας κόμβος μπορεί να μεταδώσει όχι μόνο σε έναν πατέρα αλλά σε όσους κόμβους βρίσκονται ένα επίπεδο πιο κοντά στη ρίζα και μπορούν να ακούσουν το μήνυμά του. Χρησιμοποιήσαμε συναρτήσεις ελέγχου οι οποίες είναι ορισμένες πάνω στο μέσο όσο των διανυσμάτων μέτρησης που συλλέγουν όλοι οι κόμβοι του δικτύου, ενώ για να καθορίσουμε το πότε θα μεταδώσουν οι κόμβοι εφαρμόσαμε τη γεωμετρική προσέγγιση.

Για τη δημιουργία του δέντρου συλλογής δεδομένων χρησιμοποιήσαμε τον αλγόριθμο του TAG.

5.1 Συνάρτηση συναθροίσης

Όπως έχουμε αναφέρει παραπάνω στο κείμενο, θέλουμε για οποιαδήποτε χρονική στιγμή να καθορίσουμε εάν η τιμή μιας συναθροιστικής συνάρτησης πάνω στο μέσο όρο των measurement vectors που κρατούν τα motes του δικτύου, ξεπερνά ένα προκαθορισμένο κατώφλι. Εξετάζουμε συναρτήσεις που μπορούν να εκφραστούν βάσει του μέσου όρου των μετρήσεων. Για παράδειγμα, έστω αισθητήρες σε ένα δωμάτιο που μετρούν τη θερμοκρασία και X μία τυχαία μεταβλητή που αντιπροσωπεύει ένα σύνολο αποτελούμενο από τις μετρήσεις θερμοκρασίας που έχουν λάβει τα motes. Αν ο κόμβος $s_{i,j}$ παίρνει τη μέτρηση $x_{i,j}$, τότε για να παρακολουθήσουμε αν η διακύμανση της μέτρησης βρίσκεται πάνω/κάτω από ένα όριο απαιτείται το measurement vector που θα

σηματίζει θα έχει τη μορφή $\vec{v}_{i,j} = \begin{pmatrix} x_{i,j} \\ x_{i,j}^2 \end{pmatrix}$.

Μπορούμε εύκολα να παρατηρήσουμε πως $\vec{v} = \frac{1}{N_{tot}} \sum_{i=1}^n \sum_{j=1}^{N_i} \begin{pmatrix} x_{i,j} \\ x_{i,j}^2 \end{pmatrix} = \begin{pmatrix} E[X] \\ E[X^2] \end{pmatrix}$.

Επομένως, η διακύμανση της θερμοκρασίας μπορεί να υπολογιστεί από την ακόλουθη συναθροιστική συνάρτηση:

$$f(\vec{v}) = v^{(2)} - (v^{(1)})^2 = E[X^2] - E[X]^2 = Var(X)$$

Δηλαδή, αν θέλουμε να μελετήσουμε τη διακύμανση της θερμοκρασίας σε ένα χώρο, αρκεί οι αισθητήρες να μετρούν τη θερμοκρασία και κατά το σχηματισμό του measurement vector του καθενός, να δεσμεύεται μία θέση παραπάνω στο vector ώστε να αποθηκεύουν το τετράγωνο αυτής για να μπορέσουμε στη συνέχεια να υπολογίσουμε τη διακύμανση.

5.2 Εφαρμογή γεωμετρικής προσέγγισης στην αποστολή δεδομένων μεταξύ των κόμβων του δέντρου

Σε κάθε εποχή, οι ενεργοί κόμβοι παίρνουν μία μέτρηση από το περιβάλλον. Ανάλογα με τη συνάρτηση που εξετάζουμε, κάθε κόμβος σχηματίζει το measurement vector όπως περιγράψαμε παραπάνω.

Τα φύλλα του δέντρου

- κρατούν μόνο την τελευταία επιτυχώς αναφερόμενη μέτρηση, που είχε σταλεί από έναν αισθητήρα στο γονέα της.
- Όταν μια νέα ανάγνωση \vec{v}_{new} είναι διαθέσιμη, τότε συγκρίνεται με την τελευταία αναφερόμενη ανάγνωση \vec{v}_{old} .
- Με αυτόν τον τρόπο, σχηματίζουν το διάνυσμα $\overrightarrow{DV} = \vec{v}_{new} - \vec{v}_{old}$

Οι ενδιάμεσοι κόμβοι του δέντρου

- Λαμβάνουν μερικά αποτελέσματα από τα παιδιά τους.
- Κρατούν την τελευταία αναφερόμενη μέτρηση για τον εαυτό τους, καθώς και τα τελευταία αναφερόμενα στοιχεία που έλαβαν από κάθε παιδί τους. Τα στοιχεία αυτά μπορούν να είναι είτε απλές μετρήσεις από κόμβους φύλλων είτε μερικά αποτελέσματα που αναφέρονται από έναν εσωτερικό κόμβο. Αν πρόκειται για παθητικούς κόμβους, τότε κρατούν μόνο τα τελευταία αναφερόμενα στοιχεία που έλαβαν από τα παιδιά τους.
- Με αυτόν τον τρόπο σχηματίζουν το διάνυσμα $\overrightarrow{DV} = \overrightarrow{receivedFromChildren} + \vec{v}_{new} - \vec{v}_{old}$

Όταν ένας κόμβος σχηματίσει το διάνυσμα \overrightarrow{DV} , τότε θα σχηματίσει το drift vector του ως το διάνυσμα $\vec{u} = \overrightarrow{estimateVector} + \overrightarrow{DV}$. Το estimate vector υπολογίζεται σε κάθε εποχή

από τη ρίζα του δέντρου. Την πρώτη εποχή που τρέχει ο αλγόριθμος, θα είναι μηδενικό, ενώ την εποχή που η ρίζα θα ανιχνεύσει κάποια παραβίαση θα υπολογίζεται ως:

$$\bar{e} = \bar{e} + \frac{\overline{DV}}{N_{total}}$$

λειτουργία της ρίζας καθώς και για το πότε ένας κόμβος θα στείλει δεδομένα στο γονέα του, θα μιλήσουμε παρακάτω.

Μετά το σχηματισμό και του drift vector, ο κόμβος θα φτιάξει τη σφαίρα του με κέντρο στο $\bar{e} + \frac{\overline{DV}}{2}$ και ακτίνα $\frac{\|\overline{DV}\|}{2}$. Στη συνέχεια ο κόμβος θα ελέγξει αν η σφαίρα του είναι μονοχρωματική ή όχι, δηλαδή αν για ένα προκαθορισμένο κατώφλι υπάρχει παραβίαση εντός της σφαίρας ελέγχου.

5.3 Ανίχνευση παραβίασης

Ο κόμβος ανιχνεύει παραβίαση, όταν η σφαίρα που σχηματίζει δεν είναι μονοχρωματική. Για να ελέγξουμε αν μία σφαίρα είναι μονοχρωματική, δημιουργούμε εντός αυτής ένα πλέγμα (grid). Σε κάθε σημείο $p(x,y)$ του πλέγματος εντός της σφαίρας εφαρμόζουμε τη συνάρτηση που μελετάμε. Για παράδειγμα αν η συνάρτηση που εξετάζουμε είναι η διακύμανση της θερμοκρασίας, για κάθε σημείο $p(x,y)$ θα υπολογίσουμε την τιμή $f(p)=x-y^2$. Κατόπιν βρίσκουμε τη μέγιστη και την ελάχιστη τιμή της συνάρτησης που μόλις υπολογίσαμε και ελέγχουμε αν αυτές παραβιάζουν κάποιο προκαθορισμένο κατώφλι R . Παραβίαση θα έχουμε όταν η μέγιστη τιμή $\max Value \geq R$ και η ελάχιστη τιμή $\min Value \leq R$. Σε αυτήν την περίπτωση, η σφαίρα δεν θα είναι μονοχρωματική και ο κόμβος θα χρειαστεί να στείλει το \overline{DV} του στον πατέρα του.

5.4 Η προς μετάδοση τιμή

Όταν ο κόμβος διαπιστώσει ότι η σφαίρα του δεν είναι μονοχρωματική, θα πρέπει να μεταδώσει τα δεδομένα του, δηλαδή το \overline{DV} του, στον πατέρα του.

Στην περίπτωση που έχουμε ορίσει στη υλοποίησή μας ότι ένας κόμβος θα έχει έναν μοναδικό πατέρα, τότε θα στείλει σε αυτόν τον μοναδικό του πατέρα όλο το \overline{DV} του.

Αν έχουμε ορίσει ότι ένας κόμβος μπορεί να έχει περισσότερους του ενός πατεράδες, τότε ο κόμβος σπάσει το \overline{DV} του σε τόσα μέρη όσος είναι ο αριθμός των πατεράδων του,

και θα κάνει broadcast σε αυτούς την «σπασμένη» τιμή. Σε κάθε περίπτωση, ο κόμβος που στέλνει κάνει το \overline{DV} του μηδέν.

5.5 Λειτουργία ρίζας δέντρου και συγχρονισμός

Σε περίπτωση που η ρίζα ανιχνεύσει παραβίαση, δηλαδή όταν η σφαίρα της δεν είναι μονοχρωματική, τότε ανάλογα με το είδος του συγχρονισμού που θέλουμε να λάβει χώρα, θα αναγκάσει τους κόμβους σε πλήρη ή μερικό συγχρονισμό.

5.5.1 Πλήρης συγχρονισμός

Κατά τον πλήρη συγχρονισμό, η ρίζα με broadcast μήνυμα ζητά από τους κόμβους που δεν είχαν στείλει το \overline{DV} τους στον πρώτο γύρο της συγκεκριμένης εποχής στον πατέρα ή στους πατεράδες τους, να το στείλουν υποχρεωτικά. Η τιμή αυτή δεν στέλνεται απευθείας στην ρίζα αλλά ακολουθεί τη δομή του δέντρου, δηλαδή κάθε κόμβος που δεν είχε στείλει προηγουμένως, θα στείλει στον πατέρα του κ.ο.κ. Ένας κόμβος που είχε στείλει το \overline{DV} του στον πρώτο γύρο, αλλά λαμβάνει δεδομένα από παιδιά του λόγω του συγχρονισμού, θα προωθήσει επίσης τα δεδομένα αυτά προς τα πάνω στο δέντρο. Θα πρέπει να προσέξουμε πως αν ένας κόμβος έχει πολλαπλούς πατεράδες, σε αυτό το στάδιο του συγχρονισμού, θα στείλει όλο το \overline{DV} του σε έναν μόνο πατέρα του και συγκεκριμένα σε εκείνον που θα είχε για πατέρα αν ο αλγόριθμος έτρεχε εξ αρχής με την υπόθεση ότι κάθε κόμβος θα έχει έναν μόνο γονέα. Κάθε κόμβος που θα στείλει, θα θέσει στη συνέχεια το \overline{DV} του ίσο με μηδέν.

Όταν τα δεδομένα φτάσουν ξανά στη ρίζα, τότε η ρίζα θα υπολογίσει το estimate vector

ως το $\vec{e} = \vec{e} + \frac{\overline{DV}_{total}}{N_{act}}$, όπου N_{act} ο αριθμός των κόμβων του δικτύου και \overline{DV}_{total} το

συνολικό \overline{DV} που υπολογίζει η ρίζα και για τους δύο γύρους. Τέλος, η ρίζα θα κάνει broadcast το estimate vector ώστε να το λάβουν όλοι οι κόμβοι του δικτύου για να το χρησιμοποιήσουν μελλοντικά στην κατασκευή νέας σφαίρας.

5.5.2 Μερικός συγχρονισμός

Κατά τον μερικό συγχρονισμό, η ρίζα υπολογίζει αρχικά το estimate vector $\vec{e} = \vec{e} + \frac{\overline{D\vec{V}}}{N_{act}}$

και κατόπιν το κάνει broadcast για να το λάβουν οι κόμβοι που κατά τον πρώτο γύρο της συγκεκριμένης εποχής δεν είχαν ανιχνεύσει παραβίαση και επομένως δεν είχαν μεταδώσει τις τιμές τους. Στη συνέχεια, κάθε κόμβος που δεν είχε στείλει, υπολογίζει νέο drift vector και κατ' επέκταση σχηματίζει νέα σφαίρα λαμβάνοντας υπόψιν το νέο estimate vector που μόλις έλαβε. Έπειτα, ελέγχει αν η σφαίρα του είναι μονοχρωματική σύμφωνα με τον τρόπο που περιγράψαμε στο 5.3 και στην περίπτωση που ανιχνεύσει παραβίαση προωθεί τα δεδομένα του στον μοναδικό πατέρα του. Ακόμα και αν έχουμε ορίσει ότι ένας κόμβος θα έχει περισσότερους του ενός πατεράδες, σε αυτό το στάδιο του συγχρονισμού, ο κόμβος θα στείλει τα δεδομένα του σε έναν εξ' αυτών. Όταν τα δεδομένα φτάσουν ξανά στη ρίζα, η ρίζα θα ελέγξει πάλι αν η νέα σφαίρα της είναι μονοχρωματική, και στην περίπτωση που ανιχνεύσει ξανά παραβίαση, τότε θα αναγκάσει σε πλήρη συγχρονισμό τους κόμβους που δεν έστειλαν ούτε στο δεύτερο γύρο, όπως ακριβώς περιγράφεται στο 5.5.1.

6. Πειράματα

Στα πειράματά μας χρησιμοποιήσαμε τα δεδομένα που συλλέχθηκαν από Mica2Dot κόμβους αισθητήρων με weather boards για θερμοκρασία και υγρασία στο Intel Berkley Research Lab [29]. Τα δεδομένα συλλέχθηκαν με χρήση του TinyDB σε εντός δικτύου σύστημα επεξεργασίας ερωτήματος βασιζόμενο στην πλατφόρμα TinyOS. Χρησιμοποιήσαμε αυτές τις μετρήσεις για να παράγουμε αναγνώσεις 100 κόμβων για μία χρονική περίοδο 100 εποχών.

Μελετήσαμε τον αριθμό μηνυμάτων που στέλνονται συνολικά στο δίκτυο για τις συναρτήσεις avg και variance κατά την εφαρμογή της γεωμετρικής προσέγγισης για διάφορες αποστάσεις μετάδοσης των κόμβων και για κάθε μία από τις ακόλουθες περιπτώσεις:

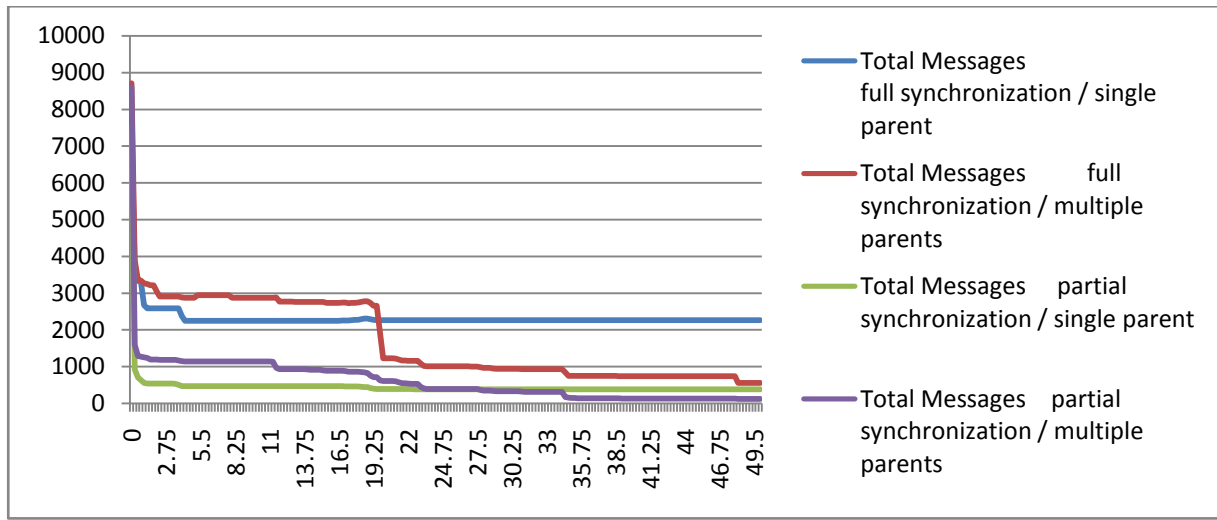
- Πλήρης συγχρονισμός και μοναδικός πατέρας για κάθε κόμβο.
- Πλήρης συγχρονισμός και κάθε κόμβος δύναται να έχει περισσότερους του ενός πατεράδες.
- Μερικός συγχρονισμός και μοναδικός πατέρας για κάθε κόμβο.
- Μερικός συγχρονισμός και κάθε κόμβος δύναται να έχει περισσότερους του ενός πατεράδες.

6.1 Συνάρτηση average

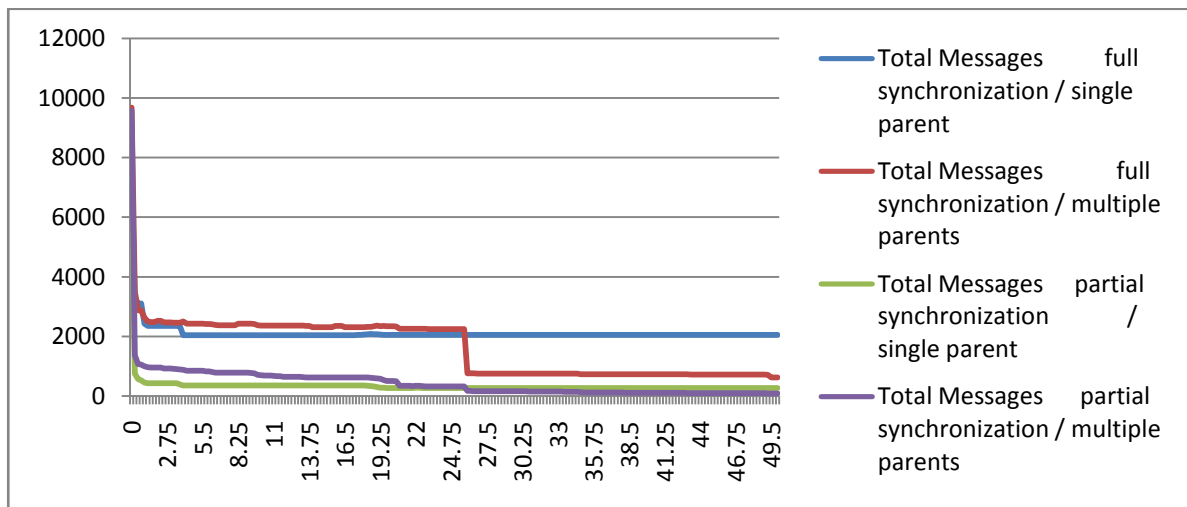
Η πρώτη συνάρτηση που μελετήσαμε ήταν το average (avg). Κατά την εφαρμογή αυτής της συνάρτησης, κάθε κόμβος ελέγχει αν το DV του είναι μεγαλύτερο από το προκαθορισμένο threshold για να προωθήσει τα δεδομένα του προς τα πάνω στο δέντρο. Η ρίζα με τη σειρά της, αν ανιχνεύσει παραβίαση, αναγκάζει τους κόμβους σε πλήρη ή σε μερικό συγχρονισμό, όπως περιγράφηκε στο 5.5.

Μελετήσαμε τον συνολικό αριθμό των μηνυμάτων που στέλνονται στο δίκτυο σε διάστημα 100 εποχών και για διαφορετικά thresholds κάθε φορά κατά την εφαρμογή της συγκεκριμένης συνάρτησης για τα μεγέθη: θερμοκρασία και υγρασία. Επιπλέον, παρατηρήσαμε πώς μεταβάλλεται ο αριθμός αυτός των μηνυμάτων για διαφορετικά transmit distance των κόμβων.

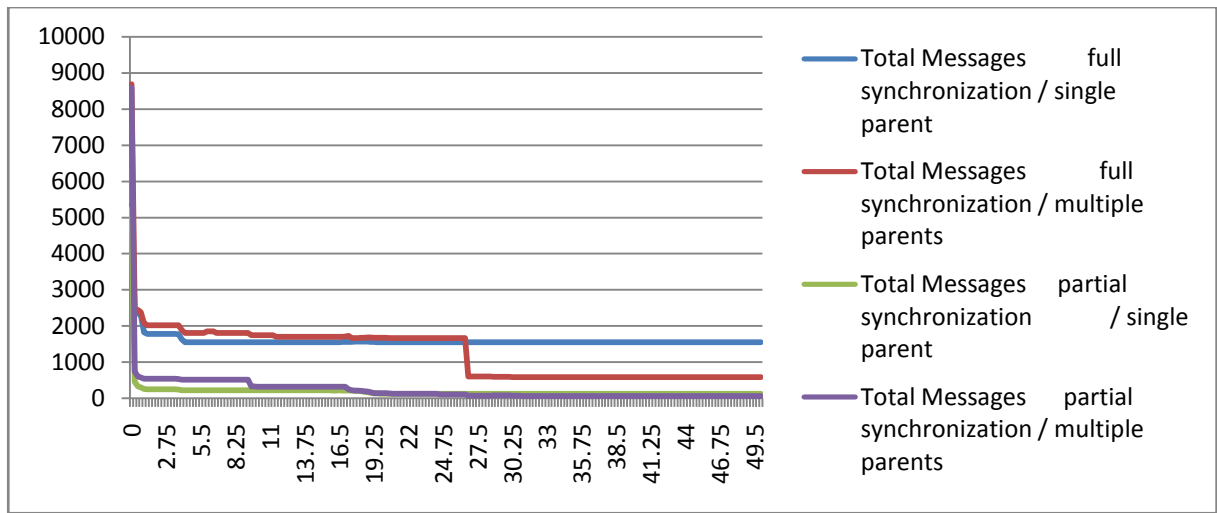
Θερμοκρασία



Εικόνα 8: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για $transmit\ distance\ \sqrt{2}$.

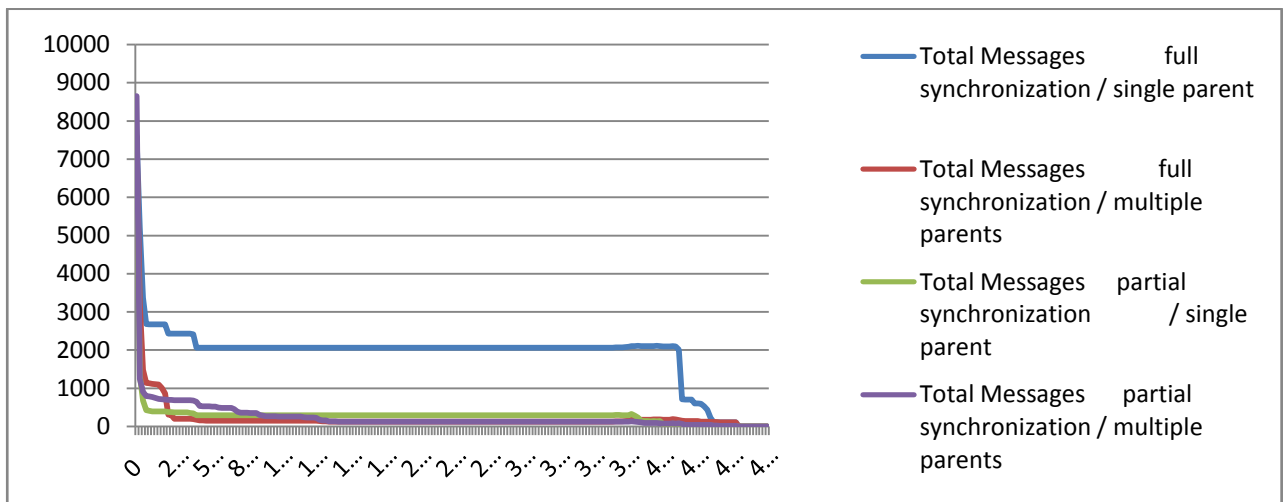


Εικόνα 9: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για $transmit\ distance\ 2$.

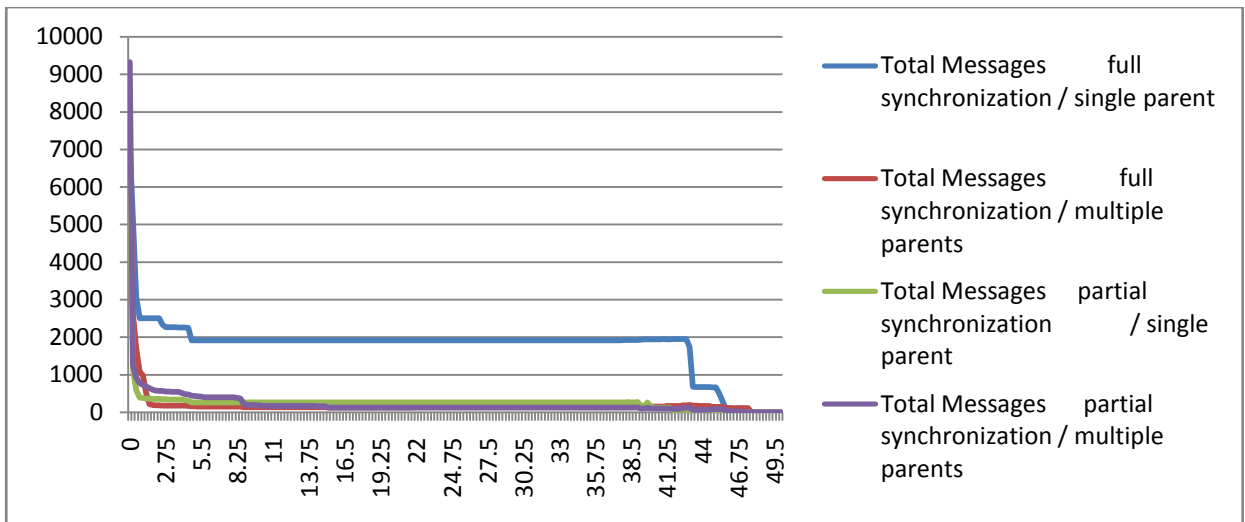


Εικόνα 10: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για transmit distance 5.

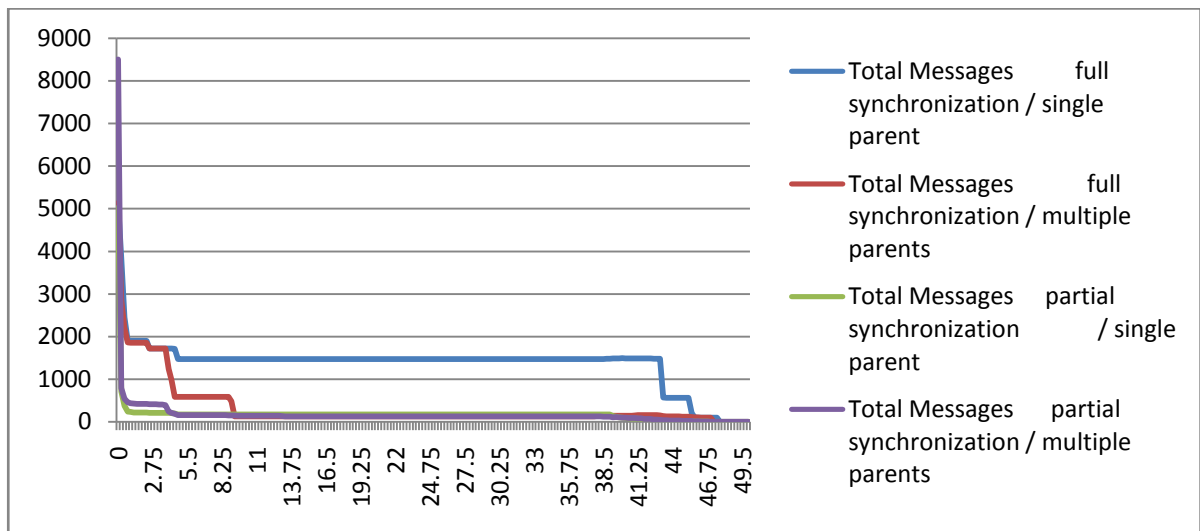
Υγρασία



Εικόνα 11: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για transmit distance $\sqrt{2}$.



Εικόνα 12: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για transmit distance 2.



Εικόνα 13: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για transmit distance 5.

Στο σημείο αυτό θα επισημάνουμε ότι η αύξηση του transmit distance των κόμβων έχει ως αποτέλεσμα ενδεχόμενη αύξηση του αριθμού των πιθανών πατεράδων ενός κόμβου, αλλά και μείωση του αριθμού των επιπέδων του δέντρου. Στην περίπτωση που το transmit distance είναι $\sqrt{2}$ ή 2, το δέντρο αποτελείται από 10 επίπεδα, ενώ για transmit distance 5 αποτελείται από 4 επίπεδα.

Στα παραπάνω σχήματα παρατηρούμε πως τόσο για τον πλήρη όσο και για τον μερικό συγχρονισμό, η μέθοδος των πολλαπλών πατεράδων έχει ως αποτέλεσμα την αποστολή τελικά μικρότερου αριθμού μηνυμάτων απ' ότι στην περίπτωση που κάθε κόμβος θα έχει έναν μοναδικό πατέρα. Κατά τη μέτρηση της θερμοκρασίας, φαίνεται ότι για μικρότερα thresholds η μέθοδος των πολλαπλών πατεράδων δίνει αυξημένο αριθμό μηνυμάτων. Αυτό οφείλεται στο γεγονός ότι οι μετρήσεις θερμοκρασίας που παίρνουν οι κόμβοι σε κάθε εποχή, ενδέχεται να διαφέρουν κατά πολύ από εποχή σε εποχή, με αποτέλεσμα οι κόμβοι να υπολογίζουν μεγάλο DV, και κατ' επέκταση να ανιχνεύουν παραβίαση. Έτσι, η τελική τιμή που θα υπολογίζει ένας πιθανός πατέρας συναθροίζοντας τα δεδομένα των πιθανών παιδιών του, θα είναι μεγάλη, με αποτέλεσμα να ανιχνεύει και εκείνος παραβίαση.

Επιπλέον, παρατηρούμε πως ο μερικός συγχρονισμός δίνει καλύτερα αποτελέσματα από τον πλήρη λόγω του ότι με τον μερικό συγχρονισμό δίνεται η ευκαιρία στους κόμβους να απορροφήσουν τυχόν αρχική παραβίαση που είχαν ανιχνεύσει σε πρώτο στάδιο χωρίς να αναγκάζονται να στείλουν εν τέλει όλοι οι κόμβοι του δικτύου, όπως συμβαίνει με τον πλήρη συγχρονισμό. Ακόμη, θα σημειώσουμε ότι η αύξηση του transmit distance των κόμβων έχει ως αποτέλεσμα τη μείωση του αριθμού των μηνυμάτων για κάθε μέθοδο σε σχέση με μικρότερο transmit distance.

6.2 Συνάρτηση variance

Για να υπολογίσουμε τη διακύμανση στις μετρήσεις ενός μεγέθους (π.χ. διακύμανση θερμοκρασίας, υγρασίας κτλ) ακολουθήσαμε το παράδειγμα που περιγράφηκε στο 5.1. Σε κάθε εποχή, κάθε κόμβος λαμβάνει μία μέτρηση δεσμεύοντας παράλληλα μία θέση μνήμης επιπλέον για να αποθηκεύσει το τετράγωνο αυτής. Όπως αναφέρθηκε στο 5.3, κατά το σχηματισμό της σφαίρας ελέγχου κάθε κόμβου, για να διαπιστώσουμε εάν ο κόμβος ανιχνεύει παραβίαση, δημιουργούμε ένα πλέγμα εντός αυτής. Σε κάθε σημείο(x,y) εντός του πλέγματος εφαρμόζουμε τη συνάρτηση variance μέσω της σχέσης:

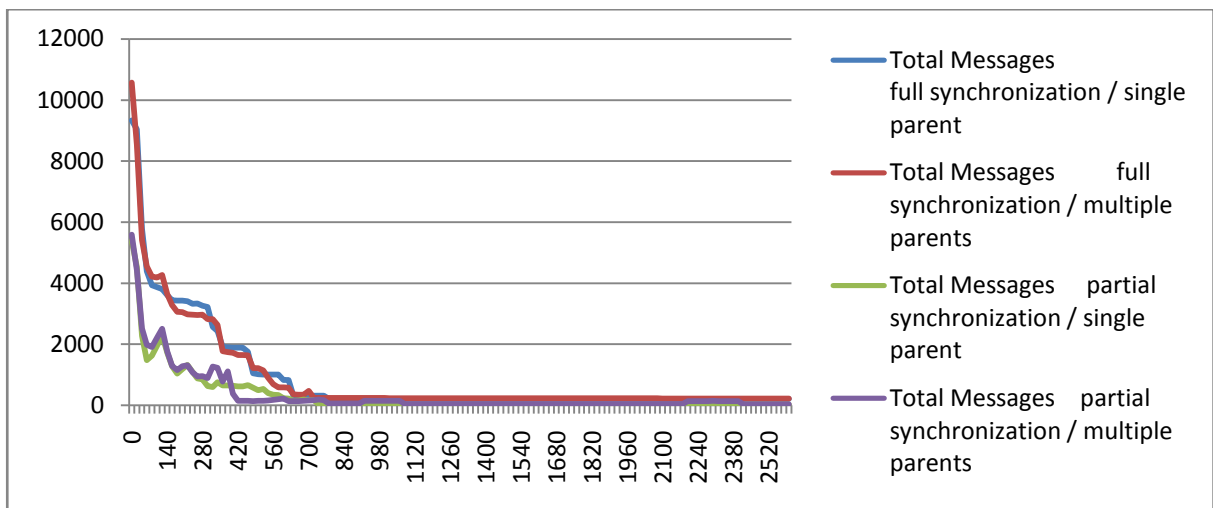
$$f(\vec{v}) = v^{(2)} - (v^{(1)})^2 = E[X^2] - E[X]^2 = Var(X)$$

Κατόπιν βρίσκουμε τη μέγιστη και την ελάχιστη τιμή της συνάρτησης που μόλις υπολογίσαμε και ελέγχουμε αν αυτές παραβιάζουν κάποιο προκαθορισμένο κατώφλι R. Παραβίαση θα έχουμε όταν η μέγιστη τιμή $\max Value \geq R$ και η ελάχιστη τιμή $\min Value$

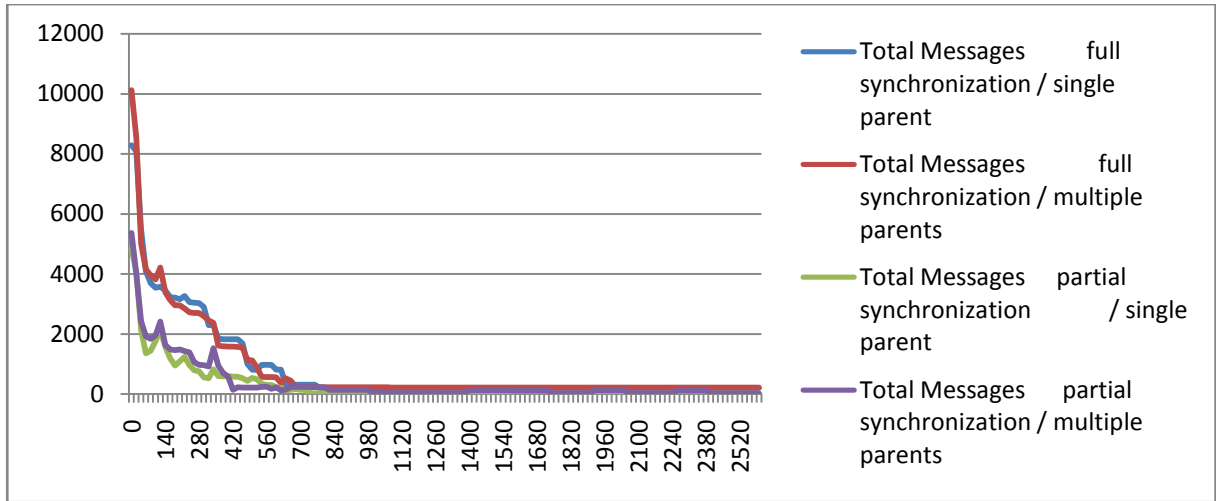
$\leq R$. Σε αυτήν την περίπτωση, η σφαίρα δεν θα είναι μονοχρωματική και ο κόμβος θα χρειαστεί να στείλει το \overline{DV} του στον πατέρα του.

Μελετήσαμε τον συνολικό αριθμό των μηνυμάτων που στέλνονται στο δίκτυο σε διάστημα 100 εποχών και για διαφορετικά thresholds κάθε φορά κατά την εφαρμογή της συγκεκριμένης συνάρτησης για τα μεγέθη: θερμοκρασία και υγρασία. Επιπλέον, παρατηρήσαμε πώς μεταβάλλεται ο αριθμός αυτός των μηνυμάτων για διαφορετικά transmit distance των κόμβων.

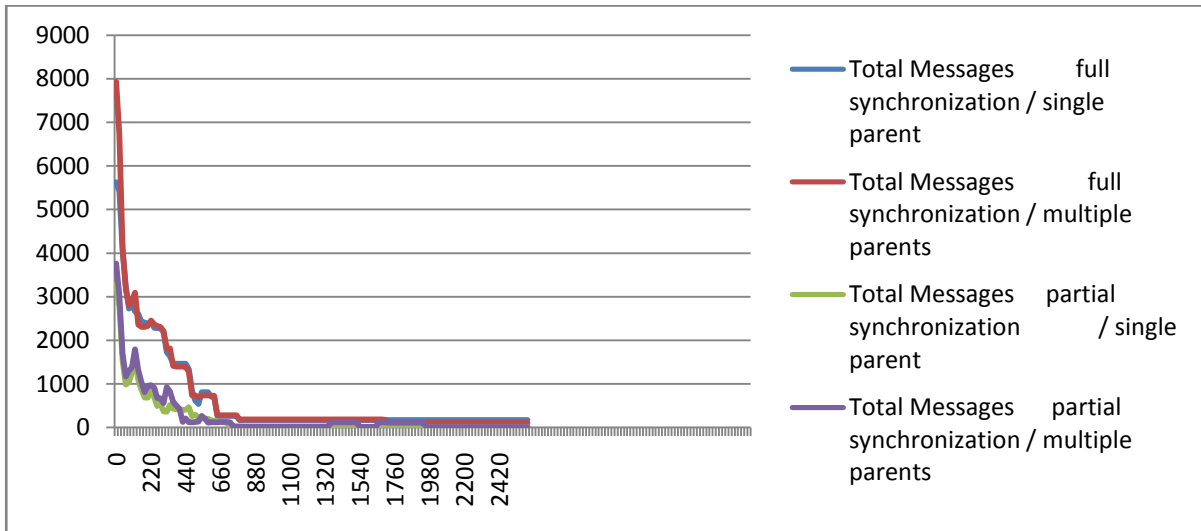
Θερμοκρασία



Εικόνα 14: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για transmit distance $\sqrt{2}$

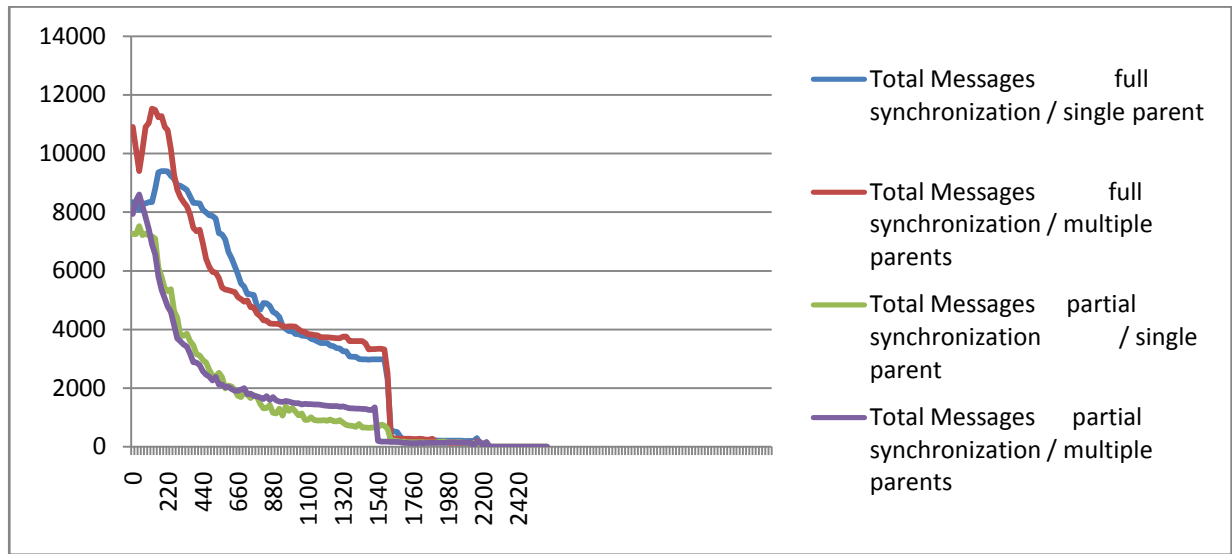


Εικόνα 15: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για *transmit distance* 2.

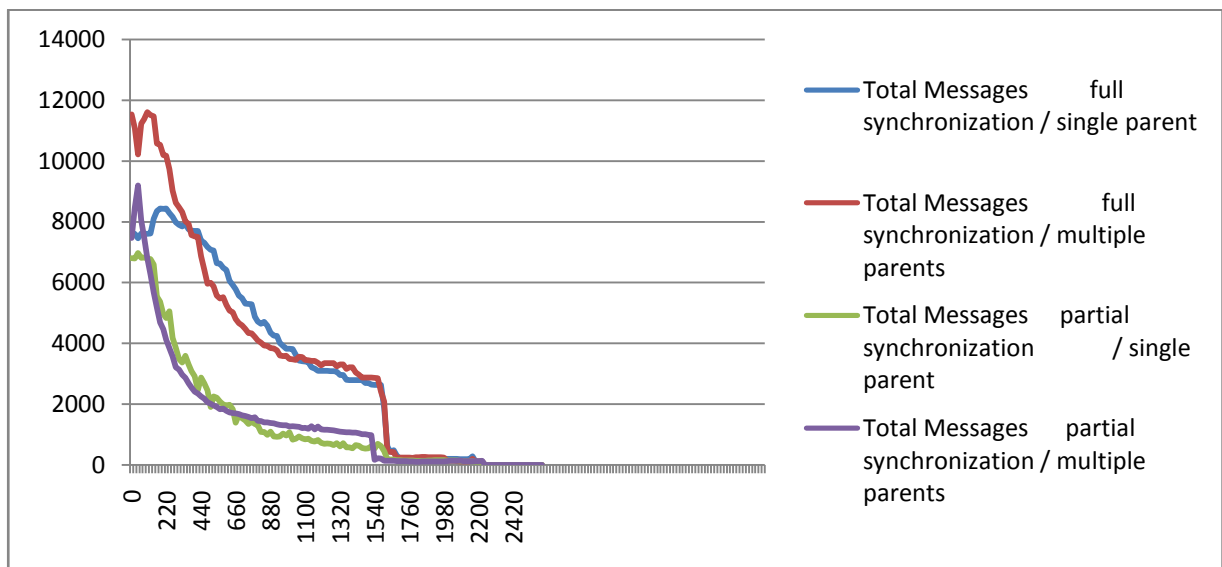


Εικόνα 16: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για *transmit distance* 5.

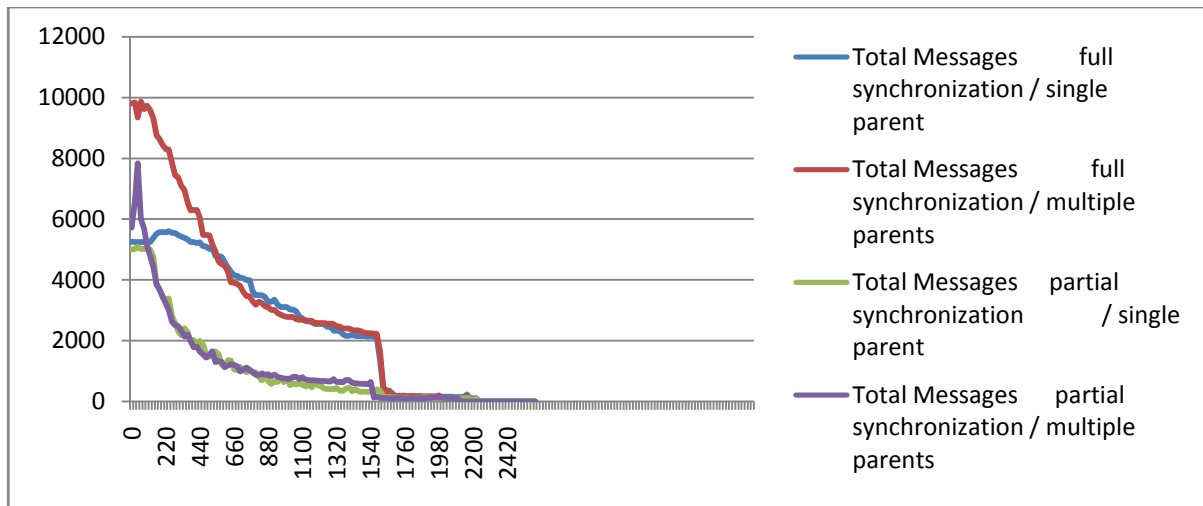
Υγρασία



Εικόνα 17: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για $transmit\ distance\ \sqrt{2}$



Εικόνα 18: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για $transmit\ distance\ 2$.



Εικόνα 19: Αριθμός μηνυμάτων που στέλνονται συνολικά στο δίκτυο για transmit distance 5.

Και για τη συνάρτηση variance, όπως και για την average, παρατηρούμε ότι η αποστολή δεδομένων σε περισσότερους του ενός πατεράδες έχει ως αποτέλεσμα την αποστολή μικρότερου αριθμού μηνυμάτων συνολικά στο δίκτυο. Είναι επίσης προφανές για ακόμη μία φορά ότι ο μερικός συγχρονισμός δίνει γενικότερα καλύτερα αποτελέσματα από τον πλήρη συγχρονισμό. Επιπλέον, παρατηρούμε ότι η αύξηση του transmit distance των κόμβων μειώνει τον αριθμό των μηνυμάτων που στέλνονται συνολικά σε σχέση με αυτό που θα είχαμε για μικρότερο transmit distance. Η αύξηση των μηνυμάτων που ενδεχομένως να παρατηρείται για μικρά thresholds στην περίπτωση αποστολής δεδομένων σε περισσότερους του ενός πατεράδων, οφείλεται στο γεγονός ότι η αύξηση του transmit distance ενός κόμβου έχει ως αποτέλεσμα περισσότεροι κόμβοι να μπορούν να ακούν το μήνυμά του, επομένως κατά το σχηματισμό του δέντρου, ο κόμβος θα έχει περισσότερους πιθανούς πατεράδες. Κατ' επέκταση, ένας πατέρας θα έχει περισσότερα πιθανά παιδιά, συνεπώς αν τα πολλά παιδιά του ανιχνεύσουν παραβίαση, εκείνος με τη σειρά του δεν θα μπορέσει να τις απορροφήσει λόγω του μικρού threshold.

7. Συμπεράσματα και μελλοντική εργασία

Οι περιορισμένες δυνατότητες των αισθητήρων σε αποθέματα ενέργειας καθώς και η ανάγκη για λήψη αξιόπιστων ειδοποιήσεων σε περιπτώσεις ανίχνευσης παραβιάσεων καθιστούν απαραίτητη την ανάπτυξη τεχνικών που θα εξασφαλίζουν στο δίκτυο μεγαλύτερη διάρκεια ζωής ενώ θα μπορούν παράλληλα να αναγνωρίσουν περιπτώσεις άμεσης ανάγκης και επέμβασης. Με την επιβολή περιορισμών (constraints) στους κόμβους και της παρακολούθησης αυτών μέσω της γεωμετρικής προσέγγισης, δώσαμε στους κόμβους τη δυνατότητα να μπορούν να ελέγχουν σε μία ευρύτερη περιοχή αν υπάρχει παραβίαση και να στέλνουν ειδοποίηση στην περίπτωση που τα constraints αυτά δεν ικανοποιούνται. Με τη δυνατότητα αποστολής μίας τιμής μέτρησης σε περισσότερους του ενός πατεράδες διαιρώντας τη σε ισάριθμα κομμάτια ανάλογα με τον αριθμό των πατεράδων κάθε κόμβου, καταφέραμε να σπάσουμε μεγάλες τιμές σε μικρότερα τμήματα οδηγώντας εν τέλει σε σημαντική μείωση του αριθμού των μηνυμάτων που στάλθηκαν στο δίκτυο, καθώς οι ενδιάμεσοι κόμβοι είχαν να χειριστούν μικρότερες τιμές, επομένως ήταν πιο εύκολο για εκείνους συναθροίζοντας τα δεδομένα τους να δημιουργήσουν μία νέα τιμή που δεν θα παραβίαζε κάποιο προκαθορισμένο κατώφλι.

Η εργασία αυτή θα μπορούσε να επεκταθεί υπολογίζοντας για ένα διάστημα εποχών τη συνολική ενέργεια που καταναλώνει ένα δίκτυο που υποστηρίζει τη δυνατότητα αποστολής δεδομένων σε περισσότερους του ενός πατεράδες. Αναμένουμε αυτή η ενέργεια να είναι χαμηλότερη από αυτή που θα δαπανούσε το δίκτυο στην περίπτωση που κάθε κόμβος είχε ένα μοναδικό γονέα, όπως συνέβη άλλωστε και με τον αριθμό μηνυμάτων, καθώς ενέργεια και μηνύματα είναι άρρηκτα συνδεδεμένα.

Επιπλέον, η εργασία θα μπορούσε να εφαρμοστεί σε ένα δίκτυο αισθητήρων αποτελούμενο από περισσότερα του ενός clusters. Σε κάθε cluster ενός τέτοιου δικτύου, οι κόμβοι θα ήταν οργανωμένοι σε δέντρο συλλογής δεδομένων και θα ακολουθούσαν τη γεωμετρική προσέγγιση για να αποφασίσουν αν θα στείλουν τα δεδομένα τους. Και σε αυτή την περίπτωση δικτύου, αναμένουμε ο αριθμός των μηνυμάτων που θα σταλούν συνολικά στο δίκτυο να είναι ελαττωμένος.

Βιβλιογραφία

- [1] I.Sharfman, A. Schuste and Daniel Keren. Aggregate threshold queries in sensor networks. In IPDPS, 2007.
- [2] J.Considine, F. Li, G. Kollios and J. Byers. Approximate aggregation techniques for sensor databases. In ICDE '04, page 449. IEEE Computer Society.
- [3] R. Govindan, E. Kohler, D.Estrin, F. Brian, K.Chintalapudi, O. Gnawali, S. Rangwala, R. Gummadi and T. Stathopoulos. Tenet: An architecture for tierd embedded networks. In CENS Technical Report 56.
- [4] I.Gupta, R. van Renesse and K. P. Birman. Scalable fault – tolerant aggregation in large process groups. In DSN '01, PAGES 433-442. IEEE Computer Society.
- [5] B. Krishnamachari, D. Estrin, and S. B. Wicker. The impact of data aggregation in wireless sensor networks. In ICDCSW '02, pages 575-578. IEEE Computer Society.
- [6] R. Kumar, V. Tsiatsis, and M. B. Srivastava. Computation hierarchy for in-network processing. In WSNA '03, pages 68-77. ACM Press.
- [7] S. Madden, M.J Franklin, J. M . Hellerstein, and W. Hong. Tag: A tiny aggregation service for ad-hoc sensor networks. In OSDI '02.
- [8] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson. Wireless sensor networks for habitat monitoring. In WSNA '02, pages 88-97. ACM Press.
- [9] A. Manjhi, S. Nath, and P. B. Gibbons. Tributaries and deltas: efficient and robust aggregation in sensor network streams. In SIGMOD '05, pages 287-298. ACM Press.
- [10] S. Nath, P. B. Gibbons, S. Seshan, and Z. R. Anderson. Synopsis diffusion for robust aggregation in sensor networks. In SenSys '04, pages 250-262. ACM Press.
- [11] I. Sharfman, A.Schuster, and D.Keren. A geometric approach to monitoring threshold functions over distributed data streams. In SIGMOD '06, pages 301-312. ACM Press.
- [12] G. Sharma and R. Mazumdar. Hybrid sensor networks: a small world. In MobiHoc '05, pages 366-377. ACM Press.
- [13] N. Shrivastava, C. Buragohain, D. Agrawal, and S. Suri. Medians and beyond: new aggregation techniques for sensor networks. In SenSys '04, pages 239-249. ACM Press.
- [14] T. Stathopoulos, L. Girod, J. Heidenmann, and D.Estrin. Mote herding for tiered wireless sensor networks. In CENS Technical Report 58.
- [15] H. Wang, D. Estrin, and L. Girod. Preprocessing in a tiered sensor network for habitat monitoring. EURASIP JASP Special Issue on Sensor Networks, pages 392-401.
- [16] M. Yarvis, N. Kushalnagar, H. Singh, A.Rangarajan, Y. Liu, and S. Singh. Exploiting heterogeneity in sensor networks. In INFOCOM '05, pages 366-377.

- [17] Y. J. Zhao, R. Govindan, and D. Estrin. Computing aggregates for monitoring wireless sensor networks. In *SNPA '03*.
- [18] G. Sagy, D. Keren, I. Sharfman, and A. Schuster. Distributed threshold querying of general functions by a difference of monotonic representation.
- [19] I. Sharfman, A. Schuster, and D. Keren. Shape sensitive geometric monitoring. In *PODS '08*.
- [20] S. Babu and J. Widom. Continuous queries over data streams. *SIGMOD Rec*, 30(3):109-120, 2001.
- [21] M. Cherniack, H. Balakrisman, M. Balazinska, D. Canrey, U. Cetintemel, Y. Xing, and S. Zdonik. Scalable Distributed Stream Processing. In *CIDR '03*.
- [22] D. Carney, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, G. Seidman, M. Stonebraker, N. Tatbul, and S.B. Zdonik. Monitoring streams – a new class of data management applications. In *VLDB, '02*.
- [23] M. Sharaf, J. Beaver, A. Labrinidis, and P. K. Chrysanthis. TiNA: a scheme for temporal coherency-aware in-network aggregation. In *MobiDE '03*.
- [24] M. Charikar, K. Chen, and M. Farach – Colton. Finding frequent items in data streams. In *ICALP '02*.
- [25] P. B. Gibbons and S. Tirthapura. Estimating simple functions on the union of data streams. In *SPAA '01*, pages 281-291, ACM Press.
- [26] A. Manjhi, V. Shkapenyuk, K. Dhamdhere, and C. Olston. Finding (recently) frequent items in distributed data streams. In *ICDE '05*, pages 767-778, IEEE Computer Society.
- [27] G. S. Manku and R. Motwani. Approximate frequency counts over data streams. In *VLDB '02*, pages 346-357.
- [28] http://courses.ced.tuc.gr/digItem.jsp?ditem_id=1710&ditem_type=1&ditem_fname=COMP516slides_21
- [29] <http://db.csail.mit.edu/labdata/labdata.html>